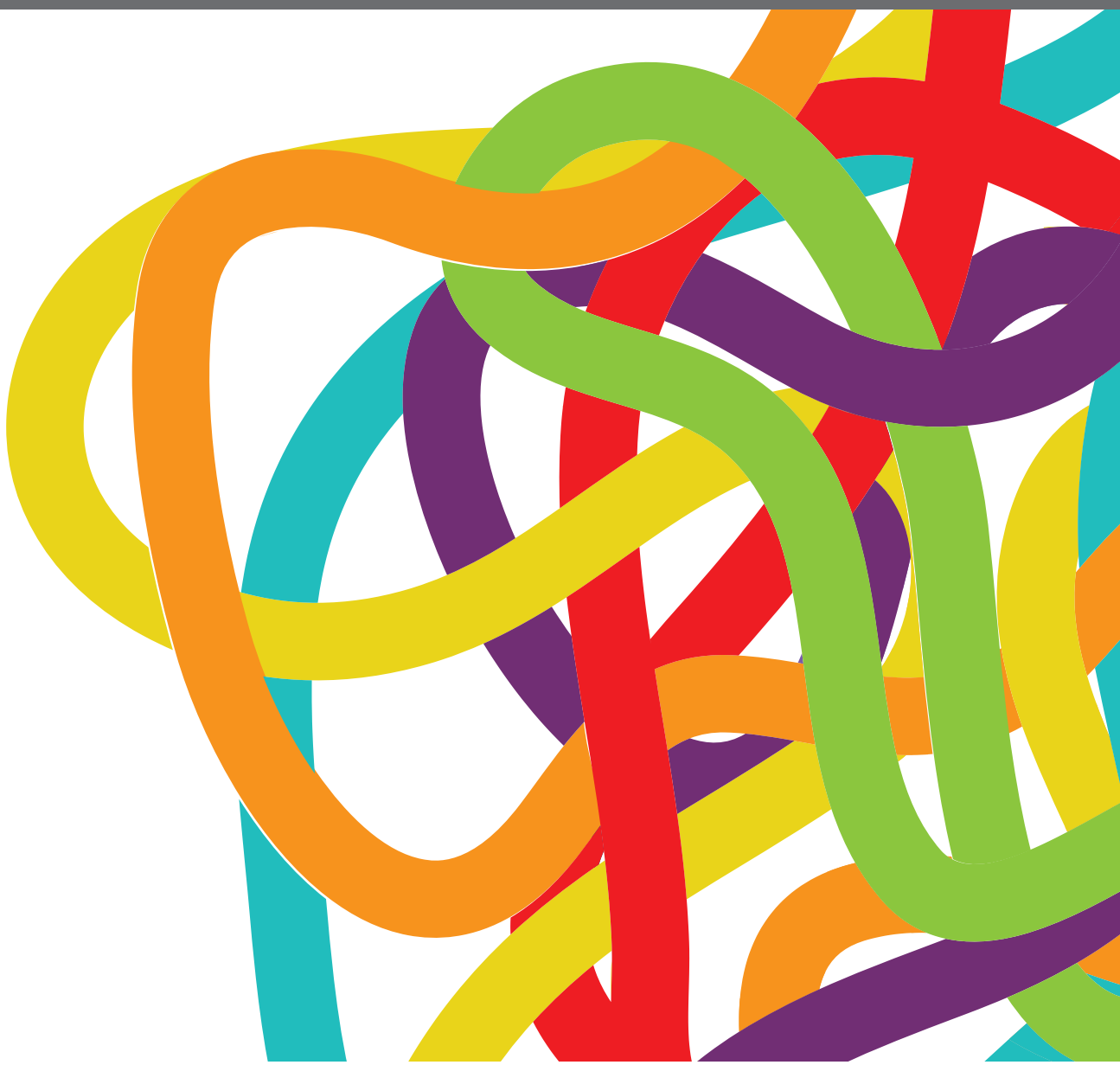
AUTOMATION AND ARTIFICIAL INTELLIGENCE IN RADIATION ONCOLOGY

EDITED BY: Savino Cilla and Jose Eduardo Villarreal Barajas
PUBLISHED IN: Frontiers in Oncology



frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-83250-594-6
DOI 10.3389/978-2-83250-594-6

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

AUTOMATION AND ARTIFICIAL INTELLIGENCE IN RADIATION ONCOLOGY

Topic Editors:
Savino Cilla, Gemalli Molise Hospital, Italy
Jose Eduardo Villarreal Barajas, Royal Devon and Exeter Hospital, United Kingdom

Citation: Cilla, S., Barajas, J. E. V., eds. (2022). Automation and Artificial Intelligence in Radiation Oncology. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-83250-594-6

Table of Contents

frontiers | Frontiers in Oncology

TYPE Editorial
PUBLISHED 12 October 2022
DOI 10.3389/fonc.2022.1038834

OPEN ACCESS

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

*CORRESPONDENCE
Savino Cilla
savinocilla@gmail.com;
savino.cilla@gemellimolise.it

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

RECEIVED 07 September 2022
ACCEPTED 26 September 2022
PUBLISHED 12 October 2022

CITATION
Cilla S and Barajas JEV (2022) Editorial: Automation and artificial intelligence in radiation oncology.
Front. Oncol. 12:1038834.
doi: 10.3389/fonc.2022.1038834

Editorial: Automation and artificial intelligence in radiation oncology

Savino Cilla[1]* and Jose Eduardo Villarreal Barajas[2]

[1]Medical Physics Unit, Gemelli Molise Hospital, Campobasso, Italy, [2]Royal Devon University Healthcare Medical Physics Department,Royal Devon and Exeter Hospital, Exeter, United Kingdom

KEYWORDS

artificial intelligence (AI), automation, machine learning, radiomics, radiation oncology

Editorial on the Research Topic
Automation and artificial intelligence in radiation oncology

The ongoing advancement of radiation oncology has always been significantly influenced by technology. A number of therapeutic advancements have been made recently as a result of technology-driven advancements in radiotherapy planning and delivery. In particular Particularly in the field of radiation oncology, artificial intelligence approaches are spreading more widely and moving from the realm of specialized research to that of accepted clinical practice. Automation and big data analysis have drawn a new era of treating cancer patients with precision and outcome prediction. The continues increase of computing power together with the improvement of treatment accuracy in fighting cancers could lead to huge progress in increasing patient's outcomes and survival rate. The integration of artificial intelligence with modern radiation therapy technologies has the potential to herald an unprecedented change for the field of radiation oncology.

The aim of this Topic was to collate original researches focusing on new developments in the application of machine learning and deep learning processes, patient outcome prediction, treatment technique improvements with automation and applications of radiomics, an emerging and promising research field based on quantitative imaging technology in the radiation oncology field. All these aspects have been well-captured in the present Research Topic which has been successfully launched in Frontiers in Oncology. We were thrilled to get a large number of contributions from authors of their most recent research findings on automation and artificial intelligence techniques for radiation oncology purposes. Twenty-one papers were finally accepted after rigorous reviews for a total of 177 authors. Contributions came from various nations and regions, including China, Italy, South Korea, Thailand, the United States and Indonesia.

Several researchers investigated the possibility for automated treatment planning solutions generated by AI algorithms to improve quality, decrease variability, and shorten planning times. One of the most time-consuming step of radiation therapy is the manual segmentation of target and normal structures, which is subject to high intra- and inter-

observer variability. Recently, numerous research groups have been focusing on the use of AI to increase the speed and accuracy in the definition of clinical target volumes for treatment. Convolutional neural network-based deep learning models have made significant advancements and demonstrated major promise as tools for automated segmentation of target volumes and organs-at-risk (OARs). For CTV delineation in cervical cancer, Liu et al. suggested a novel adversarial deep-learning-based auto-segmentation algorithm. To directly test the model and reduce inter- and intra-observer variability, a three-stage multicenter randomized controlled evaluation procedure was created. The evaluated AI model was shown to be precise and on par with the manual CTV segmentation in patients with cervical cancer. By integrating the fully convolutional network (FCN) and atrous convolution deep learning techniques, Xie et al. sought to fully automate the organs segmentation. According to the authors, this network model may efficiently increase the precision of automated segmentation of chest computed tomography images in thoracic radiotherapy. To successfully avoid radiation side effects, precise target volume and OAR delineation is essential in head-neck malignancies. In order to segment the thyroid gland on localized CT scans and to identify the gland as an OAR in radiotherapy, Wen et al. devised a model that incorporated a Spatial Squeeze and Channel Excitation Block (cSE) attention mechanism with HRNet. Due to the low contrast at the tumor's border and the wide range of tumor sizes and morphologies between different stages, the delineation of target volumes in nasopharyngeal cancer is a particularly difficult process. To solve the aforementioned issues, Yang et al. proposed a new three-dimensional (3D) automatic segmentation system that uses cascaded multiscale local augmentation of convolutional neural networks. The suggested approach may enhance and facilitate clinical applications by precisely segmenting NPC in CT scans from multi-institutional datasets. The possibility of using deep learning to automatically delineate multiple contours for breast cancer radiation therapy was examined by Dai et al. Their study showed that the devolped deep learning techniques can reliably produce target and OAR contours on planning CT and daily synthetic CT images from CBCT images, which may significantly speed up the re-planning process and satisfy the needs of online plan adaptation.

The goal of radiotherapy plan optimization is to find the optimal balance between two competing goals: delivering the highest radiation dose to the target while delivering the lowest radiation dose to nearby OARs. These OARs are typically given a numerical weight to reflect their relative importance in the optimization calculus. In order to produce a plan that fulfills the minimal acceptable threshold for each aim, physicists must repeatedly adjust the parameters that control radiation dose deposition. This fine-tuning typically goes on until time resources run out, at which point the planner is compelled to decide on the best plan he can achieve. Therefore, plan quality may strongly vary between planners and between clinical institutions. Additionally, the time and labor requirements of the existing planning paradigm can put patients at risk for delays and potentially suboptimal care while also appearing to be insurmountable barriers to adaptive radiotherapy. In this perspective, the introduction of automated systems may translate in important benefits as time saving, high quality planning, and protocol standardization, as reported by Cagni et al. Together with template-based iterative planning (1), the use of knowledge-based automated planning (KBP) techniques has recently received a special attention. Using machine learning techniques that learn from databases of previous clinically acceptable plans, KBP may assist physicists and radiation oncologists to find the best solutions for planning optimization. Castriconi et al. implemented a KBP solution for right and left-sided whole breast treatment through a new volumetric technique mimicking conventional tangential fields irradiation that can efficiently replace manually optimized plans. Xu et al. evaluated the effectiveness of a proton-specific KBP model in the development of robustly optimized intensity-modulated proton therapy plans for the treatment of advanced head and neck cancer patients, reporting that the quality of KBP plans is comparable to, and occasionally even exceeds, that of the expert plans. In this clinical setting, radiation therapy is the primary therapeutic option for early and locally progressed nasopharyngeal cancer. When compared to computed tomography, magnetic resonance imaging (MRI) has the advantage of high soft-tissue resolution, but it does not provide information on electron density (ED) for planning radiotherapy. To provide the necessary ED data for MRI-only planning, Ma et al. created a pseudo-CT generating approach. The suggested deep learning model can precisely predict CT from MRI, and the resulting pCT can be used in accurate dose estimations.

Proton therapy may also greatly benefit of using AI strategies and techniques. For example, a great interest in beam angle optimization research has been on the rise recently for proton therapy, in order to generate optimal proton plan. Cheon et al. suggested a method for beam angle optimization based on a convolutional neural network to automatically find the optimal beam angles for proton treatments set with the double-scattering delivery approach (BAODS-Net). This approach dramatically reduced the planning time increasing the potential for a real-time adaptive proton radiotherapy. Furthermore, it is well known that a double scattering proton system's beam output fluctuates depending on the beam option, range, and modulation, translating in inaccurate modeling by the treatment planning system. Because of this, the majority of proton centers with a double scattering beam system must measure the output of patient-specific proton beams in a water phantom in order to determine the necessary machine output. Three machine learning algorithms were developed by Zhu et al. to efficiently estimate the output of a proton beam using a Gaussian process regression model with various kernels. One of

these models showed accurate estimation, meeting the ±3% clinical requirement.

A second fundamental point is the growing application of artificial intelligence techniques for prediction purposes.

First, methods for machine learning have been investigated, with an emphasis on applications for machine and patient-specific quality assurance (QA) (2). The performance of various delivery system components, such as the multileaf collimator (MLC), imaging system, mechanical parameters, and dosimetric parameters, can be examined using machine learning. As a result, a "virtual" QA may forecast passing rates using different measurement techniques, different treatment planning systems, and different treatment delivery machines across multiple institutions. In this topic, Huang et al. introduced a new QA prediction model based on UNet++ using the dose distribution as input. This model was able to predict the gamma-pass rates for various gamma criteria as well as provide classification results.

Secondly, AI models have recently demonstrated the potential for effective toxicity prediction aiming to limit radiotherapy-related side effects (3). The proactive, rather than reactive, management of acute and late toxicities in patients is exacerbated by the mostly unpredictable occurrence and/or intensity of such side effects. Nevertheless, it is possible to create predictive models of radiation toxicities based on imaging data and risk variables, such as specific clinical traits, germline genetic alterations, and the radiation dose distributions, and these models can be used to guide treatment planning. Additionally, multi-omic data may capture complex tumor features, contributing to a comprehensive patient risk assessment. In particular, two complementary strategies have emerged in recent years: the integration of patient-specific biological risk factors into dose–volume-based outcome models (called radiogenomics), and the integration of imaging together with treatment-related and biological data for outcomes prediction (called radiomics). Both these approaches have the potential to develop personalized and tailored treatment plans.

The current advances and challenges in radiomics of brain tumors have been highlighted by Yi et al. The authors demonstrate how radiomics, in contrast to conventional brain imaging, offers quantitative data related to important biologic characteristics and application of deep learning which sheds light on the complete automation of imaging diagnosis.

In patients with ovarian cancer, Yu et al. assessed the accuracy of radiomics characteristics based on multiparameter magnetic resonance imaging for peritoneal carcinomatosis. A multi-factor logistic regression method was utilized to create a radiomics nomogram in combination with radiomics features and clinicopathological risk factors, reporting a better diagnostic effect than the clinical model, able to identify peritoneal carcinomatosis in ovarian cancer patients before surgery.

Gastric cancer is a typical heterogeneous malignant tumor. Chemotherapy is ineffective against this tumor and this is a common cause of tumor recurrence and metastasis. Conventional pathological TNM prediction focuses on cancer cells to predict prognosis, but they do not provide adequate prediction. Jin et al. devoloped a radiomics signature in order to predict patients' overall survival and disease-free survival after undergoing surgery for gastric cancer. The radiomics trait-associated genes identified clinically significant biological pathways and possible drug metabolic mechanisms for chemotherapy agents.

With respect to rectal cancer, although several prognosis nomograms have been established, statistical tools for predicting long-term survival in rectal cancer are lacking. Additionally, neither qualitative nor quantitative imaging findings were included in modern prognostic analyses. Nie et al. used a radiomics signatures and multiparametric MRI data to build a predictive model able to predict 5-year overall survival for patients with advanced rectal cancer. An interesting aspect of the detected radiomics signature was that it contained three from dynamic contrast-enhanced (DCE)-MRI, four from anatomical MRI, and one from functional diffusion-weighted imaging (DWI). This brought attention to how crucial multiparametric MRI is in addressing the problem of estimating long-term survival in rectal cancer.

The 5-year survival rate of lung cancer is significantly increased by early detection and treatment. Immunotherapy has recently grown quickly, caught the attention of more and more oncologists, and established itself as a significant area of study in the field of tumor therapy. The immunotherapy against programmed cell death protein 1 (programmed death-1, PD-1) and its ligand 1 (programmed death ligand-1, PD-L1) has been used in non-small cell lung cancer (NSCLC), and good results have been achieved in patients, especially in individuals with high expression of PD-L1 (4). Finding a fresh method to gauge PD-L1 expression level is thus critically required. Based on this premises, Li et al. aimed to evaluate the expression of PD-L1 in patients with NSCLC by radiomic features of 18F-FDG PET/CT and clinicopathological characteristics. In order to predict PD-L1 expression in individual NSCLC patients, the authors generated a prediction model that used both the radiomic signature and clinicopathologic risk variables. Significant correlations were found between the radiomic signature and PD-L1 expression in lung cancers. The aforementioned papers reported how the recently emerged radiomics methods are able to extract a large number of spatial features from medical images in order to predict therapeutic responses. Enlightened by those works, a recent approach called "dosiomics" has been put out, in an effort to extract spatial features from dose distribution for radiotherapy response prediction. It has been shown that the dosiomics features may be able to improve radiation therapy toxicity prediction since they have more dose distribution data than DVH features (5). For example, dosiomics informations can be used for the prediction of radiation pneumonitis.

Puttanawarut et al. investigated the feasibility of of dosiomics and radiomics features to predict the development of radiation pneumonitis over traditional dose-volume histogram. Then, four predictive models for radiation pneumonitis were compared on a esophageal and a lung cancer datasets, resulting in predictive performance of the dosiomics- and radiomics-based models significantly higher than that of the DVH-based model.

Finally, it must be underlined that studies on the dosiomics and radiomics features are still in the early stage and yet there exist some concerns regarding the stability and generalizability of this texture analysis. For example, it has been reported that some dosiomics features are unstable across various grid resolutions or dose calculation algorithms (6), showing that the reproducibility of dosiomics features depends on the process of producing images. Moreover, for many cancers, because of inter-fractional error, a different total number of fractions may induce different error behavior. These errors may also further affect the reproducibility of dosiomics features. Puttanawarut et al. investigated the stability of dosiomics features under random inter-fractional error and evaluated the uncertainties in the values of dosiomic features under inter-fractional error with IMRT and VMAT in a lung cancer dataset. The authors reported that some dosiomics features were found not reliable under inter-fractional error and with lower fraction numbers.

Conclusions

The ultimate goal of this Research Topic was to promote research and development of automation, advanced computing and AI applications in radiation oncology by publishing high-quality research articles. The 21 papers published in this Research Topic reported promising results and offered new and original perspectives regarding the role of AI in radiation oncology. We thank all the authors of the published papers for their valuable contributions and the referees for their rigorous review.

Author contributions

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

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

References

1. Cilla S, Ianiro A, Romano C, Deodato F, Macchia G, Buwenge M, et al. Template-based automation of treatment planning in advanced radiotherapy: a comprehensive dosimetric and clinical evaluation. *Sci Rep* (2020) 10:423. doi: 10.1038/s41598-019-56966-y

2. Chan MF, Witztum A, Valdes G. Integration of AI and machine learning in radiotherapy QA. *Front Artif Intell* (2020) 3:577620. doi: 10.3389/frai.2020.577620

3. Isaksson LJ, Pepa M, Zaffaroni M, Marvaso G, Alterio D, Volpe S, et al. Machine learning-based models for prediction of toxicity outcomes in radiotherapy. *Front Oncol* (2020) 10:790. doi: 10.3389/fonc.2020.00790

4. Brahmer JR, Tykodi SS, Chow LQ, Hwu WJ, Topalian SL, Hwu P, et al. Safety and activity of anti-PD-L1 antibody in patients with advanced cancer. *N Engl J Med* (2012) 366(26):2455–65. doi: 10.1056/NEJMoa1200694

5. 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

6. Placidi L, Lenkowicz J, Cusumano D, Boldrini L, Dinapoli N, Valentini V. Stability of dosiomics features extraction on grid resolution and algorithm for radiotherapy dose calculation. *Phys Med* (2020) 77:30–5. doi: 10.1016/j.ejmp.2020.07.022

frontiers in Oncology

ORIGINAL RESEARCH
published: 20 July 2021
doi: 10.3389/fonc.2021.700343

Virtual Patient-Specific Quality Assurance of IMRT Using UNet++: Classification, Gamma Passing Rates Prediction, and Dose Difference Prediction

Ying Huang[1†], Yifei Pi[2†], Kui Ma[3], Xiaojuan Miao[4], Sichao Fu[4], Hua Chen[1], Hao Wang[1], Hengle Gu[1], Yan Shao[1], Yanhua Duan[1], Aihui Feng[1], Jiyong Wang[5], Ruxin Cai[1], Weihai Zhuo[6] and Zhiyong Xu[1*]*

[1] Shanghai Chest Hospital, Shanghai Jiaotong University, Shanghai, China, [2] Department of Radiotherapy, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China, [3] Varian Medical Systems (China), Beijing, China, [4] Department of Hematology, Western Theater General Hospital, Chengdu, China, [5] PingAn Health Technology Co. Ltd, Shanghai, China, [6] Key Laboratory of Nuclear Physics and Ion Beam Application Ministry of Education, Fudan University, Shanghai, China

The dose verification in radiotherapy quality assurance (QA) is time-consuming and places a heavy workload on medical physicists. To provide a clinical tool to perform patient specific QA accurately, the UNet++ is investigated to classify failed or pass fields (the GPR lower than 85% is considered "failed" while the GPR higher than 85% is considered "pass"), predict gamma passing rates (GPR) for different gamma criteria, and predict dose difference from virtual patient-specific quality assurance in radiotherapy. UNet++ was trained and validated with 473 fields and tested with 95 fields. All plans used Portal Dosimetry for dose verification pre-treatment. Planar dose distribution of each field was used as the input for UNet++, with QA classification results, gamma passing rates of different gamma criteria, and dose difference were used as the output. In the test set, the accuracy of the classification model was 95.79%. The mean absolute error (MAE) were 0.82, 0.88, 2.11, 2.52, and the root mean squared error (RMSE) were 1.38, 1.57, 3.33, 3.72 for 3%/3mm, 3%/2 mm, 2%/3 mm, 2%/2 mm, respectively. The trend and position of the predicted dose difference were consistent with the measured dose difference. In conclusion, the Virtual QA based on UNet++ can be used to classify the field passed or not, predict gamma pass rate for different gamma criteria, and predict dose difference. The results show that UNet++ based Virtual QA is promising in quality assurance for radiotherapy.

Keywords: deep learning, radiotherapy, quality assurance, prediction model, dose difference

OPEN ACCESS

Edited by:
Jose Eduardo Villarreal Barajas, Royal Devon and Exeter Hospital, United Kingdom

Reviewed by:
Maria F. Chan, Memorial Sloan Kettering Cancer Center, United States
Yuanzeng Min, University of Science and Technology of China, China

****Correspondence:***
Weihai Zhuo
whzhuo@fudan.edu.cn
Zhiyong Xu
xzyong12vip@sina.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: *26 April 2021*
Accepted: *30 June 2021*
Published: *20 July 2021*

Citation:
Huang Y, Pi Y, Ma K, Miao X, Fu S, Chen H, Wang H, Gu H, Shao Y, Duan Y, Feng A, Wang J, Cai R, Zhuo W and Xu Z (2021) Virtual Patient-Specific Quality Assurance of IMRT Using UNet++: Classification, Gamma Passing Rates Prediction, and Dose Difference Prediction. Front. Oncol. 11:700343. doi: 10.3389/fonc.2021.700343

INTRODUCTION

Intensity-modulated radiation therapy (IMRT), a widely used treatment modality for cancer patients, provides highly conformal dose distribution to the target while sparing surrounding healthy tissues (1). Quality assurance is performed to confirm the accuracy of dose calculation, data transmission, linear accelerator performance, radiotherapy positioning, and dosimeter response accuracy (2–6). It is essential to ensure the reliability of treatment delivery and improve patient safety. The process commonly involves comparing the calculated dose distribution or fluence with the measured dose distribution or fluence (7). However, the implementation of patient-specific QA measurement is time-consuming and places a heavy workload on medical physicists (8). Additionally, the measurement work takes a lot of clinical treatment time, which is unrealistic for busy centers. Furthermore, the QA process needs to be accomplished prior to treatment. For those QA results which do not meet the predefined "pass" criteria, re-planning and verification will cause an inevitable treatment delay. Therefore, a more streamlined, less-resourced, and automated patient-specific QA method for dose verification is necessary for radiotherapy centers.

With the development of machine learning and deep learning and its application in QA results prediction, the efficiency of patient-specific QA is expected to be improved (9–16). Valdes, Lam, Li (9–12) etc. to established prediction models for the gamma passing rate (GPR) based on the complexity parameters of the TPS plan. Other researchers (13, 14) investigated the deep learning algorithms to establish GPR prediction models based on planar dose distribution. Granville (15) used support vector machines (SVMs) to classify cold, hot, and normal plans based on the plan complexity parameters and accelerator performance parameters. Li (12) discussed the prediction model of whether the plan could pass the threshold using machine learning method. The above researchers have developed an accurate prediction model of QA results, confirming the feasibility of using machine learning or deep learning for patient-specific QA.

Previous prediction models based on machine learning or deep learning were only the results of dose verification but could not provide detailed information of dose difference (9–18). Predicting the trend and position of dose difference is an important work in automatic patient-specific QA in the near future.

In this study, we proposed a novel QA prediction model based on UNet + + using the planar dose distribution as input. A model can (a) provide the classification results whether the field QA passes; (b) predict the GPRs of different gamma criteria; (c) predict the trend and position of dose difference. The prediction model allows physicists to pre-mark potentially failed fields in a proactive way, analyze dose difference simultaneously and reduce patient delays associated with unqualified measurements. Additionally, it could reduce patient-specific QA measurements to verify data transmission and delivery accuracy combination with other tools. The model is expected to be a practical clinical tool to perform patient-specific QA accurately and provide new ideas for the development of virtual QA and process optimization.

MATERIALS AND METHODS

Data Collection

109 IMRT plans (including 568 fields) from December 2019 to May 2020 were selected. All plans were generated in the Eclipse version11 (Varian Medical Systems, Palo Alto, CA). Dose distributions were calculated using the Acuros External Beam (AXB, ver.11.0.31, Varian Medical Systems, Palo Alto, CA) with a dose calculation grid of 2.5 mm. Each plan was delivered using a linear accelerator equipped with a Varian Millennium 120 MLC. Patient-specific dose verification was performed prior to treatment using the actual angle by Portal Dosimetry. Daily dose calibration was performed during the period of data collection.

PD Mining, an in-house software developed by C #, was used to register, resample, and compare the calculated and measured dose distribution. This software was developed based on the interface of Varian ESAPI portal dosimetry and verified with the manual processing results. All patient dose data was searched through patient ID. The gamma analysis results and dose difference were obtained and exported to the local file in the form of text file automatically.

As 2%/2mm was the most sensitive criterion to detect clinically relevant errors (19, 20), it was used for establishing the classification model. If the GPR was higher than 85%, the field was considered to pass the QA, vice versa. The GPRs in the criteria of 3%/3mm, 3%/2mm, 2%/3mm were calculated at the same time. Absolute dose mode and 10% dose threshold were used for the above GPR analysis.

Data Preprocessing

Pylinac library was used to extract the original image of planar dose distribution in the resolution of 1190 × 1190. Then the redundant information such as the frame and coordinate axis was cut off to get the image with the resolution 968 × 968. Flipping (horizontal random lip probability: 0.5, vertical random flip probability: 0.5) and random clipping were used to prevent overfitting. The images for the training set were randomly cropped from 968 × 968 to 960 × 960. The images in the test set were cut from the center, sizing from 968 × 968 to 960 × 960.

UNet++ Architecture

Based on the traditional medical image processing network UNet, UNet++ enables the network to learn important features of different depths through a series of nested and dense jumping connections. By adopting deep supervision, UNet++ allows model complexity tuning to balance speed and performance optimization (21).

Figure 1 shows the architecture of the UNet++ used in this study. The downsample encoding channel of $X^{0,0}\rightarrow X^{1,0}\rightarrow X^{2,0}\rightarrow X^{3,0}\rightarrow X^{4,0}$ adopted ResNet-101 architecture as the backbone network, and then the image size was restored by the corresponding upsample decoding nodes with skip connections. The planar dose distribution was input from $X^{0,0}$, and the predicted dose difference image output of the same size was obtained from $X^{0,4}$ after passing through the UNet++ network. At the same time, $X^{0,4}$ was downsampled through three max-pooling layers and two bottleneck layers, and the

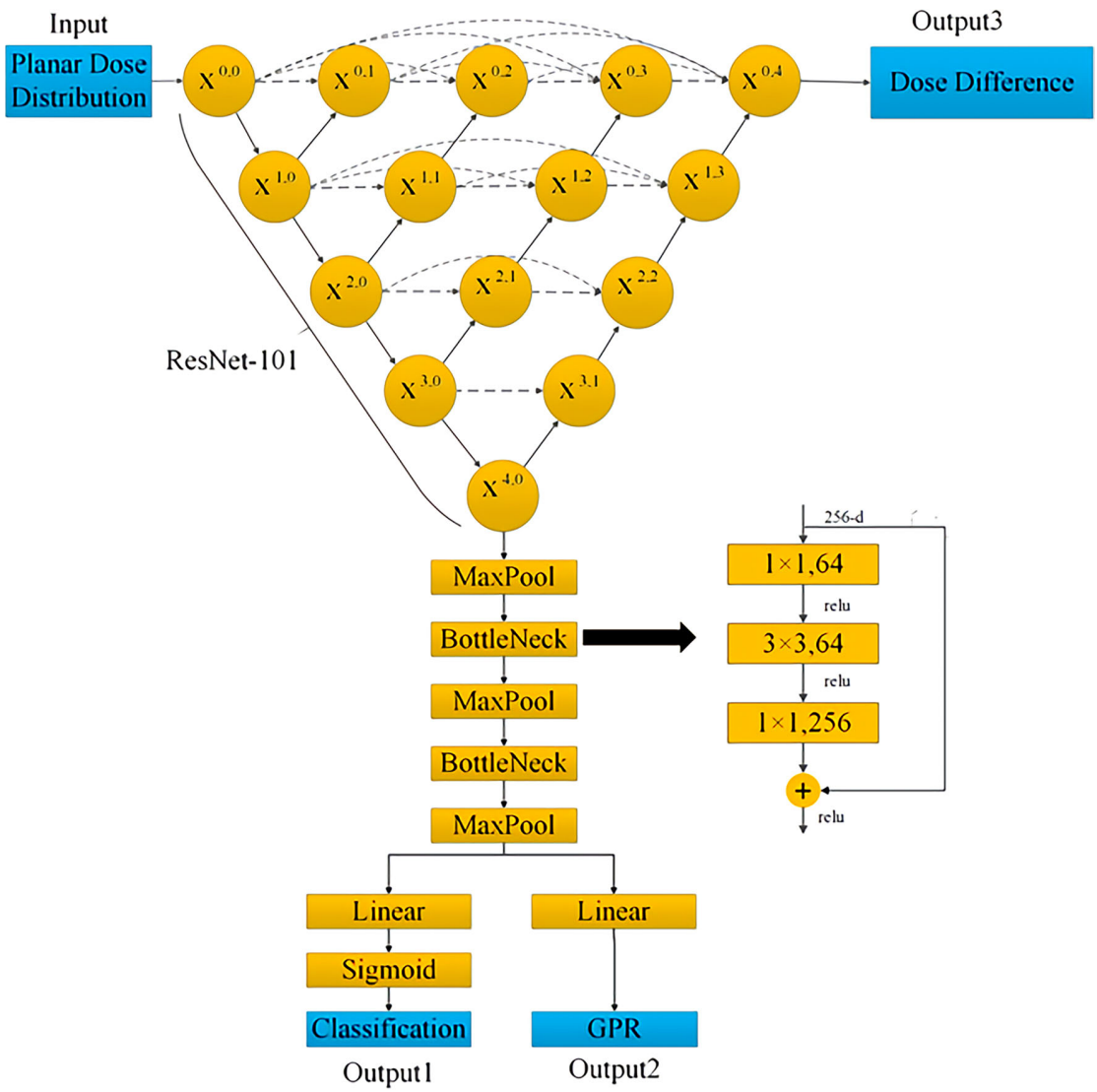


FIGURE 1 | Architecture of the UNet++ used for prediction model.

linear layer was connected in series to predict the four gamma criteria GPR and the classification results.

Model Training and Evaluation

To reduce dose difference prediction error caused by image scaling, the original resolution was adopted for the input and output. This processing would increase network memory and training time. Therefore, small batches rather than N-fold validation were selected. We randomly selected 95 fields (about 1/6) from the collected 568 samples as the test set. Four of the remaining samples (378 fields) were used as the training set and one as the validation set (95 fields).

The mean square error (MSE) loss function was used to evaluate the regression error of dose difference and GPR, and the binary cross-entropy was used to evaluate the classification error. The total loss was obtained by the weighted (1/3, 1/3, 1/3) sum of the three errors. The commonly used Adam optimizer (22) was adopted to learn the back-propagation error. The initial learning rate was set to 0.001, and it decreased exponentially as training going on, with the dropping rate setting to 0.9. A Mini batch method was used to train the model, the batch size was set to 2, and the epoch was set to 120. The prediction model was built by the open-source pytorch library. The entire training cost about 32 hours on the NVIDIA GTX-3080 GPU.

RESULTS

Learning Curve for the Prediction Model

It is expected that more epochs would give rise to higher prediction accuracy. Thus, there typically exists a minimum number of epochs beyond which the increase in prediction

accuracy would saturate. **Figure 2** shows how the number of epochs will affect the accuracy for the prediction model. With more epochs, the loss on the training data, validation data and testing data decrease. With 60 epochs, the testing, validation and training loss converge at a stable level, indicating that increasing the epochs of training sets may not yield further improvement in the accuracy of the prediction model.

Performance of Classification Model

The proportion of GPR less than 85% was 7.37% (7/95) and 8.42% (8/95) in the validation set and test set, respectively. As shown in **Table 1**, the sensitivity of the validation set is 57.14%, the specificity is 100%, and the accuracy is 96.84%. For the test set, the sensitivity is 62.50%, the specificity is 98.85%, and the accuracy of the classification model is 95.79%. From the analysis for the failed fields in measurement, while the prediction results are pass, the GPRs for these fields are near 85%, and the predicted GPRs are higher than 85%, so the classification results are pass fields.

GPR Prediction for Different Gamma Criteria

As shown in **Table 2**, the MAE and RMSE of the validation set and test set increase with stricter gamma criteria. In the test set, the smallest MAE and RMSE are 0.82 and 1.38 under 3%/3mm. The 2%/2mm gamma criteria has the largest MAE and RMSE in the test set, which are 2.52 and 3.72, respectively.

The distribution of errors between the predicted GPR and the measured GPR under different gamma criteria are shown in **Figure 3**. The prediction errors among different gamma groups (90%-100%, 80%-90%, and < 80%) are compared. The accuracy of the prediction model is affected by the measured value itself. The higher the measured GPR, the smaller prediction errors between the measured and predicted GPR are observed.

Dose Difference Prediction

In the prediction of dose difference, the dose difference and the histogram of distribution relative to TPS for the pass field and fail field are shown in **Figures 4**, **5**, respectively. The position and trend of predicted dose difference are consistent with the measured dose difference.

Figure 4 shows the dose difference for a passing field of a patient. **Figure 4A** shows the dose difference between the measured planar dose and TPS calculated, and **Figure 4B** shows the dose difference between the predicted and TPS calculated planar dose. Even for a passing field, there are still some pixels with large dose difference, and their positions can be obtained. **Figure 4C** is the histogram of the dose difference between measured and TPS calculated planar dose for one

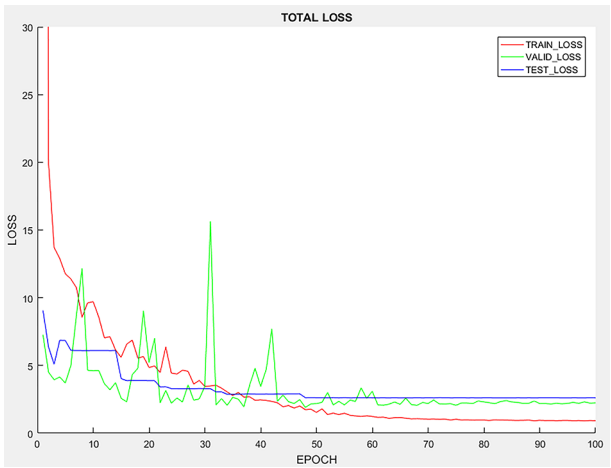


FIGURE 2 | Learning curve of the prediction model.

TABLE 1 | Results of classification model.

	Predicted-Fail	Predicted-Pass	
Validation set (95)			
Measured Fail	4	3	57.14%
Measured Pass	0	88	100.00%
Test set (95)			
Measured Fail	5	3	62.50%
Measured Pass	1	86	98.85%

The results are classified into four categories: failed measurement results and failed prediction results, TP; passed measurement results and passed prediction results (TN); passed measurement results and failed prediction results (FP); failed measurement results and passed prediction result passed (FN).

TABLE 2 | MAE and RMSE for different gamma criteria.

	MAE		RMSE	
	Validation set	Test set	Validation set	Test set
3%/3mm	0.79	0.82	1.28	1.38
3%/2mm	0.93	0.88	1.50	1.57
2%/3mm	2.01	2.11	2.31	3.33
2%/2mm	2.17	2.52	3.00	3.72

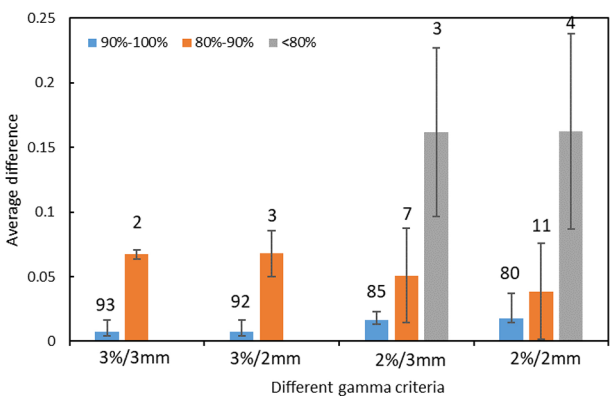


FIGURE 3 | The distribution of prediction errors in test set among different groups under different gamma criteria. (Error bar: Mean ± standard deviation. The number on the vertical axis represents the number of fields for different groups).

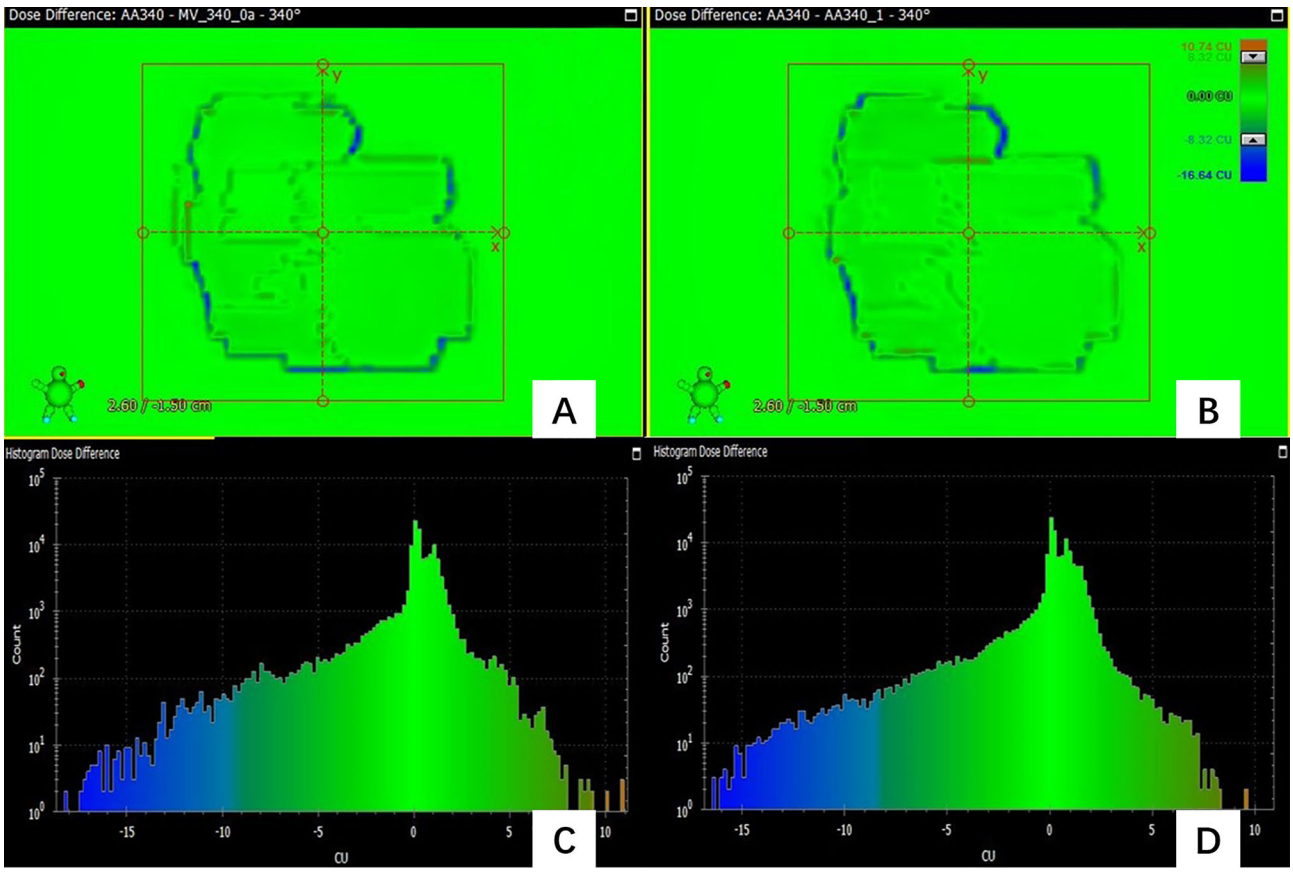


FIGURE 4 | **(A)** the dose difference between the measured planar dose and TPS calculated **(B)** the dose difference between the predicted and TPS calculated planar dose. **(C)** the histogram of the dose difference between measured and TPS calculated planar dose for one pass field **(D)** the histogram of the predicted planar dose relative to TPS calculated dose.

failed field, and **Figure 4D** is the histogram of the predicted planar dose relative to TPS calculated dose. The histogram of predicted dose difference is consistent with the measured dose difference, the maximum and the minimum dose difference are almost the same.

As **Figure 5** shows, the predicted dose difference and measured dose difference for one of the failed fields and the histogram of dose difference. Large dose difference mainly located in the edge, and the measured dose is lower than the TPS calculated dose. The trend and position of the predicted dose difference is consistent with the measured dose difference, indicating that the virtual QA results could be used as guidance for the analysis of dose difference and plan redesign. By comparing and analyzing the dose difference for pass field and failed field, we found that the number of pixels having the large dose difference in the failed field was larger than in the pass field.

DISCUSSIONS

IMRT is a complex technology in radiotherapy, so special QA is required to ensure the accuracy of dose delivery. In this study, the UNet++ was used to classify QA results, GPR prediction of different gamma criteria and accurate dose difference prediction based on planar dose distribution. The accuracy of the classification model was 95.79%; there were small RMSE (1.38-3.72) and MAE (0.82-2.52) between the measured and the predicted GPR in the test set, and the trend and position of predicted dose difference were consistent with the measured dose difference. The results showed that the prospect of realizing virtual patient-specific QA with UNet++ and provides a new idea for the optimization of the individual QA process.

QA is required for every patient before treatment to ensure the accuracy of dose delivery (2–6). Dose verification depends on the equipment highly. The resolution and energy response of the detector will affect the result of dose verification. For the fields that fail the threshold, it is necessary to adjust the radiotherapy plan repeatedly, which will bring treatment delay. The process of dose verification is labor-consuming and time-consuming, so it brings more workload to the busy center. Automatic dose verification pre-measurement can mark the plans that fail the verification in advance and predict the dose difference, which is expected to be an effective method to solve the above problems.

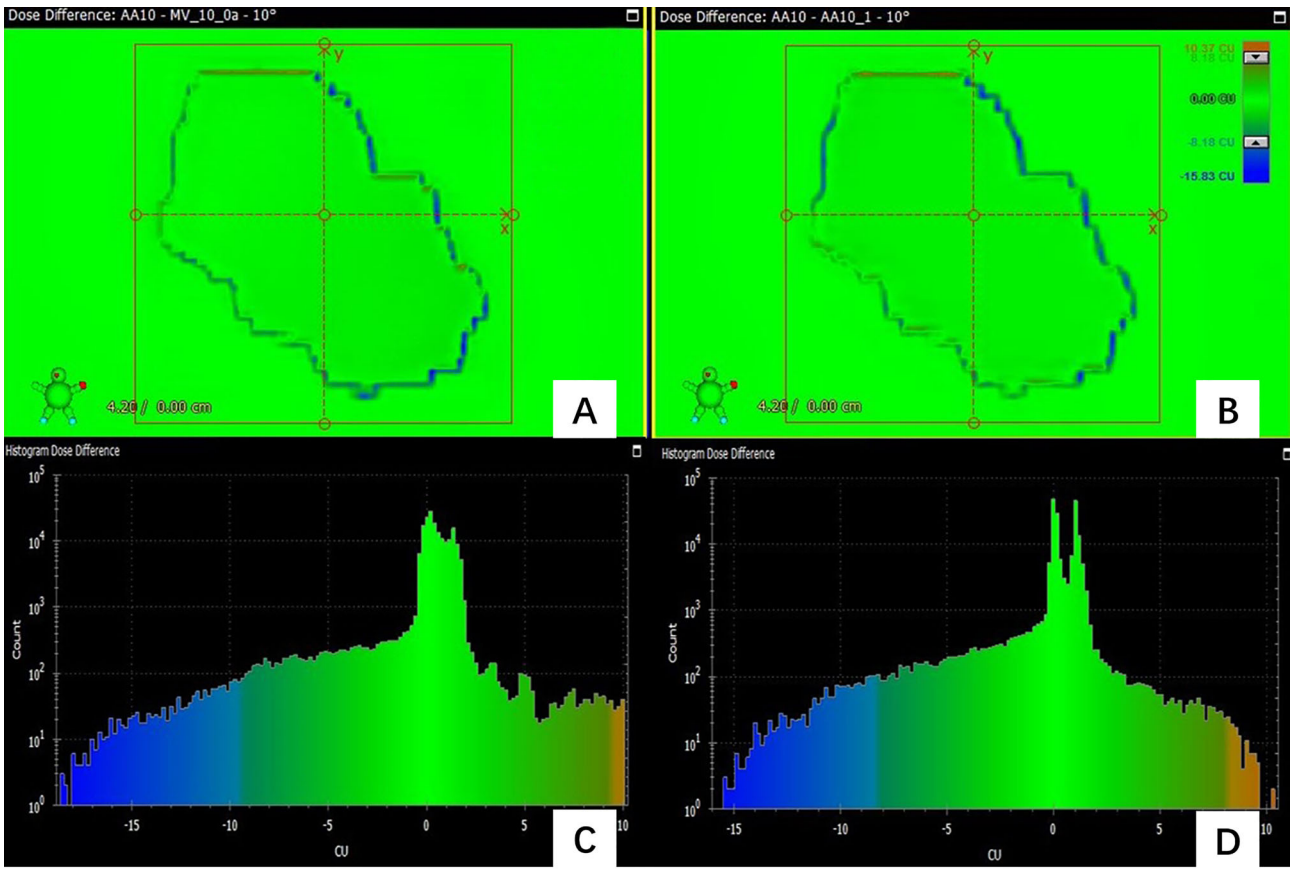


FIGURE 5 | **(A)** the dose difference between the measured planar dose and TPS calculated **(B)** the dose difference between the predicted and TPS calculated planar dose. **(C)** the histogram of the dose difference between measured and TPS calculated planar dose for one failed field **(D)** the histogram of the predicted planar dose relative to TPS calculated dose.

As suggested in the TG 218 (6), the ability of the prediction model to accurately classify plans into "pass" or "fail" based on gamma criteria used is one of the most important indicators to evaluate the clinical feasibility of the model. The prediction of GPRs under different gamma criteria could provide more comprehensive information for physicists to judge whether the plan is acceptable clinically. The dose difference prediction model predicts the trend and location of dose difference, which provides direction for physicists to modify the plan. The paradigm shift of pre-measurement QA will improve the efficiency of dose verification greatly.

Since 2%/2mm is the most sensitive to clinically relevant errors (19, 20), it is selected as the basis of classification model in this study. The accuracy of failed fields was lower compared to pass fields, as the measured GPR were near 85% for some failed fields, and the predicted GPR were higher than 85%, so the predicted results of these fields were passed. Therefore, the selection of appropriate threshold plays an important role in the accuracy of the classification model and the proportion of failed fields needed to improve the accuracy of classification. The classification model gives the physicist a more intuitive result whether the field pass the QA pre-measurement.

The prediction model in this study can give the classification results according to the standard of the classification model, and also give the GPRs of different gamma criteria (3%/3mm, 3%/2mm, 2%/3mm, 2%/2mm). As an example, for one of the fail fields under the classification result, the GPR for 3%/3mm was 98.51% (Error: 0.01%), and GPR was 97.42% (Error: 0.22%) for 3%/2mm. After discussion by physicists and the doctors, the plan can be delivered. Therefore, the prediction of the GPRs of different gamma criteria could make a comprehensive judgment for the clinical enforceability of the plan.

In the prediction of GPR under different gamma criteria, the MAE and RMSE for the measured and predicted GPR of the model increase with the stricter gamma criteria, which is caused by the increase of the uncertainty of the prediction with the stricter gamma criteria. The accuracy of the model is affected by the measured GPR itself. The higher measured GPR is, the higher accuracy. This can be explained that the data with high measured GPR accounts for a large part of our model, so the accuracy of higher GPR prediction is high.

Previous studies only predicted the results of QA using machine learning or deep learning (9–18, 23, 24), but it is impossible to predict the trend and position of dose difference.

The trend and position of dose difference plays a very important role in the analysis of the error source. However, there are no relevant reports about dose difference at present. In this study, prediction of dose difference based on the planar dose distribution is fulfilled. The result of this study that the predicted dose difference is consistent with the position and trend of the measured dose difference, can provide guidance for reason analysis.

It is noting that the pre-treatment prediction model established in this study is an auxiliary tool that needs a reparatory guarantee of the accuracy of the energy calibration and delivery process of the accelerator (2, 4). The purpose of this prediction model is not intended to replace the traditional QA, but to help physicists reduce the measurement burden of patient-specific QA when verifying the dose distribution and optimize the process of QA combined with other methods (25–28). There are still some limitations in this study: 1) As most of the clinical fields are pass fields, which leads to the unbalance of data distribution. Adequate amounts of low GPR plans for model training are needed to improve the accuracy of the model in future. 2) The model is limited to our data that all the data come from the same accelerator, the same energy, the same verification equipment. To expand the universality of the model, the research on models that include a variety of energy, different types of accelerators and dose verification equipment will be done in the future.

CONCLUSION

In this study, we developed a UNet++ based prediction model for patient-specific QA. The prediction model could classify whether the field passes or fail QA, predict GPR of different gamma criteria, predict the trend and location of dose difference and mark the position. The virtual QA tool developed in this study provides a new idea for the optimization of patient-specific QA process, and promote the development of automated patient-specific QA.

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

Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

YH and YP: study concept and design and drafting of the manuscript. YH, KM, and RC: acquisition of data. YH and KM: analysis and interpretation of data. XM, SF, and JW: develop training model. HC, HW, HG, YS, YD, and AF: critical revision of the manuscript for important intellectual content. YH and SF: statistical analysis. WZ and ZX: study supervision. All authors contributed to the article and approved the submitted version.

FUNDING

This study was funded by the Nurture projects for basic research of Shanghai Chest Hospital.

REFERENCES

1. Ma C, Wang R, Zhou S, Wang M, Yue H, Zhang Y, et al. The Structural Similarity Index for IMRT Quality Assurance: Radiomics-Based Error Classification. *Med Phys* (2021) 48:80–93. doi: 10.1002/mp.14559
2. Ezzell GA, Burmeister JW, Dogan N, LoSasso TJ, Mechalakos JG, Mihailidis D, et al. IMRT Commissioning: Multiple Institution Planning and Dosimetry Comparisons, a Report From AAPM Task Group 119. *Med Phys* (2009) 36:5359–73. doi: 10.1118/1.3238104
3. Van Esch A, Huyskens DP, Behrens CF, Samsoe E, Sjolin M, Bjelkengren U, et al. Implementing Rapidarc Into Clinical Routine: A Comprehensive Program From Machine QA to TPS Validation and Patient QA. *Med Phys* (2011) 38:5146–66. doi: 10.1118/1.3622672
4. Klein EE, Hanley J, Bayouth J, Yin FF, Simon W, Dresser S, et al. Task Group 142 Report: Quality Assurance of Medical Accelerators. *Med Phys* (2009) 36:4197–212. doi: 10.1118/1.3190392
5. Smilowitz JB, Das IJ, Feygelman V, Fraass BA, Kry SF, Marshall IR, et al. Commissioning and QA of Treatment Planning Dose Calculations - Megavoltage Photon and Electron Beams. *J Appl Clin Med Phys* (2015) 16:14–34. doi: 10.1120/jacmp.v16i5.5768
6. Miften M, Olch A, Mihailidis D, Moran J, Pawlicki T, Molineu A, et al. Tolerance Limits and Methodologies for IMRT Measurement-Based Verification QA: Recommendations of AAPM Task Group No. 218. *Med Phys* (2018) 45:e53–83. doi: 10.1002/mp.12810
7. Nyflot MJ, Thammasorn P, Wootton LS, Ford EC, Chaovalitwongse WA. Deep Learning for Patient-Specific Quality Assurance: Identifying Errors in Radiotherapy Delivery by Radiomic Analysis of Gamma Images With Convolutional Neural Networks. *Med Phys* (2019) 46:456–64. doi: 10.1002/mp.13338
8. Van Esch A, Bohsung J, Sorvari P, Tenhunen M, Paiusco M, Iori M, et al. Acceptance Tests and Quality Control (QC) Procedures for the Clinical Implementation of Intensity Modulated Radiotherapy (IMRT) Using Inverse Planning and the Sliding Window Technique: Experience From Five Radiotherapy Departments. *Radiother Oncol* (2002) 65:53–70. doi: 10.1016/s0167-8140(02)00174-3
9. Valdes G, Scheuermann R, Hung CY, Olszanski A, Bellerive M, Solberg TD. A Mathematical Framework for Virtual IMRT QA Using Machine Learning. *Med Phys* (2016) 43:4323–34. doi: 10.1118/1.4953835
10. Valdes G, Chan MF, Lim SB, Scheuermann R, Deasy JO, Solberg TD. Imrt Qa Using Machine Learning: A Multi-Institutional Validation. *J Appl Clin Med Phys* (2017) 18:279–84. doi: 10.1002/acm2.12161
11. Dao L, Zhang XZ, Li H, Deshan Y, Schott B, Zhao T, et al. Predicting Gamma Passing Rates for Portal Dosimetry-Based IMRT QA Using Machine Learning. *Med Phys* (2019) 46:4666–75. doi: 10.1002/mp.13752
12. Li JQ, Wang L, Zhang X, Liu L, Li J, Chan MF, et al. Machine Learning for Patient-Specific Quality Assurance of VMAT: Prediction and Classification Accuracy. *Int J Radiat Oncol Biol Phys* (2019) 105:893–902. doi: 10.1016/j.ijrobp.2019.07.049

13. Interian Y, Rideout V, Kearney VP, Gennatas E, Morin O, Cheung J, et al. Deep Nets vs Expert Designed Features in Medical Physics: An IMRT QA Case Study. *Med Phys* (2018) 45:2672–80. doi: 10.1002/mp.12890
14. Tomori S, Kadoya N, Takayama Y, Kajikawa T, Shima K, Narazaki K, et al. A Deep Learning-Based Prediction Model for Gamma Evaluation in Patient-Specific Quality Assurance. *Med Phys* (2018) 45:4055–65. doi: 10.1002/mp.13112
15. Granville DA, Sutherland JG, Belec JG, Russa DJL. Predicting VMAT Patient-Specific QA Results Using a Support Vector Classifier Trained on Treatment Plan Characteristics and Linac QC Metrics. *Phys Med Biol* (2019) 64:09501. doi: 10.1088/1361-6560/ab142e
16. Chan MF, Witztum A, Valdes G. Integration of AI and Machine Learning in Radiotherapy Qa. *Front Artif Intell* (2020) 3:577620. doi: 10.3389/frai.2020.577620
17. Ono T, Hirashima H, Iramina H, Mukumoto N, Miyabe Y, Nakamura M, et al. Prediction of Dosimetric Accuracy for VMAT Plans Using Plan Complexity Parameters *Via* Machine Learning. *Med Phys* (2019) 46:3823–32. doi: 10.1002/mp.13669
18. Hirashima H, Ono T, Nakamura M, Miyabe Y, Mukumoto N, Iramina H, et al. Improvement of Prediction and Classification Performance for Gamma Passing Rate by Using Plan Complexity and Dosiomics Features. *Radiother Oncol* (2020) 153:250–7. doi: 10.1016/j.radonc.2020.07.031
19. Nelms BE, Chan MF, Jarry G, Lemire M, Lowden J, Hampton C, et al. Evaluating IMRT and VMAT Dose Accuracy: Practical Examples of Failure to Detect Systematic Errors When Applying a Commonly Used Metric and Action Levels. *Med Phys* (2013) 40:111722. doi: 10.1118/1.4826166
20. Heilemann G, Poppe B, Laub W. On the Sensitivity of Common Gamma-Index Evaluation Methods to MLC Misalignments in Rapidarc Quality Assurance. *Med Phys* (2013) 40:031702. doi: 10.1118/1.4789580
21. Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J. Unet++: Redesigning Skip Connections to Exploit Multiscale Features in Image Segmentation. *IEEE Trans Med Imaging* (2020) 39:1856–67. doi: 10.1109/TMI.2019.2959609
22. He K, Zhang X, Ren S, Sun J. *Deep Residual Learning for Image Recognition [C]//Ieee Conference on Computer Vision & Pattern Recognition*. IEEE Computer Society (2016).
23. Wall P, Fontenot JD. Application and Comparison of Machine Learning Models for Predicting Quality Assurance Outcomes in Radiation Therapy Treatment Planning. *Inf Med Unlocked* (2020) 18:100292. doi: 10.1016/j.imu.2020.100292
24. Wang L, Li J, Zhang S, Zhang X, Zhang Q, Chan MF, et al. Multi-Task Autoencoder Based Classification-Regression Model for Patient-Specific VMAT Qa. *Phys Med Biol* (2020) 65:235023. doi: 10.1088/1361-6560/abb31c
25. Kerns JR, Childress N, Kry SF. A Multi-Institution Evaluation of MLC Log Files and Performance in IMRT Delivery. *Radiat Oncol* (2014) 11:176. doi: 10.1186/1748-717X-9-176
26. Ford EC, Terezakis S, Souranis A, Harris K, Gay H, Mutic S. Quality Control Quantification (QCQ): A Tool to Measure the Value of Quality Control Checks in Radiation Oncology. *Int J Radiat Oncol Biol Phys* (2012) 84:e263–9. doi: 10.1016/j.ijrobp.2012.04.036
27. Passarge M, Fix MK, Manser P, Stampanoni MF, Siebers JV. A Swiss Cheese Error Detection Method for Real-Time EPID-Based Quality Assurance and Error Prevention. *Med Phys* (2017) 44:1212–23. doi: 10.1002/mp.12142
28. Fan J, Li J, Chen L, Stathakis S, Luo W, Du Plessis F, et al. A Practical Monte Carlo MU Verification Tool for IMRT Quality Assurance. *Phys Med Biol* (2006) 51:2503–15. doi: 10.1088/0031-9155/51/10/010

Conflict of Interest: KM was employed by Varian Medical Systems. JW was employed by Ping An Healthcare Technology Co Ltd.

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

frontiers in Oncology

ORIGINAL RESEARCH
published: 19 August 2021
doi: 10.3389/fonc.2021.702270

An Adversarial Deep-Learning-Based Model for Cervical Cancer CTV Segmentation With Multicenter Blinded Randomized Controlled Validation

Zhikai Liu[1†], Wanqi Chen[2†], Hui Guan[1], Hongnan Zhen[1], Jing Shen[1], Xia Liu[1], An Liu[3], Richard Li[3], Jianhao Geng[4], Jing You[4], Weihu Wang[4], Zhouyu Li[5], Yongfeng Zhang[6], Yuanyuan Chen[7], Junjie Du[8], Qi Chen[9], Yu Chen[9], Shaobin Wang[9], Fuquan Zhang[1] and Jie Qiu[1*]*

[1] Department of Radiation Oncology, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China, [2] Department of Nuclear Medicine, Sun Yat-Sen University Cancer Center, Guangzhou, China, [3] Department of Radiation Oncology, City of Hope National Medical Center, Duarte, CA, United States, [4] Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education/Beijing), Department of Radiation Oncology, Peking University Cancer Hospital and Institute, Beijing, China, [5] Department of Radiation Oncology, Affiliated Cancer Hospital & Institute of Guangzhou Medical University, Guangzhou, China, [6] Department of Radiation Oncology, The Fourth Hospital of Jilin University (FAW General Hospital), Jilin, China, [7] Oncology Department, Cangzhou Hospital of Integrated Traditional Chinese and Western Medicine, Hebei, China, [8] Department of Radiation Oncology, Yangquan First People's Hospital, Shanxi, China, [9] Research and Development Department, MedMind Technology Co., Ltd., Beijing, China

Purpose: To propose a novel deep-learning-based auto-segmentation model for CTV delineation in cervical cancer and to evaluate whether it can perform comparably well to manual delineation by a three-stage multicenter evaluation framework.

Methods: An adversarial deep-learning-based auto-segmentation model was trained and configured for cervical cancer CTV contouring using CT data from 237 patients. Then CT scans of additional 20 consecutive patients with locally advanced cervical cancer were collected to perform a three-stage multicenter randomized controlled evaluation involving nine oncologists from six medical centers. This evaluation system is a combination of objective performance metrics, radiation oncologist assessment, and finally the head-to-head Turing imitation test. Accuracy and effectiveness were evaluated step by step. The intra-observer consistency of each oncologist was also tested.

Results: In stage-1 evaluation, the mean DSC and the 95HD value of the proposed model were 0.88 and 3.46 mm, respectively. In stage-2, the oncologist grading evaluation showed the majority of AI contours were comparable to the GT contours. The average CTV scores for AI and GT were 2.68 *vs.* 2.71 in week 0 (P = .206), and 2.62 *vs.* 2.63 in week 2 (P = .552), with no significant statistical differences. In stage-3, the Turing imitation test showed that the percentage of AI contours, which were judged to be better than GT contours by ≥5 oncologists, was 60.0% in week 0 and 42.5% in week 2. Most oncologists demonstrated good consistency between the 2 weeks (P > 0.05).

OPEN ACCESS

Edited by:
Jose Eduardo Villarreal Barajas, Royal Devon and Exeter Hospital, United Kingdom

Reviewed by:
Kevin Martell, University of Calgary, Canada
Zhang Yi, Sichuan University, China
Leyla Moghaddasi, University of Melbourne, Australia

***Correspondence:**
Fuquan Zhang
zhangfuquan3@126.com
Jie Qiu
qiujie@pumch.cn

[†]These authors share first authorship

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

Received: *29 April 2021*
Accepted: *29 July 2021*
Published: *19 August 2021*

Citation:
Liu Z, Chen W, Guan H, Zhen H, Shen J, Liu X, Liu A, Li R, Geng J, You J, Wang W, Li Z, Zhang Y, Chen Y, Du J, Chen Q, Chen Y, Wang S, Zhang F and Qiu J (2021) An Adversarial Deep-Learning-Based Model for Cervical Cancer CTV Segmentation With Multicenter Blinded Randomized Controlled Validation. Front. Oncol. 11:702270. doi: 10.3389/fonc.2021.702270

Conclusions: The tested AI model was demonstrated to be accurate and comparable to the manual CTV segmentation in cervical cancer patients when assessed by our three-stage evaluation framework.

Keywords: deep-learning, auto-segmentation, evaluation, cervical cancer, radiotherapy, clinical target volume

INTRODUCTION

Cervical cancer (CC) remains one of the leading causes of cancer-related deaths in women worldwide (1). The majority of cervical cancer cases are diagnosed at the locally advanced stage in developing countries (2). External beam radiotherapy (EBRT) with concurrent chemotherapy followed by brachytherapy, also known as radical radiotherapy (RT), is the standard treatment for locally advanced cervical cancer (3) and has been shown to be effective in decreasing the risk of pelvic and vaginal vault recurrence (4).

Accurate and individualized clinical target volume (CTV) definition is vitally important for the definitive treatment of CC (5). During the past few years, a few high-performance deep-learning models based on convolutional neural networks (CNNs) have made tremendous progress and shown promise to serve as excellent assistance for target segmentation (6–12).

A recent study has first applied a deep-learning-based method called DpnUNet to CTV segmentation in cervical cancer. The authors' previous experimental results demonstrated that 88.65% of the contours generated by DpnUNet were acceptable for clinical usage (13). The mean dice similarity coefficient (DSC) and the 95th Hausdorff distance (95HD) were 0.86 and 5.34 for the delineated CTVs. However, there are still some glaring deficits. First, performance metrics such as mean DSC and 95HD are objective and offer good reproducibility (14–17), but do not incorporate physician's judgment and may not effectively evaluate for accuracy and applicability in a practical clinical context. Second, although the subjective oncologists' assessments showed that most predicted contours were acceptable for clinical usage when a head-to-head comparison was conducted between manual and AI-generated contours in the same CT slice, the DpnUNet model performed inferiorly. Therefore, it indicated that the currently proposed models did not perform exactly comparably well to manual delineations in clinical practice. Moreover, it seems that the current evaluation system for automatic segmentation models remains limited and insufficient.

Given the aforementioned reasons, a novel adversarial deep-learning-based auto-segmentation model is hence proposed for CTV delineation in cervical cancer. Then a challenging three-stage multicenter randomized controlled evaluation system is designed to directly validate the model and to minimize the inter- and intra-observer variability. This evaluation system is a combination of objective performance metrics, subjective radiation oncologist assessment, and finally, the Turing imitation test. Accuracy and effectiveness were evaluated step by step.

Abbreviations: CC, cervical cancer; RT, radiotherapy; EBRT, external beam radiotherapy; CTV, clinical target volume; ROIs, regions of interest; CNNs, convolutional neural networks; DPN, dual path network; GT, ground truth contouring; AI, artificial Intelligence; DICOM, digital imaging and communications in medicine; ESTRO, European Society for Radiotherapy and Oncology; RTOG, Radiation Therapy Oncology Group; DSC, Dice similarity coefficient; 95HD, the 95th Hausdorff distance; IMRT, intensity modulated radiation therapy.

MATERIALS AND METHODS

Network Architecture

CTVs are challenged to be evaluated with mathematical indicators due to fuzzy boundaries and large variations among different centers and observers. Inspired by a previously described work (18), an adversarial training approach based on the typical segmentation model is proposed to achieve similar performance between CTVs delineated by the proposed model and the oncologists. The overall architecture is shown in **Figure 1**.

The proposed model is based on DpnUNet (13), which originated from the architecture of U-Net (19), but replaces all the encoder and decoder components with DPN components. Considering that the original DpnUNet is still underperforming compared with manual delineation in clinical practice, an extra convolutional layer is added at the end of DpnUNet, in which the output channels are one and the kernel size is 1 × 1. A ResNet-10 with binary classification is used as the discriminator network (20). Since the discriminator is trained to identify the input segmentation generated by the model or delineated by oncologists, it will feedback the results to the model to promote similarities between the predicted CTVs and manual delineations.

The model was trained and tested using sets of CT data from 237 patients with locally advanced cervical cancer in our center with a GTX 1080GPU. All data using oral and IV contrast were constructed with a size of 512 × 512 pixels and acquired with a Brilliance CT Big Bore (Philips Healthcare, Best, Netherlands). The proposed model was trained over 50 circles to select the best model according to the lowest validation loss score.

Data Acquisition

To perform the three-stage evaluation, CT scans of a separate set of 20 new validation patients with locally advanced cervical cancer undergoing intensity-modulated radiation therapy (IMRT) were collected from November 2018 to December 2018 at the Peking Union Medical College Hospital. All patients were diagnosed with FIGO stage IB1–IIIC1 and/or node metastasis positive (N+) CC, treated with EBRT and radical RT. The average age ± standard deviation of these patients was 51.90 ± 12.63 years old.

CTV contours of 20 patients were redefined and re-delineated manually by radiation oncologists following the updated Radiation Therapy Oncology Group (RTOG) protocols

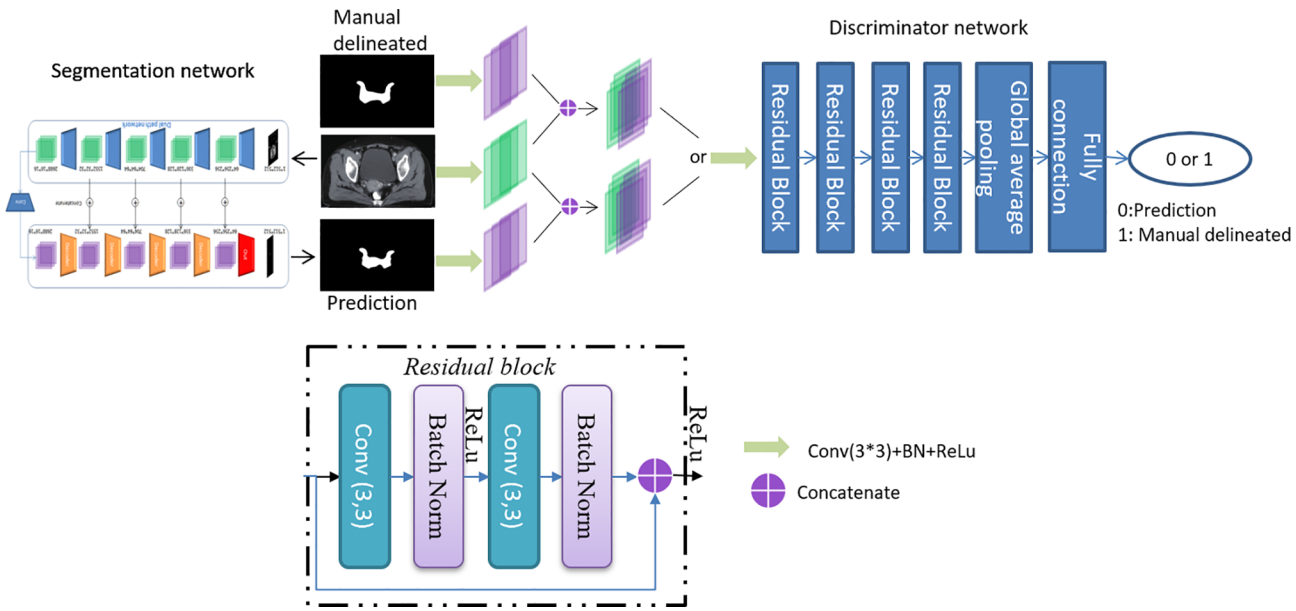


FIGURE 1 | The overall architecture of the proposed model.

(21–23). The CTV contours included the whole cervix, uterus, parametrium, vagina for 2 cm below GTV, and the elective nodal volume. All the contours were first reviewed by two senior radiation oncologists with more than 10 years of experience in radiotherapy specialized in cervical cancer at the Peking Union Medical College Hospital. To ensure the delineation quality of the human-generated CTV, the delineated contours were reviewed, modified, and approved collaboratively by a radiation oncologist committee consisting of eight senior oncologists at the Peking Union Medical College Hospital. The dataset of CT scans of 20 patients was used as a testing set of the proposed model to obtain artificial intelligence–generated contouring (AI) for performance assessment, of which 10 patients were randomly selected by Fisher-Yates shuffle for oncologist evaluation and the other 10 patients for the Turing-like test.

The Three-Stage Multicenter Randomized Controlled Evaluation

Stage 1: Performance Metrics

The flowchart of the three-level multicenter randomized controlled evaluation is shown in **Figure 2**. During the first-stage test, the Dice similarity coefficient (DSC) and the 95th percentile Hausdorff distance (95HD) were used to quantify the performance of the proposed model objectively.

The DSC was used to measure the spatial overlap between AI and GT contours, which is defined in Eq. (1).

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

Where *A* represents the volume of human-generated contour; *B* is the volume of an AI contour; and A∩B is the intersection volume that A and B have in common. The DSC value is between 0 and 1 (0 = no overlap, 1 = complete overlap).

The 95HD is defined as follows:

$$95\mathrm{HD}(\mathrm{A},\mathrm{B}) = \max\,(\mathrm{h}(\mathrm{A},\mathrm{B}),\ \mathrm{h}\,(\mathrm{B},\mathrm{A}),\ 95^{\mathrm{th}}) \tag{2}$$

$$\mathrm{HD}(\mathrm{A},\mathrm{B}) = \max(\mathrm{h}(\mathrm{A},\mathrm{B}),\mathrm{h}(\mathrm{B},\mathrm{A})) = \max(\mathrm{maxmin}||\mathrm{a}-\mathrm{b}||,\mathrm{maxmin}|\ |\mathrm{b}-\mathrm{a}||)\ \mathrm{a}\in\mathrm{Ab}\in\mathrm{B} \qquad \mathrm{b}\in\mathrm{Ba}\in\mathrm{A} \tag{3}$$

$||\bullet||$ means the Euclidean norm of the points of A and B. The HD in mm depicts the maximum mismatch between A and B. When the HD value decreases, the overlap between A and B increases. The mean and standard deviation were calculated.

Stage 2: Oncologist Evaluation

Ten cases from the testing set were randomly collected for oncologist evaluation. Twenty slices from each case were randomly extracted by Fisher-Yates shuffle, of which 10 slices were randomly selected to show GT contours, and the others were overlaid with AI contours. In total, 200 slices were obtained (AI: 10 × 10 = 100 slices *vs.* GT: 10 × 10 = 100 slices) and then randomly assigned to nine experienced radiation oncologists from six different cancer centers with more than 10 years of clinical experience in cervical cancer. The dataset of 200 randomized slices was evaluated by each oncologist slice by slice. The contours were graded in four scores: 3 points (No revision), 2 points (Minor revision), 1 point (Major revision), and 0 points (Rejection). The rubric is shown in **Table 1**.

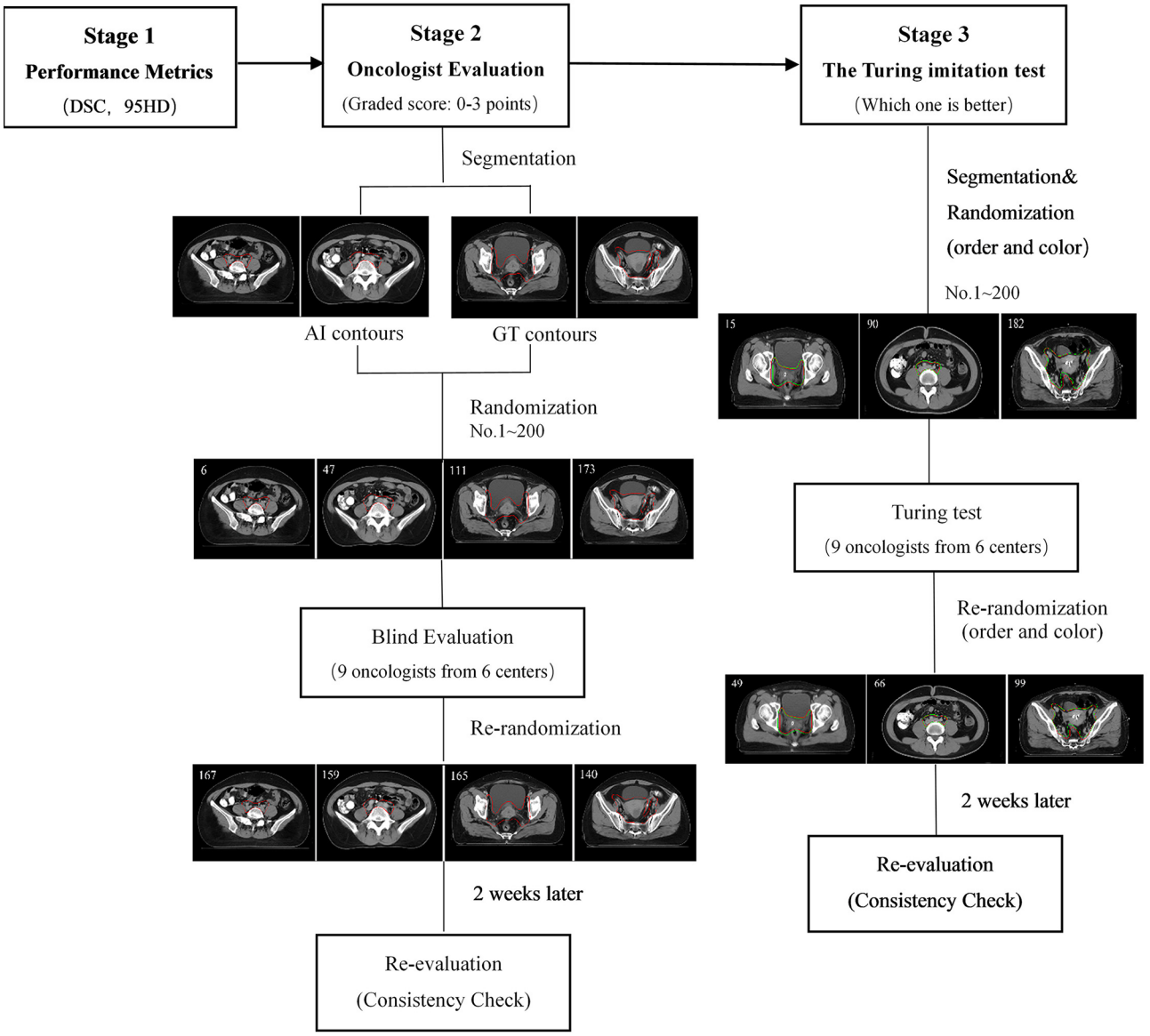


FIGURE 2 | The flowchart of the three-stage multicenter randomized controlled evaluation.

TABLE 1 | Criteria for the radiation oncologist evaluation.

Score	Grade	Criteria
3	No revision	The segmentation is perfect and completely acceptable for treatment.
2	Minor revision	The segmentation needs a few minor edits but has no significant clinical impact without correction.
1	Major revision	The segmentation needs significant revision. Treatment planning should not proceed without contour correction.
0	Rejection	The segmentation is unacceptable and needs to be redrawn.

The steps are outlined as follows:

1. **Data acquisition:** Twenty slices containing CTV from 10 patients' planning CT scans were randomly selected to generate a CT dataset consisting of 200 axial slices.
2. **Segmentation:** Both machine AI and human GT contours were generated for each dataset. Ten slices of each patient were randomly selected and overlaid with AI contours, while the other 10 slices were overlaid with GT contours. The contour color of the two groups was intentionally made the same for the blind test.

3. **Randomization:** The 200 CT slices were randomized by Fisher-Yates shuffle with an assigned unique ID so that the study authors could later distinguish whether each contour was an AI or GT.
4. **Blind evaluation:** The dataset of 200 randomized slices were distributed to the nine radiation oncologists. Each slice was scored from 0 to 3 blindly.
5. **Consistency evaluation:** After 2 weeks, the same dataset assigned in a new random order was distributed to the nine radiation oncologists for a second grading.
6. **Analysis:** The mean scores and the percentage of clinical acceptance of the AI and GT groups were calculated.

Stage 3: The Turing Imitation Test

The Turing imitation test is a subjective head-to-head comparison between GT- and AI-generated contours. In this test, the participant was presented with two contours overlaid simultaneously in the same CT slice, one of which was generated by the AI. The radiation oncologist was requested to choose which contour was better for clinical application. The steps are outlined as follows:

1. **Data acquisition:** We randomly extracted 20 CTV containing axial CT slices from each of the remaining 10 test patients to generate a 200-slice dataset.
2. **Segmentation:** For each slice, the AI and GT contours of CTV were generated randomly in a different color (red or green). The structure colors were randomized on a per-slice basis so as not to bias the Turing imitation test.
3. **Randomization:** AI- and GT-generated CTV slices were randomized by Fisher-Yates shuffle and anonymized to facilitate the blind evaluation. Each slice was assigned a unique ID so images could be de-anonymized later to analyze.
4. **Turing test:** The dataset was distributed to the test team, consisting of nine radiation oncologists from six different centers. Each radiation oncologist was requested to compare the AI and GT delineations and select the one that was more suitable for clinical application. The evaluation time for each slice was limited to 30 s to prevent the observer from seeking additional visual clues regarding the source of the contour.
5. **Consistency evaluation:** After 2 weeks, the same dataset assigned in a new random order and color was distributed to the radiation oncologists for a new comparison.
6. **Analysis:** If the AI contours received a better evaluation, the result would be considered positive. The positive rates of the entire test set and of each oncologist were calculated. Following the original Turing proposal (24), the threshold of the overall positive result rate was set to 30%. Above that, the AI model is considered to have passed the Turing imitation test.

Statistical Analysis

The mean and standard deviation of DSC and 95HD were calculated. The Wilcoxon matched-pairs signed-rank test was used to compare the AI and GT contours in the oncologist evaluation and the Turing imitation test. The score difference between AI and GT contours evaluated by each oncologist was performed by Mann-Whitney U test. The Wilcoxon paired signed-rank test was used to compare the agreement of the oncologist evaluation between 2 weeks for each oncologist. The McNemar test was used to compare the consistency of the Turing test between 2 weeks. Statistical significance was set at two-tailed $P < .05$.

RESULTS

Stage 1: Quantitative Performance Metrics

All slices of the 20 testing patients were evaluated with the quantitative performance metrics, which is shown and compared with DpnUNet in **Table 2**. The DSC and 95HD values of the proposed model were 0.88 ± 0.03 and 3.46 ± 1.88 mm, respectively.

Stage 2: Oncologist Evaluation

Table 3 shows oncologist evaluation results of CTV contours. Score ≥2 was defined as suitable for clinical application. Using these scoring criteria for contour evaluation, most CTV contours were clinically acceptable by all the oncologists. For AI contours, the percentage of clinically acceptable scores was 97.4%, compared to the 98.3% of GT contours. We also compared AI and GT scores with a separate Mann-Whitney test for each oncologist and found that there was no significant difference between the week 0 timepoint and the after-2-weeks timepoint. **Figure 3** shows the CTV scores for AI and GT contours. The overall average scores for AI and GT were 2.68 *vs.* 2.71 in week 0 ($P = .206$) and 2.62 *vs.* 2.63 in week 2 ($P = .552$), respectively. The intra-observer consistency analyses between 2 weeks were performed by the Wilcoxon paired signed-rank test. It was found that the consistency of two oncologists was poor, while the others had good consistency between 2 weeks ($P > .05$).

Stage 3: The Turing Imitation Test

When considering physician selection of the AI contour as preferred over the GT contour as a positive result, the overall positive rate in week 0 was 54.17% compared with 45.83% negative rate ($P = .139$), while in week 2 the positive rate was 54% *vs.* the negative rate of 46% ($P = .128$), which demonstrated the proposed deep machine learning model performed equally well or even better than human delineation. Furthermore, the consistency evaluation was performed by repeating the same dataset in different random order and colors to the test team after 2 weeks. The results are shown in **Table 4**. Subclass analysis was performed to evaluate individual oncologists and CT slices. The results showed that six slices (3.0% in week 0) of AI contours were scored to be better than GT by all the oncologists. The percentage of AI contours that were approved to be better by ≥5 oncologists was 60.0% in week 0 and 42.5% in week 2. The distribution map is shown in **Figure 4**. Sample CTV delineations are presented in **Figure 5**.

TABLE 2 | The comparison of DSC and 95HD value of our proposed model and DpnUNet.

Test	Patient (No.)	Proposed Model		DpnUNet	
		DSC	95HD (mm)	DSC	95HD (mm)
Stage 2 patient cohort :	1	0.9	1.95	0.84	2.09
Oncologist Evaluation	2	0.91	2.34	0.84	2.38
	3	0.9	3.68	0.89	3.61
	4	0.9	1.95	0.90	1.85
	5	0.83	7.68	0.75	8,84
	6	0.88	2.98	0.81	3.10
	7	0.84	7.07	0.80	8.10
	8	0.9	2.55	0.93	2.45
	9	0.89	2.83	0.83	3.85
	10	0.88	3.35	0.86	3.41
Stage 3 patient cohort :	11	0.85	5.1	0.75	6.17
The Turing Test	12	0.91	2.83	0.90	3.48
	13	0.81	7.76	0.84	7.92
	14	0.91	2.24	0.89	2.33
	15	0.91	2.21	0.94	1.97
	16	0.89	2.24	0.87	2.06
	17	0.9	2.83	0.82	2.49
	18	0.89	2.45	0.92	2.88
	19	0.93	2.25	0.94	2.26
	20	0.85	2.93	0.84	2.25
	Mean ± STD	0.88 ± 0.03	3.46 ± 1.88	0.86 ± 0.06	3.67 ± 2.22

TABLE 3 | Graded oncologist evaluation for AI and GT contours.

	Week 0																	
Oncologist	**A**		**B**		**C**		**D**		**E**		**F**		**G**		**H**		**I**	
Score	**AI**	**GT**	**AI**	**GT**	**AI**	**GT**	**AI**	**GT**	**AI**	**GT**	**AI**	**GT**	**AI**	**GT**	**AI**	**GT**	**AI**	**GT**
3	89%	97%	93%	95%	30%	37%	71%	74%	54%	57%	75%	85%	94%	94%	82%	77%	45%	37%
2	11%	3%	7%	5%	61%	56%	28%	25%	46%	42%	25%	15%	6%	6%	18%	23%	45%	57%
1	0	0	0	0	9%	6%	1%	1%	0	1%	0	0	0	0	0	0	10%	6%
0	0	0	0	0	0	1%	0	0	0	0	0	0	0	0	0	0	0	0%
Mean Score	2.89	2.96	2.93	2.95	2.21	2.29	2.70	2.73	2.54	2.56	2.75	2.85	2.94	2.94	2.82	2.77	2.35	2.31
P value	0.061		0.553		0.282		0.640		0.719		0.078		1.000		0.382		0.494	
	Week 2																	
3	93%	92%	88%	93%	29%	37%	78%	77%	42%	50%	78%	69%	94%	96%	33%	33%	50%	42&
2	7%	8%	12%	7%	63%	54%	22%	21%	57%	50%	22%	31%	5%	4%	62%	64%	39%	56%
1	0	0	0	0	8%	9%	0	1%	1%	0	0	0	1%	0	5%	3%	7%	2%
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4%	0
Mean Score	2.93	2.92	2.88	2.93	2.21	2.28	2.78	2.74	2.41	2.5	2.78	2.69	2.93	2.96	2.28	2.3	2.35	2.40
***P* value**	0.789		0.229		0.352		0.940		0.230		0.150		0.509		0.846		0.728	
Consistency (*P* value)		0.782		0.108		0.907		0.064		**0.007**		0.118		0.491		**0.000**		0.170

P < 0.05, the results are statistically significant.

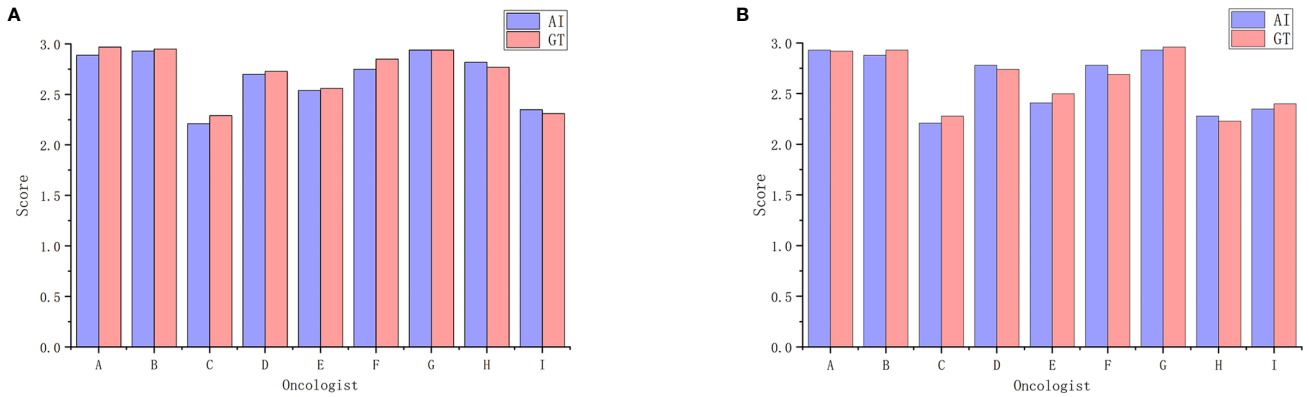


FIGURE 3 | Average scores for AI and GT by the nine oncologists. **(A)** Week 0. **(B)** Week 2.

TABLE 4 | The results of the Turing-like imitation test.

Oncologist	Week 0		Week 2		Consistency (*P* value)
	Positive	Negative	Positive	Negative	
A	130 (65%)	70 (35%)	137 (68.5%)	63 (31.5%)	0.296
B	92 (46%)	108 (54%)	100 (50%)	100 (50%)	0.461
C	106 (53%)	94 (47%)	116 (58%)	84 (42%)	0.134
D	107 (53.5%)	93 (46.5%)	100 (50%)	100 (50%)	0.510
E	98 (49%)	102 (51%)	114 (57%)	86 (43%)	**0.034**
F	111 (55.5%)	89 (44.5%)	102 (51%)	98 (49%)	0.508
G	122 (61%)	78 (39%)	117 (58.5%)	83 (41.5%)	0.712
H	119 (59.5%)	81 (40.5%)	95 (47.5%)	105 (52.5%)	0.101
I	90 (45%)	110 (55%)	89 (44.5%)	111(55.5%)	0.815
***P* value**	0.139		0.128		

P < 0.05, results are statistically significant.

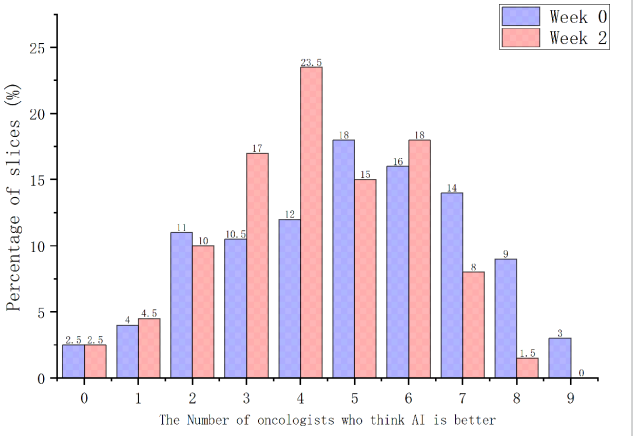


FIGURE 4 | The distribution map of the positive results.

DISCUSSION

Segmentation of CTV is an essential step for successful radiotherapy delivery (16). However, manual delineation is time-consuming and subjective, with considerable inter- and intra-observer variability (25–28). Therefore, accurate and consistent automated segmentation methods are highly desirable and useful for pretreatment radiotherapy planning. Automatic segmentation techniques especially based on CNN models have made significant progress with increasing reliability and accuracy in recent years, thus potentially relieving radiation oncologists from the time-cost of contouring. To the authors' knowledge, very few studies were reported on the automatic delineation of the CTV (29–32) due to the ambiguous and blurred boundaries between the CTV and normal tissues, the potential for tumor spread or subclinical diseases in the CT images, and the inter-observer variability in recognition of anatomical structures. The current most common approach to evaluate automatic delineation of the CTV is to compare with GT contours using quantitative measures such as DSC and HD (33, 34). However, this mathematical evaluation is basic and depends only on the geometrical properties of the organ being delineated. This approach does not incorporate clinical judgment and may not adequately extract the main characteristics and the core elements of the image.

Given the clinical application, an authors' previous study added subjective oncologist evaluation to the proposed model,

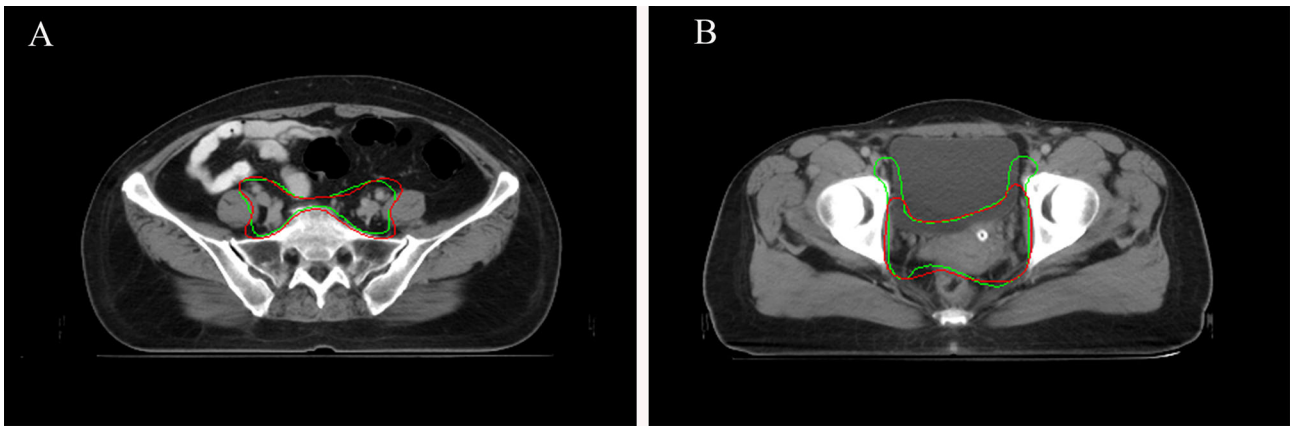


FIGURE 5 | **(A)** Sample CTV where the AI contour was approved by all the oncologists. AI contours in green line. GT contours in red line. **(B)** Sample CTV where the GT contour was approved by all the oncologists. AI contours in green line. GT contours in red line.

and the result showed that more than 88% of the slices predicted from DpnUNet were evaluated as "No revision" or "Minor revision" (13). However, when radiation oncologists were presented with AI and GT contours overlaid simultaneously in the same CT slice, the GT contour was always the one chosen to be better. Therefore, a novel auto-segmentation model that indeed performs comparably well or even better to manual delineation for CTV delineation is desirable. Moreover, the current evaluating performance of segmentation, particularly CTV segmentation, can be challenging due to the large variations among different centers and observers (35, 36). Therefore, a three-stage randomized controlled evaluation framework was proposed, combining the three elements of traditional performance metrics, oncologist evaluation, and the Turing imitation test, for a comprehensive assessment of the proposed model in cervical cancer CTV segmentation.

During stage-1 evaluation, the mean DSC value of CTV of the proposed model was 0.88, which was higher compared with the acceptable threshold of 0.80 to 0.86 used in other studies (13, 37–39). The average 95HD value was 3.46 mm compared to 5.34 mm by the DpnUNet model (13). The results indicated a strong concordance between the proposed automatic model and human experts for CTV contouring.

In stage-2 evaluation, a multicenter randomized controlled evaluation involving nine radiation oncologists from six different centers was designed to examine the model's clinical utility and generalization. The anonymized CT slices were randomly distributed with AI or GT contours to experienced radiation oncologists for assessment. The choice of a random design instead of using entire connected slices is mainly because AI sometimes has obvious characteristics at certain levels, especially at the beginning and the end, which do not affect the accuracy of target delineation but make it more easy to be distinguished. Moreover, the evaluation is more clinically relevant and minimizes assessment bias as oncologists are blinded to the source of the contours. The results showed that our proposed model was highly acceptable for clinical application and treatment. There was no significant difference in physician acceptability ratings between scores of AI and GT contours, which means our model can provide consistent segmentation and performed well with good agreement to the manual contours. However, there were still 2.6% of cases where the AI contours were judged by some oncologists to require major revision. We retrospectively analyzed these outlier cases and found that most of them were in the middle level of the pelvic cavity; thus, the ROIs had very unclear boundaries and massive diversity of sizes, shapes with low contrast to the rectum, bladder, and small intestines. The circumstances mentioned above limit the generalizability of the AI model, and therefore more caution is warranted.

In 1950, Alan Turing proposed an influential test for how to evaluate artificial intelligence: an imitation is successful when we cannot distinguish it from the real thing (24). Here, this analogous logic was applied to the artificial segmentation technology, and a similar Turing imitation test was proposed. The variant of the Turing imitation test used in this study is a randomized blinded evaluation. In contrast with the stage-2 task, in which evaluators viewed individual stimuli and made categorical judgments, the radiation oncologists were presented with AI and GT contour masks on the same slice and were requested to choose which was better. If the positive rate of AI is more than 30%, then the AI model was considered to have passed the test. It is a straightforward head-to-head comparison, which compares two contours in the exact same condition to minimize the interference factors such as scanning conditions, anatomical variations, and severity of disease in different patients. As shown in **Table 4**, the segmentation model passed the Turing test with overall positive rates much higher than 30%. The overall positive rate was 54.17% in week 0 and 54% in week 2, which demonstrated that the AI segmentation model performed equally well as humans (P = .139, P = .128). Moreover, correlations were observed between the objective and subjective

measures. Those with lower DSC and 95HD values were also more likely to be flagged as requiring revision or inferior performance during the subjective evaluation.

Subjective assessment still has drawbacks. Oncologists involved in this study stated that they might change their opinion of the grading score if they viewed it at a later point, and may not be able to definitively decide between two contours if they showed a high degree of overlap. Therefore, the intra-observer consistency analyses between 2 weeks were performed during stage-2 and stage-3 evaluation. Most oncologists were found to maintain good consistency between 2 weeks without significant difference. Considering that good scores or positive rates could have resulted from a range of factors affecting how the contours were evaluated, a distribution map across all images involved in the Turing imitation test was additionally generated, to evaluate the number of oncologists who consistently thought AI contours were better. The results showed that the percentage of AI contours to be better than GT by ≥5 oncologists was 60.0% in week 0 and 42.5% in week 2, which further demonstrated the excellent performance of the proposed segmentation model.

CONCLUSION

In this study, a novel deep-learning-based CNN model for fully automatic and accurate CTV segmentation in cervical cancer was proposed. Then a comprehensive three-stage randomized controlled evaluation framework was performed to validate the model. This evaluation system is a combination of objective and subjective evaluation and can diminish the risk of bias and enhance real-world clinical relevance compared to the most commonly used evaluation method of applying performance metrics alone. The tested AI model was demonstrated to be accurate and comparable to the manual CTV segmentation in cervical cancer patients. Furthermore, this study provided guidelines for each step, which can be referred to by other centers according to their sample size limitation. While this study focuses only on cervical cancer, the methodology and general learnings may translate to other tumor sites. Moreover, this comprehensive assessment of contouring performance may also be referenced as a base framework for evaluating the clinical utility of automatic segmentation methods in the future.

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 evaluation was reviewed and approved by the Peking Union Medical College Hospital Institutional Review board. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

FZ and JQ had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Concept and design: FZ, JQ, ZKL, WC. Acquisition, analysis, or interpretation of data: HG, HZ, JS, XL, JG, JY, WW, ZYL, YZ, YYC, JD. AI model design: SW, QC, YC. Drafting of the manuscript: WC, ZKL. Critical revision of the manuscript for important intellectual content: AL, RL. All authors contributed to the article and approved the submitted version.

FUNDING

This work was funded by the following grants from the Non-profit Central Research Institute Fund of Chinese Academy of Medical Sciences (grant number 2019XK320014). FZ takes responsibility for the integrity of the data and the accuracy of the data analysis.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2021.702270/full#supplementary-material

REFERENCES

1. Koh WJ, Abu-Rustum NR, Bean S, Bradley K, Campos SM, Cho KR, et al. Cervical Cancer, Version 3.2019, NCCN Clinical Practice Guidelines in Oncology. *J Natl Compr Cancer Network* (2019) 17:64–84. doi: 10.6004/jnccn.2019.0001
2. Chibwesha CJ, Stringer JSA. Cervical Cancer as a Global Concern: Contributions of the Dual Epidemics of HPV and HIV. *Jama* (2019) 322:1558–60. doi: 10.1001/jama.2019.16176
3. Peters WA, Liu PY, Barrett RJ, Stock RJ, Monk BJ, Berek JS, et al. Concurrent Chemotherapy and Pelvic Radiation Therapy Compared With Pelvic Radiation Therapy Alone as Adjuvant Therapy After Radical Surgery in High-Risk Early-Stage Cancer of the Cervix. *J Clin Oncol* (2000) 18:1606–13. doi: 10.1200/JCO.2000.18.8.1606
4. Han K, Milosevic M, Fyles A, Pintilie M, Viswanathan AN. Trends in the Utilization of Brachytherapy in Cervical Cancer in the United States. *Int J Radiat Oncol Biol Phys* (2013) 87:111–9. doi: 10.1016/j.ijrobp.2013.05.033
5. Monk BJ, Tewari KS, Koh WJ. Multimodality Therapy for Locally Advanced Cervical Carcinoma: State of the Art and Future Directions. *J Clin Oncol* (2007) 25:2952–65. doi: 10.1200/JCO.2007.10.8324
6. Sarıgül M, Ozyildirim BM, Avci M. Differential Convolutional Neural Network. *Neural Networks* (2019) 116:279–87. doi: 10.1016/j.neunet.2019.04.025
7. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, et al. A Guide to Deep Learning in Healthcare. *Nat Med* (2019) 25:24–9. doi: 10.1038/s41591-018-0316-z
8. 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) 291:677–86. doi: 10.1148/radiol.2019182012
9. Lustberg T, van Soest J, Gooding M, Peressutti D, Aljabar P, van der Stoep J, et al. Clinical Evaluation of Atlas and Deep Learning Based Automatic Contouring for Lung Cancer. *Radiother Oncol* (2018) 126:312–7. doi: 10.1016/j.radonc.2017.11.012
10. Ahn SH, Yeo AU, Kim KH, Kim C, Goh Y, Cho S, et al. Comparative Clinical Evaluation of Atlas and Deep-Learning-Based Auto-Segmentation of Organ Structures in Liver Cancer. *Radiat Oncol* (2019) 14:213. doi: 10.1186/s13014-019-1392-z
11. Ayyalusamy A, Vellaiyan S, Subramanian S, Ilamurugu A, Satpathy S, Nauman M, et al. Auto-Segmentation of Head and Neck Organs at Risk in Radiotherapy and Its Dependence on Anatomic Similarity. *Radiat Oncol J* (2019) 37:134–42. doi: 10.3857/roj.2019.00038
12. Kim N, Chang JS, Kim YB, Kim JS. Atlas-Based Auto-Segmentation for Postoperative Radiotherapy Planning in Endometrial and Cervical Cancers. *Radiat Oncol* (2020) 15:106. doi: 10.1186/s13014-020-01562-y
13. Liu Z, Liu X, Guan H, Zhen H, Sun Y, Chen Q, et al. Development and Validation of a Deep Learning Algorithm for Auto-Delineation of Clinical Target Volume and Organs at Risk in Cervical Cancer Radiotherapy. *Radiother Oncol* (2020) 153:172–9. doi: 10.1016/j.radonc.2020.09.060
14. van Rooij W, Dahele M, Ribeiro Brandao H, Delaney AR, Slotman BJ, Verbakel WF. Deep Learning-Based Delineation of Head and Neck Organs at Risk: Geometric and Dosimetric Evaluation. *Int J Radiat Oncol Biol Phys* (2019) 104:677–84. doi: 10.1016/j.ijrobp.2019.02.040
15. van der Veen J, Willems S, Deschuymer S, Robben D, Crijns W, Maes F, et al. Benefits of Deep Learning for Delineation of Organs at Risk in Head and Neck Cancer. *Radiother Oncol* (2019) 138:68–74. doi: 10.1016/j.radonc.2019.05.010
16. Men K, Zhang T, Chen X, Chen B, Tang Y, Wang S, et al. Fully Automatic and Robust Segmentation of the Clinical Target Volume for Radiotherapy of Breast Cancer Using Big Data and Deep Learning. *Phys Med* (2018) 50:13–9. doi: 10.1016/j.ejmp.2018.05.006
17. Men K, Chen X, Zhang Y, Zhang T, Dai J, Yi J, et al. Deep Deconvolutional Neural Network for Target Segmentation of Nasopharyngeal Cancer in Planning Computed Tomography Images. *Front Oncol* (2017) 7:315. doi: 10.3389/fonc.2017.00315
18. Moeskops P, Veta M, Lafarge MW, Eppenhof KAJ, Pluim JPW. Adversarial Training and Dilated Convolutions for Brain MRI Segmentation In: *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Springer: Cham. (2017) p. 56–64. doi: 10.1007/978-3-319-6755897
19. Norman B, Pedoia V, Majumdar S. Use of 2D U-Net Convolutional Neural Networks for Automated Cartilage and Meniscus Segmentation of Knee MR Imaging Data to Determine Relaxometry and Morphometry. *Radiology* (2018) 288:177–85. doi: 10.1148/radiol.2018172322
20. Zhou LQ, Wu XL, Huang SY, Wu GG, Ye HR, Wei Q, et al. Lymph Node Metastasis Prediction From Primary Breast Cancer US Images Using Deep Learning. *Radiology* (2020) 294:19–28. doi: 10.1148/radiol.2019190372
21. Harris VA, Staffurth J, Naismith O, Esmail A, Gulliford S, Khoo V, et al. Consensus Guidelines and Contouring Atlas for Pelvic Node Delineation in Prostate and Pelvic Node Intensity Modulated Radiation Therapy. *Int J Radiat Oncol Biol Phys* (2015) 92:874–83. doi: 10.1016/j.ijrobp.2015.03.021
22. Lim K, Small W Jr, Portelance L, Creutzberg C, Jürgenliemk-Schulz IM, Mundt A, et al. Consensus Guidelines for Delineation of Clinical Target Volume for Intensity-Modulated Pelvic Radiotherapy for the Definitive Treatment of Cervix Cancer. *Int J Radiat Oncol Biol Phys* (2011) 79:348–55. doi: 10.1016/j.ijrobp.2009.10.075
23. Small WJr., Bosch WR, Harkenrider MM, Strauss JB, Abu-Rustum N, Albuquerque KV, et al. NRG Oncology/RTOG Consensus Guidelines for Delineation of Clinical Target Volume for Intensity Modulated Pelvic Radiation Therapy in Postoperative Treatment of Endometrial and Cervical Cancer: an Update. *Int J Radiat Oncol Biol Phys* (2021) 109:413–24. doi: 10.1016/j.ijrobp.2020.08.061
24. Turing AM. Computing Machinery and Intelligence. *Mind* (1950) 433–60. doi: 10.1093/mind/LIX.236.433
25. Keenan LG, Rock K, Azmi A, Salib O, Gillham C, McArdle O. An Atlas to Aid Delineation of Para-Aortic Lymph Node Region in Cervical Cancer: Design and Validation of Contouring Guidelines. *Radiother Oncol* (2018) 127:417–22. doi: 10.1016/j.radonc.2018.02.013
26. Small WJr., Mell LK, Anderson P, Creutzberg C, De Los Santos J, Gaffney D, et al. Consensus Guidelines for Delineation of Clinical Target Volume for Intensity-Modulated Pelvic Radiotherapy in Postoperative Treatment of Endometrial and Cervical Cancer. *Int J Radiat Oncol Biol Phys* (2008) 71:428–34. doi: 10.1016/j.ijrobp.2007.09.042
27. Apolle R, Appold S, Bijl HP, Blanchard P, Bussink J, Faivre-Finn C, et al. Inter-Observer Variability in Target Delineation Increases During Adaptive Treatment of Head-and-Neck and Lung Cancer. *Acta Oncol* (2019) 58:1378–85. doi: 10.1080/0284186X.2019.1629017
28. Bø HK, Solheim O, Jakola AS, Kvistad KA, Reinertsen I, Berntsen EM. Intra-Rater Variability in Low-Grade Glioma Segmentation. *J Neuro-oncol* (2017) 131:393–402. doi: 10.1007/s11060-016-2312-9
29. Larsson R, Xiong JF, Song Y, Ling-Fu, Chen YZ, Xiaowei X, et al. Automatic Delineation of the Clinical Target Volume in Rectal Cancer for Radiation Therapy Using Three-Dimensional Fully Convolutional Neural Networks. *Annu Int Conf IEEE Eng Med Biol Soc* (2018) 2018:5898–901. doi: 10.1109/EMBC.2018.8513506
30. Song Y, Hu J, Wu Q, Xu F, Nie S, Zhao Y, et al. Automatic Delineation of the Clinical Target Volume and Organs at Risk by Deep Learning for Rectal Cancer Postoperative Radiotherapy. *Radiother Oncol* (2020) 145:186–92. doi: 10.1016/j.radonc.2020.01.020
31. Shusharina N, Söderberg J, Edmunds D, Löfman F, Shih H, Bortfeld T. Automated Delineation of the Clinical Target Volume Using Anatomically Constrained 3D Expansion of the Gross Tumor Volume. *Radiother Oncol* (2020) 146:37–43. doi: 10.1016/j.radonc.2020.01.028
32. Liu Z, Liu X, Xiao B, Wang S, Miao Z, Sun Y, et al. Segmentation of Organs-at-Risk in Cervical Cancer CT Images With a Convolutional Neural Network. *Phys Med* (2020) 69:184–91. doi: 10.1016/j.ejmp.2019.12.008
33. 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
34. van Dijk LV, Van den Bosch L, Aljabar P, Peressutti D, Both S, Steenbakkers RJHM, 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
35. Louie AV, Rodrigues G, Olsthoorn J, Palma D, Yu E, Yaremko B, et al. Inter-Observer and Intra-Observer Reliability for Lung Cancer Target Volume Delineation in the 4D-CT Era. *Radiother Oncol* (2010) 95:166–71. doi: 10.1016/j.radonc.2009.12.028
36. Federico M, Hernandez-Socorro CR, Ribeiro I, Martin JG, Oramas MDR, Saez-Bravo ML, et al. Prospective Intra/Inter-Observer Evaluation of Pre-Brachytherapy Cervical Cancer Tumor Width Measured in TRUS and MR Imaging. *Radiat Oncol* (2019) 14:173. doi: 10.1186/s13014-019-1352-7
37. Kurata Y, Nishio M, Kido A, Fujimoto K, Yakami M, Isoda H, et al. Automatic Segmentation of the Uterus on MRI Using a Convolutional Neural Network. *Comput Biol Med* (2019) 114:103438. doi: 10.1016/j.compbiomed.2019.103438
38. Chen L, Shen C, Zhou Z, Maquilan G, Albuquerque K, Folkert MR, et al. Automatic PET Cervical Tumor Segmentation by Combining Deep Learning and Anatomic Prior. *Phys Med Biol* (2019) 64:085019. doi: 10.1088/1361-6560/ab0b64
39. Ghose S, Holloway L, Lim K, Chan P, Veera J, Vinod SK, et al. A Review of Segmentation and Deformable Registration Methods Applied to Adaptive Cervical Cancer Radiation Therapy Treatment Planning. *Artif Intell Med* (2015) 64:75–87. doi: 10.1016/j.artmed.2015.04.006

Conflict of Interest: Authors SW, QC, and YC were employed by the company MedMind Technology Co.

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.

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: 24 August 2021
doi: 10.3389/fonc.2021.712423

Replacing Manual Planning of Whole Breast Irradiation With Knowledge-Based Automatic Optimization by Virtual Tangential-Fields Arc Therapy

Roberta Castriconi[1†], Pier Giorgio Esposito[1†], Alessia Tudda[1], Paola Mangili[1], Sara Broggi[1], Andrei Fodor[2], Chiara L. Deantoni[2], Barbara Longobardi[1], Marcella Pasetti[2], Lucia Perna[1], Antonella del Vecchio[1], Nadia Gisella Di Muzio[2] and Claudio Fiorino[1]*

[1] Medical Physics, San Raffaele Scientific Institute, Milano, Italy, [2] Radiotherapy, San Raffaele Scientific Institute, Milano, Italy

OPEN ACCESS

Edited by:
Savino Cilla,
Gemalli Molise Hospital, Italy

Reviewed by:
Raquel Bar-Deroma,
Rambam Health Care Campus, Israel
Tsair-Fwu Lee,
National Kaohsiung University of Science and Technology, Taiwan

***Correspondence:**
Claudio Fiorino
fiorino.claudio@hsr.it

[†]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: 20 May 2021
Accepted: 02 August 2021
Published: 24 August 2021

Citation:
Castriconi R, Esposito PG, Tudda A, Mangili P, Broggi S, Fodor A, Deantoni CL, Longobardi B, Pasetti M, Perna L, del Vecchio A, Di Muzio NG and Fiorino C (2021) Replacing Manual Planning of Whole Breast Irradiation With Knowledge-Based Automatic Optimization by Virtual Tangential-Fields Arc Therapy. Front. Oncol. 11:712423. doi: 10.3389/fonc.2021.712423

Purpose: To implement Knowledge Based (KB) automatic planning for right and left-sided whole breast treatment through a new volumetric technique (ViTAT, Virtual Tangential-fields Arc Therapy) mimicking conventional tangential fields (TF) irradiation.

Materials and Method: A total of 193 clinical plans delivering TF with wedged or field-in-field beams were selected to train two KB-models for right(R) and left(L) sided breast cancer patients using the RapidPlan (RP) tool implemented in the Varian Eclipse system. Then, a template for ViTAT optimization, incorporating individual KB-optimized constraints, was interactively fine-tuned. ViTAT plans consisted of four arcs (6 MV) with start/stop angles consistent with the TF geometry variability within our population; the delivery was completely blocked along the arcs, apart from the first and last 20° of rotation for each arc. Optimized fine-tuned KB templates for automatic plan optimization were generated. Validation tests were performed on 60 new patients equally divided in R and L breast treatment: KB automatic ViTAT-plans (KB-ViTAT) were compared against the original TF plans in terms of OARs/PTVs dose-volume parameters. Wilcoxon-tests were used to assess the statistically significant differences.

Results: KB models were successfully generated for both L and R sides. Overall, 1(3%) and 7(23%) out of 30 automatic KB-ViTAT plans were unacceptable compared to TF for R and L side, respectively. After the manual refinement of the start/stop angles, KB-ViTAT plans well fitted TF-performances for these patients as well. PTV coverage was comparable, while PTV $D_{1\%}$ was improved with KB-ViTAT by R:0.4/L:0.2 Gy ($p < 0.05$); ipsilateral OARs D_{mean} were similar with a slight (i.e., few % volume) improvement/worsening in the 15–35 Gy/2–15 Gy range, respectively. KB-ViTAT better spared contralateral OARs: D_{mean} of contralateral OARs was 0.1 Gy lower ($p < 0.05$); integral dose was R:5%/L:8% lower ($p < 0.05$) than TF. The overall time for the automatic plan optimization and final dose calculation was 12 ± 2 minutes.

Conclusions: Fully automatic KB-optimization of ViTAT can efficiently replace manually optimized TF planning for whole breast irradiation. This approach was clinically

implemented in our institute and may be suggested as a large-scale strategy for efficiently replacing manual planning with large sparing of time, elimination of inter-planner variability and of, seldomly occurring, sub-optimal manual plans.

Keywords: breast cancer, radiation oncology, automation, plan optimization, tangential field, knowledge-based

INTRODUCTION

Post-operative irradiation of the whole breast is a well assessed and effective therapeutic option in the treatment of localized breast cancer (1). Typically, more than 70% of women submitted to breast-conserving surgery receives post-operative radiotherapy and a large fraction of them is treated to sterilize the whole breast, typically delivering 38–40 Gy in 15–16 fractions or 50 Gy in 25 fractions. Due to the large and increasing incidence of breast cancer in the female population (2), this treatment represents a quite relevant fraction of the patients daily treated in the radiation oncology departments worldwide. Despite the evolution toward more personalized approaches including the possibility to reduce the treated volumes (i.e., partial breast irradiation) or to deliver higher dose to the tumor bed or to include selected nodal regions at risk, whole breast irradiation still maintains a central role in the treatment of breast cancer. Nowadays, different techniques are used; most Institutes still prefer the conventional tangential fields (TF) arrangement, either using 3-dimensional conformal radiation therapy (3DCRT) with wedges (physical or dynamic) or intensity modulated radiotherapy (IMRT), often delivered with a few segments per beam, mostly manually optimized (3–6). Rotational techniques (7, 8) generally showed better performances in better tailoring the dose distribution to the Planning Target Volume (PTV) shape with a consequent improved sparing of organs at risk (OARs) at high-intermediate dose levels, especially in the case of concave-shaped PTVs. Nevertheless, the issues related to the potential clinical meaning of the higher low-dose spread to the heart, lungs, and contralateral breast with rotational techniques is still open (9, 10). Due to this, the TF arrangement, limiting any relevant low-dose spread, is expected to stably remain among the most used techniques to treat breast cancer also in the next decade. On the other hand, manual (and also inverse planned) optimization is time consuming and highly dependent on the planner skill (11). Standardization in radiotherapy treatment planning is an important goal aimed to guarantee to all patients a high quality treatment, independent of the planner time and skills; this seems still more urgent in countries with a rapidly growing incidence of cancer and low/middle income (2). Automated planning solutions were recently introduced (12–25) with the aim of reducing planning time and inter-operator variability while conserving (or improving) a high quality plan (23, 26–29). Many systems were largely investigated for a different clinical situation including the breast site (11, 17, 30–35). Regarding the TF approach for whole breast irradiation, a relatively weak point of auto-planning is the difficulty to take into account the inter-patient variations in assessing the best position of the fields to limit the dose to the adjacent organs, concomitantly assuring PTV coverage and highly homogenous dose distribution within PTV. Consequently, automatic solutions for this application were rarely reported using in-house systems, intrinsically difficult to extend on a larger scale (11, 30, 32).

Automating the optimization of TF by knowledge-based (KB) approaches would be suitable as the (largely available) past information could be modelled to be optimally applied on new patients. The RapidPlan® system (Varian Inc.) is available commercially and widely tested in many clinical scenarios, including breast VMAT (17, 31, 33). However, no clinical examples for the TF irradiation of whole breast are reported likely due to the configuration of the system implemented for IMRT/VMAT inverse-planning optimization. In order to obtain this objective, within a project for the large-scale implementation of automatic KB plan optimization (MIKAPOCo, Multi-Institutional Knowledge-based Approach to Plan Optimization for the Community), we previously demonstrated the possibility to mimic (and slightly improve) the performances of TF irradiation through a partially blocked multi-arcs approach using RapidArc®, named ViTAT [Virtual Tangential-fields Arc Therapy (36)]. Aims of the present work were:

1. to develop KB-models based on TF plans aimed to drive inverse-planned ViTAT plans; and
2. to demonstrate the possibility of replacing TF manual optimization with an automatic KB-ViTAT plan optimization.

MATERIAL AND METHODS

Clinical and Planning Procedures for Tangential Field Planning

At our institute, breast cancer patients treated with whole breast irradiation receives 40 Gy in 15 fractions (2.67 Gy/fr), prescribed as a mean dose to PTV. CTV and PTV are defined according to the AIRO national guidelines (36): in short, CTV is contoured excluding the skin, with a 5-mm margin from the surface while PTV is obtained by a 5-mm isotropic expansion. PTV is finally cropped with a 5-mm margin from the body surface. During the last 10 years, patients were treated mostly with opposed or quasi-opposed, BEV-based (Beam's Eye View) optimized, fields, using physical wedges. Manually optimized segments were generally added to improve the PTV coverage and dose distribution homogeneity. Starting from 2018, this technique was gradually replaced by the manually optimized field-in-field technique (3), avoiding wedges. The number of segments for each field ranged between one (only open fields) and four, with most patients

optimized using 2–3 segments per field. Moreover, for both techniques, the gantry position of few segments could be slightly changed (between 5° and 10°) with respect to the "prevalent" field, aiming to limit hot spots and improve homogeneity within PTV and OARs sparing. Based on the internally conducted plan comparison performed in 2017, the differences between FIF and our TF approach using physical wedges were clinically negligible, in slight favor of the FIF. Patients were treated with 6 MV beams, with few exceptions when 18 MV and 6 MV fields/segments were combined. During the last 5 years, all patients were treated with a Varian CLINAC-IX 2300 equipped with a 5-mm Millenium-MLC system; daily image-guidance with CBCT was performed for all patients; plans were optimized using the Varian Eclipse TPS system (v. 13.6) using the AcurosXB® algorithm for dose calculation.

The planning goals for PTV were: $V_{95\%} > 95\%$ (the fraction of PTV volume receiving more than 95% of the prescribed dose higher that 95%) and $D_{max} < 108\%$. OARs were defined according to the AIRO national guidelines and always included the contralateral breast, the contralateral lung, the ipsilateral lung, and the heart: constraints were $V_{20Gy} < 20\%$ for the ipsilateral lung and $V_{5Gy} < 40\%$, $D_{mean} < 5$ Gy for the heart in the left breast case. Independently from these constraints, planners always tried to reduce the dose to OARs as much as possible while respecting the goals for PTV. Contralateral breast was always avoided by the medial entry beam, while the lateral beam could include small portions at the exit. In the case the constraints could not be respected, PTV coverage received a higher priority against OARs sparing, with few exceptions. Alternatively, a VMAT plan could also be optimized and chosen by the physician.

The ViTAT Technique

As largely explained in Esposito et al. (36), we previously demonstrated the feasibility to mimic the TF irradiation performances through a multi-arcs approach where the delivery was partly blocked using RapidArc® (RA), named Virtual Tangential-fields Arc Therapy (ViTAT). **Figure 1** shows the planning ViTAT set-up.

In short, ViTAT consisted of four arcs (6 MV) with a collimator rotation angle of ±10° and optimized with a RA technique with start/stop angles of 60°/220° for the right side and 300°/135° for the left one, consistent with the TF geometry in our population. Irradiation through the arcs was completely blocked, apart in the first and the last 20° of rotation. As for the breast VMAT-optimization (17), Target structure for plan optimization is obtained expanding the PTV outside the body (1.5 cm expansion in the external-lateral and anterior direction) to account for any breathing and inter-fraction deformation effect. A virtual bolus (1.5 cm thickness, -500 HU density) (17, 37) is linked to the fields during optimization to fill the Target structure during the optimization in order to avoid effects due to the electronic non-equilibrium and to assure a proper safety set-up margin (38, 39). Of course, the bolus is removed before the final dose calculation. Field dimensions are BEV-based adapted to the individual patient anatomy. Plans are optimized with the Rapid Arc (RA) optimization module by inverse planning. We implemented the ViTAT technique in this study with the goal of replacing the manual planning optimization in TF irradiation through a KB-optimization approach.

KB-Model Generation

In order to guarantee a better plan homogeneity, TF clinical plans of the period 2016–2019 were considered for building the model, resulting in 90 plans (70 with wedged fields and 20 with FIF) for the right-side breast. In the same way, 103 (88 with wedged fields and 15 FIF) TF plans in the range 2016–2020 were selected for the left-side breast. These plans were used to train two KB-models, one for the right-side and one for the left-side breast using the RapidPlan (RP) tool implemented in the Eclipse TPS (v 13.6, Varian Inc.). This choice also balanced the need to train a sufficient number of plans and to keep a sufficiently large validation group, considering both the recent FIF (2019–2020) and older TF (before 2016) plans, as explained later. RP is

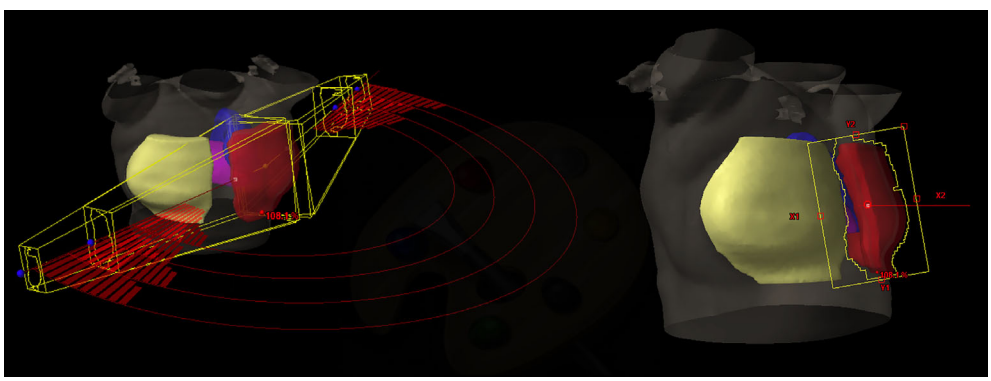


FIGURE 1 | ViTAT setup: 3D view of a right side ViTAT plan showing the geometry of the four arcs used (on the left)—the segments show the beam delivery while the rest of the arcs are blocked for delivery; beam eye view of the medial angle (on the right).

configured to model inverse plans delivered with the IMRT/VMAT technique. Hence, each TF dose distribution was linked to a virtual RA-plan (40) consisting of two partially reverse arcs in the range 60°/220° for the right-side and 300°/135° for the left-side with a collimator rotation angle of ±10°. The choice of the start/stop angles was optimized according to the distribution of the beam angles of the original plans, as explained in (36). During KB modelling, arcs were not considered to be blocked (no avoidance sector were used), so the entire geometry of the structure is seen from the Beams Eye View (BEV) of the arc. Before accomplishing model configuration, the impact of using multiple arcs for building the models on DVH-estimates was investigated. As the partitions of the OAR during the training phase are the same for equal arcs, no differences were observed in the DVH-estimates for model configuration using two arcs instead of four ones (as used for the ViTAT techniques). The OARs considered for training the models were the ipsilateral lung and contralateral breast for the right side plus the heart for the left side.

Right-side heart and contralateral lung were not modelled as they were considered to be more easily managed in the final template (with the same fixed constraints for all patients) without any support of the prediction of the model, due to their anatomical position treatment and to the ballistic of the arc orientations. The tuning of the models was performed by using statistical tools available in the RP system and the Model Analytic platform: the methods followed to limit the impact of outliers were reported elsewhere (25, 41, 42). In short, for each OARs, the features that exceeded by >2 SD the principal components fitted with the KB-models were identified and the original clinical plan were individually re-evaluated; they were excluded as "dosimetric outliers" only in case of recognized sub-optimal planning. All the other potential outliers were found to be representative of an "uncommon" but clinically suitable geometry/anatomy condition and were kept in the models.

Template for Automatic ViTAT Plan Optimization

Based on the resulting prediction models, the system generates the confidence intervals of the expected DVH for each OAR and suggests the lower confidence values to be used for plan optimization. In order to obtain a robust and efficient template, the choice of the position and penalty of the generated DVH constraints needs to be optimized. The optimization of the KB-based template for each side was carried out by several fine-tuning tests based on repetitive automatic plans for five sample patients for each of the two situations (right and left), testing the impact of the position of DVH constraints and their penalties, as similarly reported in previous studies (23–25). Based on our experience, once the PTV priorities are fixed, the OARs priorities were gradually increased through three KB-test templates. Selected relevant dose-volume parameters were analyzed for the five sample patients and compared against the original clinical plans to assess the performances of the three KB-based templates and to assess the one showing the best performances.

Validation Tests

Validation tests were performed on 60 new patients equally divided between right-side and left-side breast. The external validation of the models was performed on patients selected in the range 2019–2020 and 2013–2016.In this way, we also intended to validate the model separately for wedged fields plans (the oldest group) and for the more recent FIF plans (delivered in 2019–2020). The number of FiF plans is 15 for right-sided breast patients and 11 for the left-sided ones, while the number of wedged fields is 15 for right-sided breast patients and 19 for left-sided ones. Importantly, the validation assessed the performances of the model to properly adapt the anatomical/morphological features of each individual patient and the feasibility of using TF DVH-prediction to fully automate the ViTAT optimization. According to others (17, 19, 23, 40), the validation was performed by re-optimizing a number of clinical plans and by comparing them against the original ones. All KB fully-automatic ViTAT-plans (KB-ViTAT) were compared against the original plans (TF) in terms of OARs/PTVs dose-volume parameters. The comparison was based on the analysis of the mean dose, maximum dose ($D_{1\%}$ for PTV and $D_{2\%}$ for OARs), and selected dose-volume parameters extracted from DVH. All selected parameters and DVHs were semi-automatically exported *via* ESAPI scripts and saved in spreadsheets for analysis. Wilcoxon-tests were performed to assess the statistically significant differences ($p < 0.05$).

RESULTS

In total, 10 patients (out of 90) for the right side and 18 patients (out of 103) for the left side were excluded by the model resulting in sub-optimal plans for at least one of the considered OARs or for PTV coverage. The final model returned χ^2 and R^2 parameters for the predicted OARs is shown in **Table 1**. The resulting DVH-estimates were used to generate an individually optimized KB-template for the ViTAT optimization. For the not trained OARs, the position of DVH constraints and their penalties were fixed and tuned as previously explained. Given the overall results, the KB-based template for automatic planning optimization was finally generated and is shown in **Table 2** for the right-side case and in **Table 3** for the left-side one.

Mean DVHs comparison between KB-ViTAT plans and clinical ones are shown in **Figures 2** and **3** for right and left side, respectively. On 30 KB automatic re-optimized cases per each side, only one for the right case and seven for the left case

TABLE 1 | Final model goodness parameter for models for both right and left sided breast cancer patients.

Model	Structure	χ^2	R^2
Right-side breast	Ipsilateral Lung	1.043	0.604
	Contralateral Breast	1.050	0.511
Left-side breast	Ipsilateral Lung	1.043	0.723
	Heart	1.035	0.672
	Contralateral Breast	1.046	0.505

TABLE 2 | The KB-based template for automatic planning optimization for the ViTAT technique for right-sided breast cancer treatment.

Organs	Objectives	Volume (%)	Dose (Gy)	Priority	gEUD a
PTV	Upper	0	40	500	
	Lower	100	40	500	
Target	Upper	0	40	500	
	Lower	100	40	500	
Contralateral Lung	Upper	0	3	600	
	Upper	2.5	1	150	
	Upper	10	0.7	150	
	gEUD		0.3	200	1
Contralateral	Upper	Generated	1.5	600	
Breast	Upper	Generated	1	200	
	Upper	0	Generated	400	
	gEUD		0.5	500	1
Heart	Upper	0	3	600	
	Upper	2.5	1	150	
	Upper	10	0.7	150	
	gEUD		0.5	200	1
Ipsilateral Lung	Upper	0	40	200	
	Upper	Generated	30	200	
	Upper	Generated	20	200	
	Upper	Generated	16	500	
	Upper	Generated	10	400	
	Upper	Generated	5	500	
	Upper	Generated	2	500	

The parameters obtained by the RapidPlan prediction automatically replace the "Generated" placeholder.

TABLE 3 | The KB-based template for automatic planning optimization for the ViTAT technique for left-sided breast cancer treatment.

Organs	Objectives	Volume (%)	Dose (Gy)	Priority	gEUD a
PTV	Upper	0	40	600	
	Lower	100	40	600	
Target	Upper	0	40	600	
	Lower	100	40	600	
Contralateral Lung	Upper	0	3	600	
	Upper	2.5	1	150	
	Upper	10	0.7	150	
	gEUD		0.3	200	1
Contralateral	Upper	Generated	1	500	
Breast	Upper	Generated	1.5	250	
	Upper	0	Generated	450	
	gEUD		0.5	550	1
Heart	Upper	0	40	250	
	Upper	Generated	30	250	
	Upper	Generated	20	250	
	Upper	Generated	16	450	
	Upper	Generated	10	450	
	Upper	Generated	5	500	
	Upper	Generated	2	500	
	gEUD		3.4	500	1
Ipsilateral Lung	Upper	0	40	180	
	Upper	Generated	30	180	
	Upper	Generated	20	180	
	Upper	Generated	16	450	
	Upper	Generated	10	400	
	Upper	Generated	5	450	
	Upper	Generated	2	450	

The parameters obtained by the RapidPlan prediction automatically replace the "Generated" placeholder.

resulted in plans unacceptability in terms of the PTV coverage and/or ipsilateral lung constraints. The reason for this has to be found in the shape and position of the PTV with respect to OARs, causing an uncovering in the medial part of the PTV. However, after the manual refinement of the start/stop angles of 5°/10° by an expert planner, all automatic KB-ViTAT plans well fitted TF-performances. A manual refinement of 5° for the medial angle was necessary for the right-breast case and four of the left-side patients getting the following angles respectively: 65°/220° and 295°/135°. In the remaining three cases for left-side patients, it was necessary to also change the distal angle by 5° and 10° for two and one patients, respectively. Making the modification in the start/stop angles and re-starting the automated optimization, the PTV coverage and the ipsilateral lung constraints were satisfied and comparable to original clinical TF plan. As an example, **Supplementary Figures 1** and **2** (**Supplementary Material**) show the comparison of the originally automatic KB-ViTAT plan with the original start/stop angle and the refined KB-ViTAT with start angle modified by 5° (lower) for the right-side patient and for one of the left-side patients with manual modification of gantry angles.

Quantitative analyses are shown in **Tables 4** and **5**. Overall, differences between TF and KB-ViTAT were small and in slight favor of ViTAT. PTV coverage was similar, while PTV $D_{1\%}$ was improved with automatic optimization (p-value < 0.05). KB-ViTAT better spared contralateral OARs with respect to TF: mean dose was lowered by 0.1 Gy for all contralateral OARs, resulting in a decrease of 33% of the mean dose to the contralateral lung, 20% for the right and 20% for the left side contralateral breast, and 14% for the right-side heart.

The KB-ViTAT integral (i.e., body) dose was 5% lower than TF for the right case and 8% for the left case. Ipsilateral lung mean dose was identical (Right: 6.8 Gy, p > 0.05; Left: 5.7 Gy, p > 0.05): there was a modest worsening (i.e., few % volume) for the ipsilateral lung in the 2–15 Gy range and a slight improvement in the 20-35 Gy range when considering KB-ViTAT vs TF. The same behaviour was found for the left-sided

TABLE 4 | Dose-volume parameters comparison (TF vs ViTAT) for the validation cohort test of 30 plans for the right-sided breast case.

Organs	Features	*TF*	*KB-ViTAT*	*ΔP*
PTV	$V_{95\%}$ (%)	96.7 ± 1.3	96.7 ± 0.9	0.0
	$D_{1\%}$ (Gy)	42.3 ± 0.3	41.8 ± 0.3	**0.5**
	SD (Gy)	1.1 ± 0.1	1.0 ± 0.1	**0.1**
Body	D_{mean} (Gy)	4.2 ± 1.0	4.0 ± 1.0	**0.2**
	$D_{2\%}$ (Gy)	40.4 ± 0.4	40.2 ± 0.4	**0.2**
Heart	D_{mean} (Gy)	0.7 ± 0.2	0.6 ± 0.1	**0.1**
	$D_{2\%}$ (Gy)	1.8 ± 0.4	1.9 ± 0.6	-0.1
Contralateral Lung	D_{mean} (Gy)	0.3 ± 0.1	0.2 ± 0.1	**0.1**
	$D_{2\%}$ (Gy)	0.9 ± 0.4	0.7 ± 0.3	**0.2**
Contralateral Breast	D_{mean} (Gy)	0.5 ± 0.2	0.4 ± 0.2	**0.1**
	$D_{2\%}$ (Gy)	1.7 ± 0.6	1.9 ± 1.3	-0.2
Ipsilateral Lung	V_{5Gy} (%)	25.4 ± 4.8	28.0 ± 3.7	**-2.6**
	V_{20Gy} (%)	14.2 ± 2.7	13.7 ± 2.6	0.5
	D_{mean} (Gy)	6.8 ± 1.1	6.8 ± 1.0	0.0
	$D_{2\%}$ (Gy)	38.0 ± 1.1	37.8 ± 1.1	0.2

Parameters are presented as mean value ± standard deviation and differences ΔP. Values with a statistically significant difference (p-value < 0.05) are in bold.

TABLE 5 | Dose-volume parameters comparison (TF vs ViTAT) for the validation cohort test of 30 plans for the left-sided breast case.

Organs	Features	*TF*	*KB-ViTAT*	*ΔP*
PTV	$V_{95\%}$ (%)	96.6 ± 1.5	96.3 ± 0.9	0.31
	$D_{1\%}$ (Gy)	41.9 ± 0.3	41.8 ± 0.3	**0.1**
	SD (Gy)	1.0 ± 0.3	1.0 ± 0.1	0.0
Body	D_{mean} (Gy)	3.9 ± 0.9	3.6 ± 0.8	**0.3**
	$D_{2\%}$ (Gy)	40.3 ± 0.3	40.1 ± 0.4	**0.2**
Heart	V_{3Gy} (%)	12.1 ± 6.1	16.5 ± 6.5	**-4.4**
	V_{16Gy} (%)	4.1 ± 2.1	3.6 ± 2.1	**0.5**
	D_{mean} (Gy)	2.7 ± 0.9	2.7 ± 0.9	0.0
	$D_{2\%}$ (Gy)	27.4 ± 9.9	23.3 ± 9.0	**4.1**
Contralateral Lung	D_{mean} (Gy)	0.3 ± 0.2	0.2 ± 0.1	**0.1**
	$D_{2\%}$ (Gy)	1.1 ± 0.4	1.0 ± 0.4	0.1
Contralateral Breast	D_{mean} (Gy)	0.6 ± 0.3	0.5 ± 0.2	**0.1**
	$D_{2\%}$ (Gy)	2.1 ± 1.2	2.8 ± 1.2	**-0.7**
Ipsilateral Lung	V_{5Gy} (%)	19.5 ± 5.9	23.8 ± 6.8	**-4.3**
	V_{20Gy} (%)	11.9 ± 3.9	10.9 ± 4.5	**1.0**
	D_{mean} (Gy)	5.7 ± 1.6	5.7 ± 1.7	0.0
	$D_{2\%}$ (Gy)	37.6 ± 3.1	35.9 ± 5.1	**1.7**

Parameters are presented as mean value ± standard deviation and differences ΔP. Values with a statistically significant difference (p-value < 0.05) are in bold.

case for the heart with a modest worsening in the range 2–10 Gy and a slight improvement in the range 15–35 Gy, while on average, delivering the same mean heart dose (2.7 Gy).

Furthermore, in **Figures 4** and **5**, the histograms representing the distribution of the differences between TF and KB-ViTAT were reported for the selected PTV/OARs dose-volume parameters. Overall, the time for automatic plan optimization and final dose calculation was registered and found to be 12 ± 2 minutes.

Importantly, the model performances were also evaluated separately for wedged fields plans (the oldest group) and the more recent FIF plans (delivered in 2019–2020). No statistically significant differences resulted in terms of the PTV/OARs dose-volume parameters when comparing the differences between KB-ViTAT and TF in the two cohorts of plans.

The KB-based automatic approach for ViTAT was clinically implemented, first for the right-sided breast and, more recently, for the left-sided one. Five plans for the right-side and five ones for left-side breast were preliminarily verified in terms of dose distribution in a planar phantom using a detector matrix. All

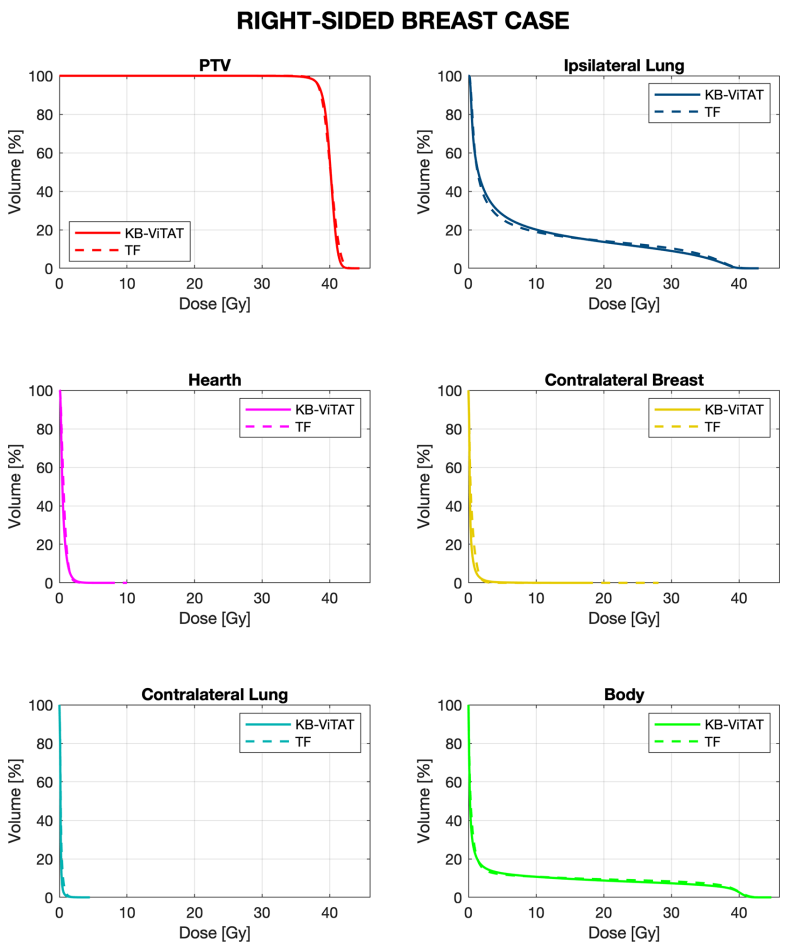


FIGURE 2 | Mean-DVH comparison of 30 right-sided breast patient tests between the original clinical TF (solid lines) and fully automatic KB-ViTAT (dashed lines) plans.

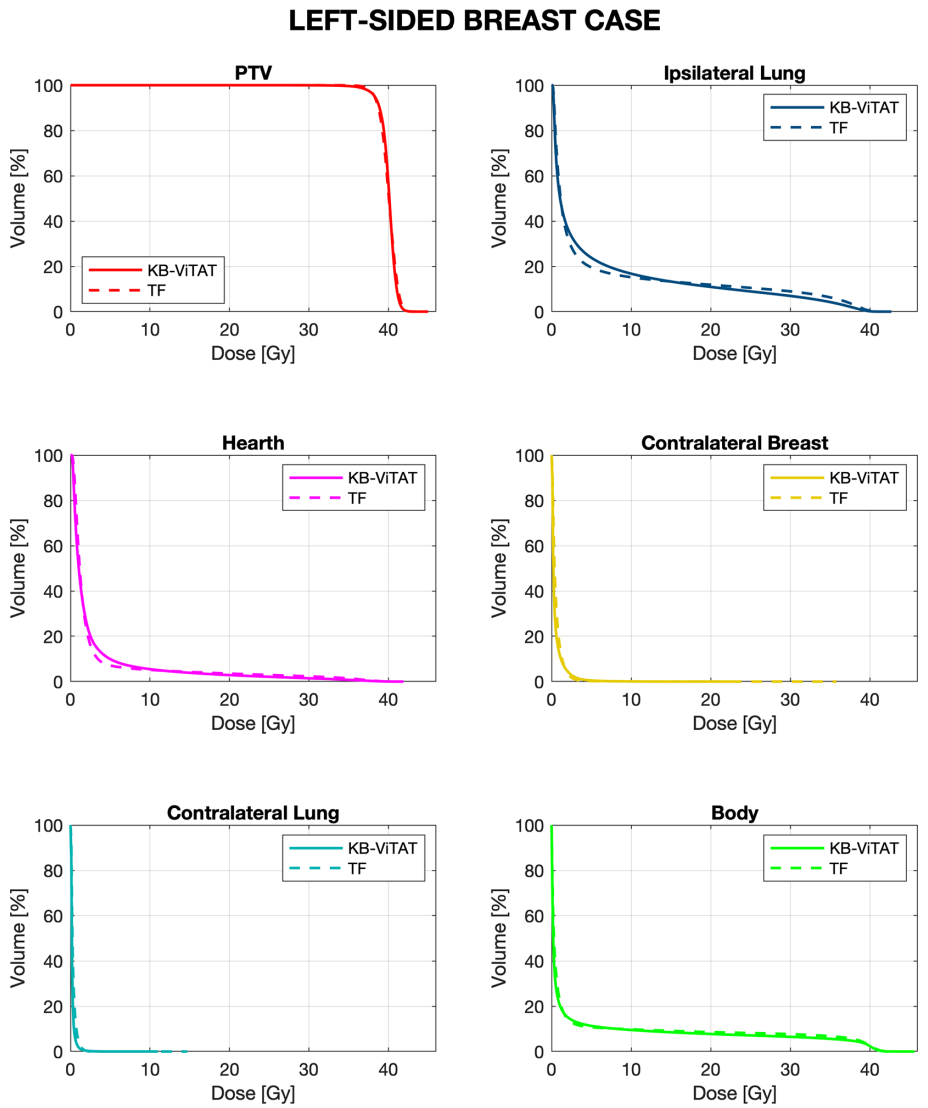


FIGURE 3 | Mean-DVH comparison of 30 left-sided breast patient tests between the original clinical TF (solid lines) and fully automatic KB-ViTAT (dashed lines) plans.

clinical KB-VITAT plans underwent the same dosimetric verification before treatment. The gamma passing rate in comparing the calculated *vs* delivered dose maps was larger than 98% for all plans (in total n = 30), considering 3% - 3 mm as criteria, in agreement with our experience (43).

DISCUSSION

Scope of this work was to fully automate the planning of the whole breast treatment through the KB-optimization approach to mimic the performances of the tangential beam technique. The study aimed at assessing the capability of the RapidPlan tool to handle the dose distributions treated without using the VMAT technique. Moreover, we also explored the feasibility of translating the DVH-prediction model into a fully automatic optimized template for ViTAT planning. As previously shown (36), the ViTAT technique was able to generate dose distributions comparable to TF performances, with some slight improvement in the PTV coverage and homogeneity, in the sparing of contralateral OARs and with a mild reduction of the integral body dose. Then, as demonstrated here, once configured and validated, a KB model trained with TF dose distribution can be efficiently implemented for an automatic ViTAT optimization, completely and efficiently replacing manual plan optimization. The entire workflow was optimized aiming to automatize the selection of the beam angles through the VMAT approach but at the same time to avoid the dose-bath at intermediate–low doses typical of the rotational techniques (7). As a matter of fact, the concerns related to the increase of the mean dose to OARs dealt with the risk of increasing

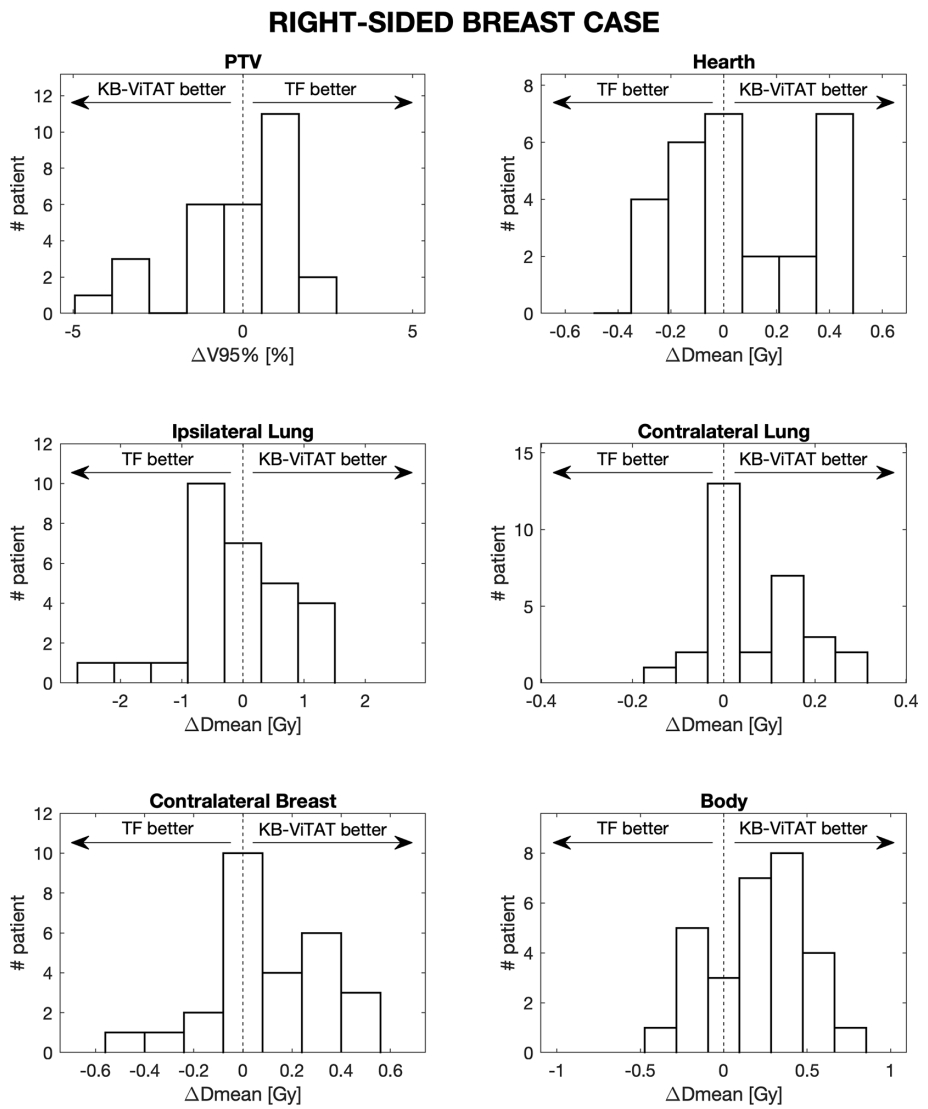


FIGURE 4 | Population histograms of the differences between the clinical tangential field TF and automated re-optimized KB-ViTAT plans for the investigated dosimetric parameters for right-sided breast case, showing the V95% parameter for PTV and mean doses for OARs.

radiation-induced secondary malignancies and late cardiac events (1, 7, 44, 45). Few authors suggested a compromise between better conformity/OARs sparing at high doses against OARs sparing at intermediate–low doses to limit the dose bath associated to VMAT by the partial blocking of arcs (35, 46). Although all these approaches obtained a significant reduction of the low-dose bath, none of them followed the goal of mimicking tangential field irradiation, differently from our ViTAT approach. Other authors demonstrated the possibility of automatizing the angle selection for tangential fields (11) but none investigated the possibility of using the KB-approach. With the approach presented here, we demonstrated the feasibility of the complete replacement of manual TF plans with automatic plans with huge improvements in efficiency, in reducing/eliminating inter-planner variability, and in avoiding sub-optimal plans. Concerning this last point, despite a careful quantitative evaluation was not accomplished, it is relevant that about 15% of the clinical TF plans were *a priori* excluded when building the KB models as due to sub-optimal planning.

As a matter of fact, the automatic workflow that involved the fixed selection of the start/stop angles had proven to be an efficient way to reproduce the TF performances well. Only in about 13% of cases this approach had failed: on 60 KB automatic re-optimized cases, only eight resulted in plans that were unacceptable in terms of the PTV coverage and ipsilateral lung constraints. Seven out of eight referred to the left side were found to be more challenging from the point of view of the choice of the start/stop angles. However, with the manual refinement of the start/stop angles by an expert planner,

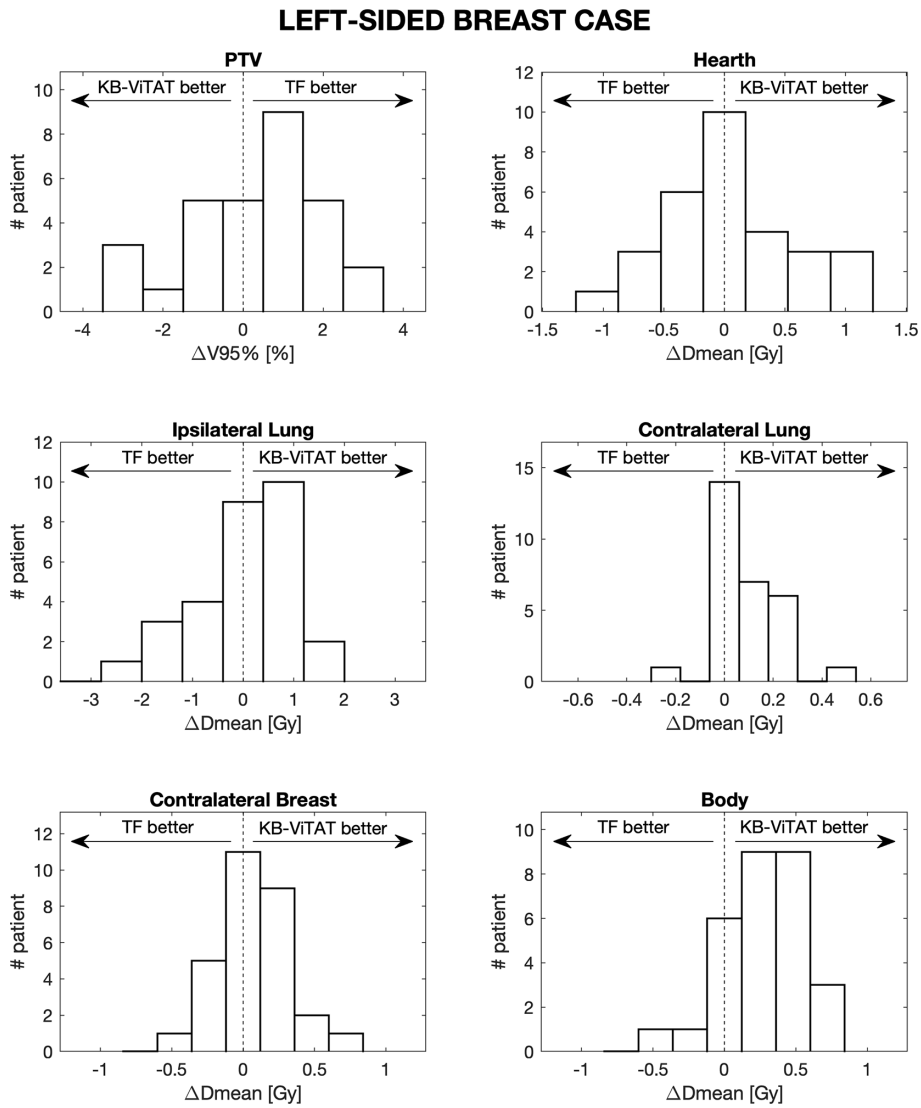


FIGURE 5 | Population histograms of the differences between the clinical tangential field TF and automated re-optimized KB-ViTAT plans for the investigated dosimetric parameters for left-sided breast case, showing the V95% parameter for PTV and mean doses for OARs.

the resulting automatic KB-ViTAT plans well fitted TF-performances also for these patients. Furthermore, plans were deliverable showing excellent dosimetric verification performances in phantoms. Importantly, a large sparing of planning time was obtained, with an overall time for automatic plan optimization and final dose calculation of 12 ± 2 minutes.

A highly relevant point, worthy to be underlined here, concerns the demonstration of the automation of a largely used technique by using a commercially available tool, making its potential adoption easy for Varian users. This issue is of primary importance, opening the possibility to a large-scale implementation with a consequent large reduction of the repetitive, manual procedures usually followed during a whole breast plan optimization. The possible sharing of KB-models, permitted by this system, should also be considered as an additional opportunity to rapidly spread this approach primarily to Varian users but in principle extendible/adaptable to other delivery systems (24, 40).

From a wider point of view, the large availability of TF plans and the "limited" inter-institute variability of PTV/OARs can make the development of robust KB DVH prediction models easier and, more importantly, the possibility of extending this approach on a multi-institutional scale. This issue is worthy of investigation and is currently under study within the MIKAPOCo (Multi-Institutional Knowledge-based approach to plan optimization for the community) consortium, joining several Italian Institutes.

Another relevant point concerns the possibility to exploit the ViTAT approach more to obtain further improved plan

performances by the possibility of training models using the planning data obtained from clinical plans optimized and delivered at our Institute by applying more stressed intensity-modulation techniques: a project to train KB-models using the planning data of patients treated with Tomotherapy (using the Tomo-Direct module, resulting in IMRT-like TF plans) is currently ongoing.

CONCLUSIONS

The approach followed here demonstrated the possibility of the complete replacement of manual tangential breast planning with automatic planning, including beam angle choice. Automatic fixed selection of the start/stop angles and KB-driven optimization were found to be an efficient way to well fit TF planning, with evident advantages in terms of time sparing, elimination of inter-planner variability, and of sub-optimal planning. Manual refinement of the start/stop angles was necessary in 13% of patients, with a large unbalance between the right (3%) and left (23%) sides, resulting in an additional 15–20 minutes more compared to the 12 ± 2 minutes spent for planning optimization and dose calculation in the remaining 87% of the patients. Due to its versatility and the use of a commercial system, this approach shows promising applications for a large-scale implementation.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because they are data of patients treated in a single center and cannot be shared without transfer agreement. Requests to access the datasets should be directed to fiorino.claudio@hsr.it.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

RC, PE, and CF contributed with the conception of the study. RC, PE, and AT worked on data base and modeling. PM, SB, BL, and LP contributed in data integrity evaluation. CF, AF, SB, PM, NM, RC, and PE discussed the interpretation of the results. AF, CDA, NM, and MP were clinical responsible for the treatment of the patient. RC, PE, and CF wrote the manuscript. AV, NM, and CF reviewed and edited the manuscript. CF was responsible for the scientific coordination, managing, and project funding. All authors contributed to the article and approved the submitted version.

FUNDING

The work was supported by an AIRC (Associazione Italiana Ricerca sul Cancro) grant (IG 23150).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2021.712423/full#supplementary-material

Supplementary Figure 1 | Comparison of plans between the originally automatic KB-ViTAT plan with start/stop angle of 60°/220° (upper) and the refined KB-ViTAT with start angle of 65° (lower) for the right-sided breast case. Starting angles are modified in order to obtain an acceptable coverage to PTV.

Supplementary Figure 2 | Comparison of plans between the originally automatic KB-ViTAT with start/stop angle of 300°/135° (upper) and the refined KB-ViTAT with start angle of 295° (lower)for the left-sided breast case. Starting angles are modified in order to obtain an acceptable coverage to PTV and better hearth and lung dose distribution similarity between TF and KB-ViTAT case.

REFERENCES

1. Darby S, McGale P, Correa C, Taylor C, Arriagada R, Clarke M, et al. Effect of Radiotherapy After Breast-Conserving Surgery on 10-Year Recurrence and 15-Year Breast Cancer Death: Meta-Analysis of Individual Patient Data for 10 801 Women in 17 Randomised Trials. *Lancet* (2011) 378:1707–16. doi: 10.1016/S0140-6736(11)61629-2
2. WHO. *Who Report on Cancer: Setting Priorities, Investing Wisely and Providing Care for All*. Geneva: World Health Organization (2020) Licence: CC BY-NC-SA 3.0 IGO.
3. Chui C-S, Hong L, Hunt M, McCormick B. A Simplified Intensity Modulated Radiation Therapy Technique for the Breast. *Med Phys* (2002) 29:522–9. doi: 10.1118/1.1460875
4. Pignol J-P, Olivotto I, Rakovitch E, Gardner S, Sixel K, Beckham W, et al. A Multicenter Randomized Trial of Breast Intensity-Modulated Radiation Therapy to Reduce Acute Radiation Dermatitis. *J Clin Oncol* (2008) 26:2085–92. doi: 10.1200/JCO.2007.15.2488
5. Donovan EM, Yarnold JR, Adams EJ, Morgan A, Warrington APJ, Evans PM. An Investigation Into Methods of IMRT Planning Applied to Breast Radiotherapy. *Br J Radiol* (2008) 81:311–22. doi: 10.1259/bjr/28583675
6. Smith W, Menon G, Wolfe N, Ploquin N, Trotter T, Pudney D. IMRT for the Breast: A Comparison of Tangential Planning Techniques. *Phys Med Biol* (2010) 55:1231–41. doi: 10.1088/0031-9155/55/4/022
7. Nichols GP, Fontenot JD, Gibbons JP, Sanders ME. Evaluation of Volumetric Modulated Arc Therapy for Postmastectomy Treatment. *Radiat Oncol* (2014) 9:66. doi: 10.1186/1748-717X-9-66
8. Johansen S, Cozzi L, Olsen DR. A Planning Comparison of Dose Patterns in Organs at Risk and Predicted Risk for Radiation Induced Malignancy in the Contralateral Breast Following Radiation Therapy of Primary Breast Using Conventional, IMRT and Volumetric Modulated Arc Treatment Techniqu. *Acta Oncol (Madr)* (2009) 48:495–503. doi: 10.1080/02841860802657227
9. Darby SC, Ewertz M, McGale P, Bennet AM, Blom-Goldman U, Brønnum D, et al. Risk of Ischemic Heart Disease in Women After Radiotherapy for Breast Cancer. *N Engl J Med* (2013) 368:987–98. doi: 10.1056/nejmoa1209825
10. Johansen H, Kaae S, Johansen H, Kaae S, Jensen M-B, Mouridsen HT. Extended Radical Mastectomy Versus Simple Mastectomy Followed by Radiotherapy in Primary Breast Cancer. A Fifty-Year Follow-Up to the

Copenhagen Breast Cancer Randomised Study. *Acta Oncol (Madr)* (2008) 47:633–8. doi: 10.1080/02841860801989753
11. Penninkhof J, Spadola S, Breedveld S, Baaijens M, Lanconelli N, Heijmen B. Individualized Selection of Beam Angles and Treatment Isocenter in Tangential Breast Intensity Modulated Radiation Therapy. *Int J Radiat Oncol Biol Phys* (2017) 98:447–53. doi: 10.1016/j.ijrobp.2017.02.008
12. Yan H, Yin F-F, Willett C. Evaluation of an Artificial Intelligence Guided Inverse Planning System: Clinical Case Study. *Radiother Oncol* (2007) 83:76–85. doi: 10.1016/j.radonc.2007.02.013
13. Monz M, Küfer KH, Bortfeld TR, Thieke C. Pareto Navigation - Algorithmic Foundation of Interactive Multi-Criteria IMRT Planning. *Phys Med Biol* (2008) 53:985–98. doi: 10.1088/0031-9155/53/4/011
14. Teichert K, Süss P, Serna JI, Monz M, Küfer KH, Thieke C. Comparative Analysis of Pareto Surfaces in Multi-Criteria IMRT Planning. *Phys Med Biol* (2011) 56:3669–84. doi: 10.1088/0031-9155/56/12/014
15. Breedveld S, Storchi PRM, Voet PWJ, Heijmen BJM. ICycle: Integrated, Multicriterial Beam Angle, and Profile Optimization for Generation of Coplanar and Noncoplanar IMRT Plans. *Med Phys* (2012) 39:951–63. doi: 10.1118/1.3676689
16. Xhaferllari I, Wong E, Bzdusek K, Lock M, Chen JZ. Automated IMRT Planning With Regional Optimization Using Planning Scripts. *J Appl Clin Med Phys* (2013) 14:176–91. doi: 10.1120/jacmp.v14i1.4052
17. Fogliata A, Nicolini G, Bourgier C, Clivio A, De Rose F, Fenoglietto P, et al. Performance of a Knowledge-Based Model for Optimization of Volumetric Modulated Arc Therapy Plans for Single and Bilateral Breast Irradiation. *PloS One* (2015) 10:e0145137. doi: 10.1371/journal.pone.0145137
18. Wu H, Jiang F, Yue H, Zhang H, Wang K, Zhang Y. Applying a RapidPlan Model Trained on a Technique and Orientation to Another: A Feasibility and Dosimetric Evaluation. *Radiat Oncol* (2016) 11:1–7. doi: 10.1186/s13014-016-0684-9
19. Hussein M, South CP, Barry MA, Adams EJ, Jordan TJ, Stewart AJ, et al. Clinical Validation and Benchmarking of Knowledge-Based IMRT and VMAT Treatment Planning in Pelvic Anatomy. *Radiother Oncol* (2016) 120:473–9. doi: 10.1016/j.radonc.2016.06.022
20. Delaney AR, Tol JP, Dahele M, Cuijpers J, Slotman BJ, Verbakel WFAR. Effect of Dosimetric Outliers on the Performance of a Commercial Knowledge-Based Planning Solution. *Int J Radiat Oncol Biol Phys* (2016) 94:469–77. doi: 10.1016/j.ijrobp.2015.11.011
21. Moore KL. Automated Radiotherapy Treatment Planning. *Semin Radiat Oncol* (2019) 29:209–18. doi: 10.1016/j.semradonc.2019.02.003
22. Ge Y, Wu QJ. Knowledge-Based Planning for Intensity-Modulated Radiation Therapy: A Review of Data-Driven Approaches. *Med Phys* (2019) 46:2760–75. doi: 10.1002/mp.13526
23. Castriconi R, Fiorino C, Broggi S, Cozzarini C, Di Muzio N, Calandrino R, et al. Comprehensive Intra-Institution Stepping Validation of Knowledge-Based Models for Automatic Plan Optimization. *Phys Med* (2019) 57:231–7. doi: 10.1016/j.ejmp.2018.12.002
24. Castriconi R, Cattaneo GM, Mangili P, Esposito PG, Broggi S, Cozzarini C, et al. Clinical Implementation of Knowledge-Based (KB) Automatic Plan Optimization for Helical Tomotherapy. *Pract Radiat Oncol* (2021) 11(2): E236–44. doi: 10.1016/j.prro.2020.09.012
25. Castriconi R, Fiorino C, Passoni P, Broggi S, Di Muzio NG, Cattaneo GM, et al. Knowledge-Based Automatic Optimization of Adaptive Early-Regression-Guided VMAT for Rectal Cancer. *Phys Med* (2020) 70:58–64. doi: 10.1016/j.ejmp.2020.01.016
26. 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
27. Nawa K, Haga A, Nomoto A, Sarmiento RA, Shiraishi K, Yamashita H, et al. Evaluation of a Commercial Automatic Treatment Planning System for Prostate Cancers. *Med Dosim* (2017) 42:203–9. doi: 10.1016/j.meddos.2017.03.004
28. Schubert C, Waletzko O, Weiss C, Voelzke D, Toperim S, Roeser A, et al. Intercenter Validation of a Knowledge Based Model for Automated Planning of Volumetric Modulated Arc Therapy for Prostate Cancer. The Experience of the German RapidPlan Consortium. *PloS One* (2017) 12(5):e0178034. doi: 10.1371/journal.pone.0178034
29. Good D, Lo J, Lee WR, Wu QJ, Yin F-F, Das SK, et al. A Knowledge-Based Approach to Improving and Homogenizing Intensity Modulated Radiation Therapy Planning Quality Among Treatment Centers: An Example Application to Prostate Cancer Planning. *Int J Radiat Oncol Biol Phys* (2013) 87:176–81. doi: 10.1016/j.ijrobp.2013.03.015
30. Zhao X, Kong D, Jozsef G, Chang J, Wong EK, Formenti SC, et al. Automated Beam Placement for Breast Radiotherapy Using a Support Vector Machine Based Algorithm. *Med Phys* (2012) 39:2536–43. doi: 10.1118/1.3700736
31. Wang W, Purdie TG, Rahman M, Marshall A, Liu F-F, Fyles A. Rapid Automated Treatment Planning Process to Select Breast Cancer Patients for Active Breathing Control to Achieve Cardiac Dose Reduction. *Int J Radiat Oncol Biol Phys* (2012) 82:386–93. doi: 10.1016/j.ijrobp.2010.09.026
32. Purdie TG, Dinniwell RE, Fyles A, Sharpe MB. Automation and Intensity Modulated Radiation Therapy for Individualized High-Quality Tangent Breast Treatment Plans. *Int J Radiat Oncol Biol Phys* (2014) 90:688–95. doi: 10.1016/j.ijrobp.2014.06.056
33. Nicolini G, Clivio A, Fogliata A, Vanetti E, Cozzi L. Simultaneous Integrated Boost Radiotherapy for Bilateral Breast: A Treatment Planning and Dosimetric Comparison for Volumetric Modulated Arc and Fixed Field Intensity Modulated Therapy. *Radiat Oncol* (2009) 4:27. doi: 10.1186/1748-717X-4-27
34. Fogliata A, Seppälä J, Reggiori G, Lobefalo F, Palumbo V, De Rose F, et al. Dosimetric Trade-Offs in Breast Treatment With VMAT Technique. *Br J Radiol* (2017) 90:20160701. doi: 10.1259/bjr.20160701
35. Tsai PF, Lin S-M, Lee S-H, Yeh C-Y, Huang Y-T, Lee C-C, et al. The Feasibility Study of Using Multiple Partial Volumetric-Modulated Arcs Therapy in Early Stage Left-Sided Breast Cancer Patients. *J Appl Clin Med Phys* (2012) 13:62–73. doi: 10.1120/jacmp.v13i5.3806
36. Esposito PG, Castriconi R, Mangili P, Fodor A, Pasetti M, Di Muzio NG, et al. Virtual Tangential-Fields Arc Therapy (ViTAT) for Whole Breast Irradiation: Technique Optimization and Validation. *Phys Med* (2020) 77:160–8. doi: 10.1016/j.ejmp.2020.08.011
37. Lizondo M, Latorre-Musoll A, Ribas M, Carrasco P, Espinosa N, Coral A, et al. Pseudo Skin Flash on VMAT in Breast Radiotherapy: Optimization of Virtual Bolus Thickness and HU Values. *Phys Med* (2019) 63:56–62. doi: 10.1016/j.ejmp.2019.05.010
38. Ricotti R, Ciardo D, Fattori G, Leonardi MR, Morra A, Dicuonzo S, et al. Intra-Fraction Respiratory Motion and Baseline Drift During Breast Helical Tomotherapy. *Radiother Oncol* (2017) 122:79–86. doi: 10.1016/j.radonc.2016.07.019
39. Goddu SM, Yaddanapudi S, Pechenaya OL, Chaudhari SR, Klein EE, Khullar D, et al. Dosimetric Consequences of Uncorrected Setup Errors in Helical Tomotherapy Treatments of Breast-Cancer Patients. *Radiother Oncol* (2009) 93:64–70. doi: 10.1016/j.radonc.2009.07.013
40. Cagni E, Botti A, Micera R, Galeandro M, Sghedoni R, Orlandi M, et al. Knowledge-Based Treatment Planning: An Inter-Technique and Inter-System Feasibility Study for Prostate Cancer. *Phys Med* (2017) 36:38–45. doi: 10.1016/j.ejmp.2017.03.002
41. Alpuche Aviles JE, Cordero Marcos MI, Sasaki D, Sutherland K, Kane B, Kuusela E. Creation of Knowledge-Based Planning Models Intended for Large Scale Distribution: Minimizing the Effect of Outlier Plans. *J Appl Clin Med Phys* (2018) 19:215–26. doi: 10.1002/acm2.12322
42. Cagni E, Botti A, Wang Y, Iori M, Petit SF, Heijmen BJM. Pareto-Optimal Plans as Ground Truth for Validation of a Commercial System for Knowledge-Based DVH-Prediction. *Phys Med* (2018) 55:98–106. doi: 10.1016/j.ejmp.2018.11.002
43. Rinaldin G, Perna L, Agnello G, Pallazzi G, Cattaneo GM, Fiorino C, et al. Quality Assurance of Rapid Arc Treatments: Performances and Pre-Clinical Verifications of a Planar Detector (MapCHECK2). *Phys Med* (2014) 30:184–90. doi: 10.1016/j.ejmp.2013.05.004
44. Grantzau T, Thomsen MS, Væth M, Overgaard J. Risk of Second Primary Lung Cancer in Women After Radiotherapy for Breast Cancer. *Radiother Oncol* (2014) 111:366–73. doi: 10.1016/j.radonc.2014.05.004
45. Virén T, Heikkilä J, Myllyoja K, Koskela K, Lahtinen T, Seppälä J. Tangential Volumetric Modulated Arc Therapy Technique for Left-Sided Breast Cancer Radiotherapy. *Radiat Oncol* (2015) 10:79. doi: 10.1186/s13014-015-0392-x

46. Yu AF, Jones LW. Breast Cancer Treatment-Associated Cardiovascular Toxicity and Effects of Exercise Countermeasures. *Cardiooncology* (2016) 2:1. doi: 10.1186/s40959-016-0011-5

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: 08 September 2021
doi: 10.3389/fonc.2021.713617

MRI-Only Radiotherapy Planning for Nasopharyngeal Carcinoma Using Deep Learning

Xiangyu Ma[1†], Xinyuan Chen[1†], Jingwen Li[2], Yu Wang[1], Kuo Men[1] and Jianrong Dai[1*]*

[1] National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China, [2] Cloud Computing and Big Date Research Institute, China Academy of Information and Communications Technology, Beijing, China

Background: Radical radiotherapy is the main treatment modality for early and locally advanced nasopharyngeal carcinoma (NPC). Magnetic resonance imaging (MRI) has the advantages of no ionizing radiation and high soft-tissue resolution compared to computed tomography (CT), but it does not provide electron density (ED) information for radiotherapy planning. Therefore, in this study, we developed a pseudo-CT (pCT) generation method to provide necessary ED information for MRI-only planning in NPC radiotherapy.

Methods: Twenty patients with early-stage NPC who received radiotherapy in our hospital were investigated. First, 1433 sets of paired T1 weighted magnetic resonance (MR) simulation images and CT simulation images were rigidly registered and preprocessed. A 16-layer U-Net was used to train the pCT generative model and a "pix2pix" generative adversarial network (GAN) was also trained to compare with the pure U-Net regrading pCT quality. Second, the contours of all target volumes and organs at risk in the original CT were transferred to the pCT for planning, and the beams were copied back to the original CT for reference dose calculation. Finally, the dose distribution calculated on the pCT was compared with the reference dose distribution through gamma analysis and dose-volume indices.

Results: The average time for pCT generation for each patient was 7.90 ± 0.47 seconds. The average mean (absolute) error was –9.3 ± 16.9 HU (102.6 ± 11.4 HU), and the mean-root-square error was 209.8 ± 22.6 HU. There was no significant difference between the pCT quality of pix2pix GAN and that of pure U-Net ($p > 0.05$). The dose distribution on the pCT was highly consistent with that on the original CT. The mean gamma pass rate (2 mm/3%, 10% low dose threshold) was 99.1% ± 0.3%, and the mean absolute difference of nasopharyngeal PGTV $D_{99\%}$ and PTV $V_{95\%}$ were 0.4% ± 0.2% and 0.1% ± 0.1%.

Conclusion: The proposed deep learning model can accurately predict CT from MRI, and the generated pCT can be employed in precise dose calculations. It is of great significance to realize MRI-only planning in NPC radiotherapy, which can improve structure delineation and considerably reduce additional imaging dose, especially when an MR-guided linear accelerator is adopted for treatment.

Keywords: nasopharyngeal carcinoma, radiotherapy, MRI-only planning, pseudo CT, deep learning, dosimetric evaluation

OPEN ACCESS

Edited by:
Jose Eduardo Villarreal Barajas, Royal Devon and Exeter Hospital, United Kingdom

Reviewed by:
Ruijie Yang, Peking University Third Hospital, China
Joshua Pohyun Kim, Henry Ford Health System, United States

***Correspondence:**
Kuo Men
menkuo126@126.com
Jianrong Dai
dai_jianrong@cicams.ac.cn

[†]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: *23 May 2021*
Accepted: *25 August 2021*
Published: *08 September 2021*

Citation:
Ma X, Chen X, Li J, Wang Y, Men K and Dai J (2021) MRI-Only Radiotherapy Planning for Nasopharyngeal Carcinoma Using Deep Learning. Front. Oncol. 11:713617. doi: 10.3389/fonc.2021.713617

INTRODUCTION

Nasopharyngeal carcinoma (NPC) is the most common malignant tumor in the head and neck (HN), especially in southern China and Southeast Asia. Radical radiotherapy (RT) is the main treatment modality for early or locally advanced NPC, and computed tomography (CT) is necessary for patient positioning and RT planning, since it provides electron density (ED) information for dose calculation. Magnetic resonance imaging (MRI) has the advantages of high soft-tissue resolution and no additional imaging dose compared to CT. With the development of the MR-guided linear accelerator (MR-linac), an *MRI-only* RT-planning workflow is desirable. However, MRI does not provide ED information, which hinders its application in RT planning. Therefore, there is a need for a reliable and effective method to predict ED information based on MR images.

Currently, this issue is addressed using three main methods. The primary one is to simply segment soft tissue and bone (1, 2) and assign the densities of water and bone to them, respectively. However, it is difficult to distinguish between bone and air in MRI.

The second is the atlas-based pseudo-CT (pCT) generation. It requires a deformable registration from an MRI atlas to the patient MRI to obtain a special transformation, which is then applied to a paired CT atlas to generate pCT images (3–8). However, when the patient's MRI is quite different from the image in the atlas library, and there is a special anatomical structure (such as a large tumor or surgical cavity), leading to deformation registration errors, which affect the accuracy of pCT.

The third is the voxel-wise pCT generation (9–12). By establishing a voxel-wise pCT generation model, point-by-point prediction is performed. This method prevents manual or semi-manual segmentation of soft tissue and bone, and it is not sensitive to abnormal anatomical structures. Earlier studies employed machine learning methods, such as cluster analysis, Gaussian regression, and principal component analysis, to establish such a generative model. However, some of them still need manual or semi-automatic delineation of bone and air cavity, and the prediction accuracy still needs to be improved.

Recently, convolutional neural network (CNN) and its derivate deep learning models have been widely used for cross-modality image generation owing to their ability to automatically extract multilevel features of data. At present, most studies on MR-pCT generation focus on brain and prostate RT (11, 13–16), and promising accuracy has been achieved. However, there is a need for further studies to develop and verify deep learning based pCT generative models for treatment sites with more intertwined air cavities and bony structures, such as HN. A previous study (17) used U-Net and T2-weighted (T2w) MRI to generate HN pCT for NPC and reported a promising image quality. In this study, we adopted U-Net but with T1w MRI, another routine clinical MRI modality, to train an MR-pCT generative model for NPC. We not only evaluated the CT number prediction accuracy but also systematically analyzed the dosimetric difference between the obtained pCT and the corresponding original CT with the same beam layout. Besides, we compared the performance of the generative adversarial network (GAN), another popular deep learning network, with U-Net for pCT generation.

MATERIALS AND METHODS

Image Collection

The image data in this study were obtained from 20 patients with NPC who received RT in our hospital from September 2017 to April 2018. All data are retrospective and nonidentifiable so that the institutional ethics review and written consent are exempted.

Before treatment, all patients underwent CT and MR simulation scanning in our department within very close time and with the same fixing devices for each patient. CT scanning was performed using a CT simulator (SOMATOM Definition AS 40, Brilliance CT big bore, Philips) with the acquisition parameters (voltage: 120 kV; exposure: 240 mAs; pitch: 0.94; image size: 512×512; pixel spacing: 0.96 mm; slice thickness: 3.0 mm). MRI scanning was performed using a 3.0-T MR simulator (Discovery MR750w, GE Healthcare) with a 6-channel split head coil and T1-FSE sequence with the acquisition parameters (repetition time (TR): 834 ms; echo time (TE): 7.96 ms; flip angle: 111°; image size: 512×512; pixel spacing: 0.55 mm; and slice thickness: 3 mm). All patients were fixed with head-neck-shoulder thermoplastic film. The upper boundary of the scanning range is half of the frontal sinus, and the lower boundary extends to the supraclavicular region.

Data Preprocessing

Due to the design and characteristics of the coil, the signal intensity distribution of the same tissue might be uneven. We used an N3 algorithm to calibrate the bias field and performed gray value normalization and histogram matching of the MR images. Then, the MR and CT images were rigidly registered, and outlines were drawn on the aligned CT and MR images, respectively, using the thresholding segmentation algorithm. The overlap of the two outlines was used to generate a mask, and the density outside the mask was set equal to that of air.

Deep Learning Architecture for MR-pCT Generation

Two deep learning models, CNN and conditional GAN (cGAN), were adopted in this study for comparison. For both models, we used several data enhancement techniques, including random clipping and flipping, to expand the number of data. The Adam method was used to optimize the loss function. The initial learning rate was set to 0.0002, and the maximum number of iterations was set to 40000. The network training and test were based on the Tensorflow platform and NVIDIA Tesla K80 GPU.

The architecture of the CNN model was 16-layer U-Net, which was developed for MR-pCT generation (**Figure 1**, *red box*). As shown in **Table 1**, the modules of convolution-batch

normalization and rectified linear units (ReLU) were used for the encoder–decoder network. The kernel size was 4 × 4 for the convolutional and deconvolutional layers. Skip connections were added between each mirrored encoder and decoder layers for better recovery of image details.

The architecture of cGAN, as shown in **Figure 1** *(green box)*, was the "pix2pix" model, which used paired MR and CT images as input and ground truth, respectively. It learned a loss that uses a discriminator to determine if an output image is real or fake while simultaneously training a generator to minimize the loss. GAN is supposed to have the ability to overcome problems such as image blurring. For better comparison with the performance of U-Net, the generator part adopted the same U-Net architecture as aforementioned. The patch size of the discriminator was set to 70 × 70. We adopted "cGAN + L1" as a loss function, as suggested in (18). It comprises a standard cGAN loss function and a weighted L1 distance term.

To get a reliable and stable model based on a small sample size, a 10-fold cross-validation method was used to train the pCT generation model. Through cross-validation, optimal model parameters were determined and then used to generate pCT images for the 20 patients. Then, voxel-wise Hounsfield units (HU) comparison was performed between the pCT and the original CT for each patient, considering the mean error (ME), mean absolute error (MAE), and root-mean-square error (RMSE) (Equations 1–3).

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|p(i) - g(i)| \tag{1}$$

$$ME = \frac{1}{N}\sum_{i=1}^{N}(p(i) - g(i)) \tag{2}$$

TABLE 1 | Sixteen-layer U-Net architecture.

Encoder	Decoder
Conv 1 + BN + ReLU (512 × 512 × 64)	De_Conv 9 + BN + ReLU (4 × 4 × 512)
Conv 2 + BN + ReLU (256 × 256 × 128)	De_Conv 10 + BN + ReLU (8 × 8 × 512)
Conv 3 + BN + ReLU (128 × 128 × 256)	De_Conv 11 + BN + ReLU (16 × 16 × 512)
Conv 4 + BN + ReLU (64 × 64 × 512)	De_Conv 12 + BN + ReLU (32 × 32 × 512)
Conv 5 + BN + ReLU (32 × 32 × 512)	De_Conv 13 + BN + ReLU (64 × 64 × 256)
Conv 6 + BN + ReLU (16 × 16 × 512)	De_Conv 14+ BN + ReLU (128 × 128 × 128)
Conv 7 + BN + ReLU (8 × 8 × 512)	De_Conv 15 + BN + ReLU (256 × 256× 64)
Conv 8 + BN + ReLU (4 × 4 × 512)	De_Conv 16 + BN + ReLU (512 × 512 × 1)

The encoder input and decoder output image sizes are both 512 × 512. Conv, convolution; De_Conv, deconvolution; BN, Batch normalization; ReLU, rectified linear units.

$$RMSE = \sqrt{\sum_{i=1}^{N}\frac{p(i)^2 - g(i)^2}{N}} \tag{3}$$

where N is the total number of voxels of interest, $p(i)$ the value of the i-th voxel in the pseudo-CT, and $g(i)$ the corresponding voxel value in the ground truth (original) CT.

Dosimetric Evaluation

For each patient, contours of target volumes and organs at risk (OARs) were transferred from the original planning CT by rigid fusion to the corresponding pCT. RT plans were first designed with a Pinnacle treatment planning system (Philips) based on the pCT, and the beams were then copied to the original CT with the

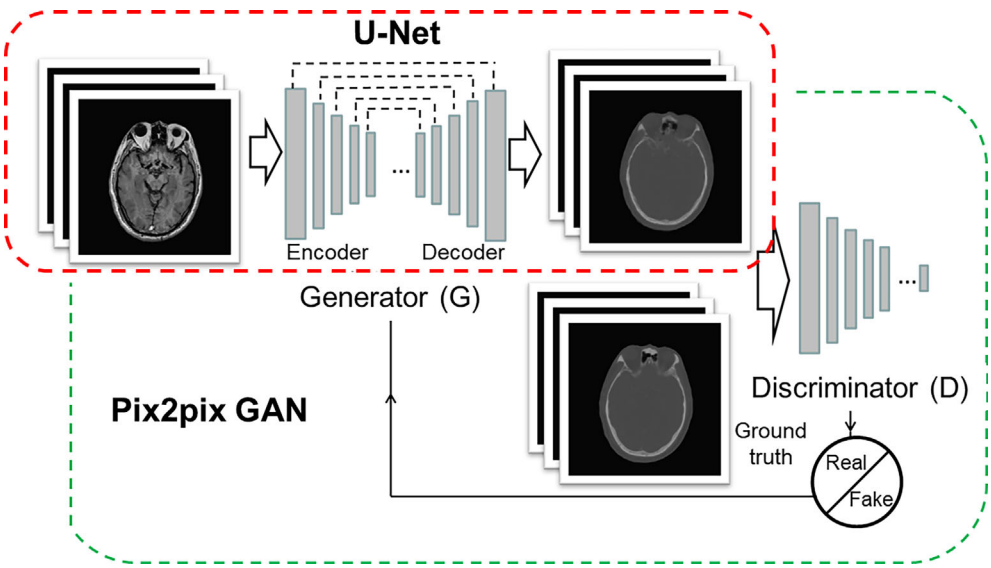


FIGURE 1 | Architectures of U-Net (red box) and pix2pix GAN (green box). The U-Net is composed of an encoder and a decoder, and each of them has eight mosaic layers, which are detailed in **Table 1**.

same isocenter for the ground truth dose calculation. Gamma analysis was performed using Sun Nuclear Patient (SNC Patient) software with a 2 mm/3% (global mode, 10% low dose threshold) criterion to compare dose distributions of pCT and original CT. The gamma criterion for the calculation is inconsistent in the literature, ranging from 1 mm/1% to 3 mm/3% (14–16). Hence, we used an intermediate value herein. For each patient, 10 slices with a 10-mm interval near the image central slice were selected for 2D gamma analysis, and the mean gamma pass rate of the slices was calculated for dosimetry consistency assessment. Besides, a dose-volume histogram (DVH) comparison was performed to evaluate the accuracy of the clinically concerned dosimetry metrics of PTVs and OARs.

RESULTS

Performance Comparison Between U-Net and pix2pix GAN

The quality of pCT generated by U-Net and pix2pix GAN has no statistical difference (paired t-test, $p > 0.05$) in terms of ME, MAE, and RMSE (**Table 2**), whereas the performance of U-Net was slightly better than that of pix2pix GAN. A visual comparison of the pCTs generated by the two types of networks is shown in **Figure 2**, and a spatial discrepancy map for the U-Net is shown in **Figure 3**. Since there was no significant difference between the performance of U-Net and pix2pix GAN in this task, we adopted the simpler-structured U-Net for the further dosimetric comparison.

Dosimetric Consistency Between Real CT and pCT-Based RT Planning

The spatial dose distribution of the pCT-based RT plan could be replicated very well on real CT with the same beam layout, and it demonstrates a good overlap between pCT and CT, regarding the structure DVH (**Figure 4**). Detailed DVH metric comparison is shown in **Table 3**.

The mean (± standard deviation) gamma pass rate of all the patients was 99.1% ± 0.3%, and the median gamma pass rate of all the selected slices was 99.3%, demonstrating a high consistency between the real CT and pCT-based RT planning. The worst slice pass rate was 95.9%, and the best was 100%. An exemplary gamma analysis result is shown in **Figure 5**. Notably, the positions that failed to pass the analysis are all in the peripheral areas because the CT and MR images were not acquired simultaneously, thus they could not be perfectly registered, especially near the outline.

TABLE 2 | Prediction performance comparison of U-Net and pix2pix GAN.

Quality metrics	U-Net	GAN	*p-value
Average ME (HU)	–9.3 ± 16.9	–8.7 ± 17.3	0.325
Average MAE (HU)	102.6 ± 11.4	104.2 ± 12.5	0.051
Average RMSE (HU)	209.8 ± 22.6	213.2 ± 24.1	0.067

*Paired t-test.

DISCUSSION

In this study, deep CNN (U-Net) was used to generate pCT from T1 weighted MRI for NPC radiotherapy, and the dosimetric accuracy was assessed for pCT-based RT planning. We proved that the deep learning network can reliably convert MR images to pCT for HN position to provide ED information. Although deep learning network usually needs much training data, we achieved promising intensity and dosimetry prediction accuracy with limited data.

Previous studies on MR-pCT generation mostly focused on the brain or prostate, and only a few considered HN. The variance of existing brain pCT generation quality recorded by atlas-based studies is relatively large (average MAE from 85 to 184 HU) (3, 6, 19), which is attributed to the different data and image processing algorithms. For machine learning-based studies, Gudur et al. (20) used a Bayesian probability model and realized an average MAE of 126 HU for brain pCT generation. Despite the larger structure density variance around NPC than brain tumors, the pCT quality of the proposed model (MAE of 102.6 HU) is comparable, even superior to that in previous brain tumor studies. A direct comparison between the atlas and CNN-based pCT generation was performed by Han et al. (13), and the MAE of the CNN-based method was 10.26% lower than that of the atlas-based method, demonstrating the advantage of CNN-based pCT generation.

T1w MRI was used in this study, whereas a previous study on HN pCT generation (17) adopted T2w MRI for pCT generation using U-Net, and good pCT quality (MAE of 131 HU) was achieved. A dosimetric comparison was performed on a single patient regarding DVH metrics, and the difference between the minimum-dose-of-98%-volume (D98%) of high-risk, intermediate-risk, and low-risk PTVs on true CT and pCT was less than 1%. Herein, we further evaluated dosimetric accuracy in terms of global dose distribution consistency and OAR dose-volume metrics statistically. Combining our results and their study, regardless of the different network architecture, the performance of deep CNN on pCT generation from the two routine clinical MR modalities (i.e., T1w and T2w) prove to be promising, especially in clinical practice when it comes to MR-linac based adaptive RT, where both T1w and T2w MRI are possible to be adopted for online planning for each individual patient. Besides, although GAN is supposed to have a strong nonlinearity modeling ability (17), our results show that there is no significant difference in the performance of pix2pix GAN and U-Net. The training process of GAN can be improved *via* some sophisticated strategies, such as using other kinds of activation functions, cost functions, normalization, or optimizers, but this is beyond the scope of this study; thus, there is a need for further studies.

As for the computational efficiency of the proposed model, the average pCT generation time is 7.9 s using GPU acceleration, in contrast to several minutes or a few hours in the aforementioned atlas-based studies. The speed advantage of CNN is more important in MR-linac based radiotherapy, where online adaptive RT planning is needed.

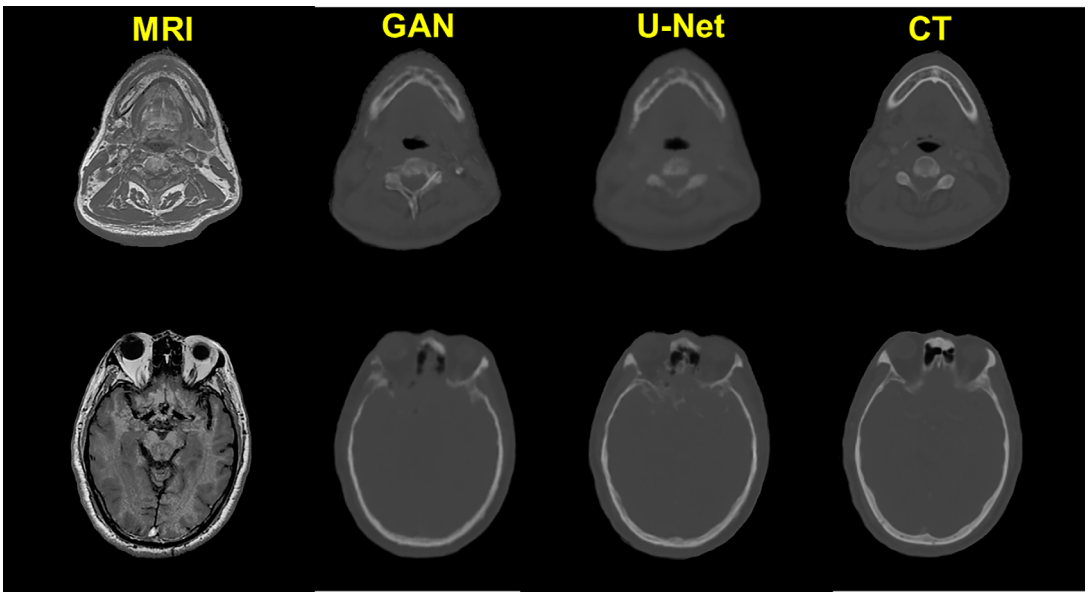


FIGURE 2 | Comparison of the prediction results of U-Net and pix2pix GAN on two exemplary slices.

For the key factors for CNN-based MRI-pCT generative model training, Andres et al. (21) evaluated the influence of training set size, MR sequence (T1w or contrast-enhanced T1w), MRI standardization approach, bias field correction, and architecture of CNN on the quality of brain pCT generation. They found that larger training set sizes result in higher pCT quality, whereas the other factors have no significant effect on the dosimetry quality, and all the candidate methods are relevant for potential use in clinical practice. The best MAE obtained using the slightly optimized preprocessing method was 78 ± 22 HU. Regarding the network architectures, 3%/3-mm gamma indices of 99.83% 0.19% and 99.74% 0.24% were obtained for HighRes-Net and 3D U-Net, respectively. Largent et al. (14) evaluated and compared U-Net and GAN using various loss functions (L2, single-scale perceptual loss (PL), multiscale PL, weighted multiscale PL), and patch-based method (PBM) based on T2w MRIs in prostate cancer. They found that GAN L2 and U-Net L2 show a lower MAE (≤34.4 HU) than U-Net PL, GAN PL, and PBM. The gamma pass rates were greater than 99% for all DLMs. GAN L2 and U-Net L2 provided the lowest dose uncertainties together with a low computation time. Their results show that the performance of U-Net and GAN is similar for pCT generation, which is consistent with our findings, although their study was for prostate pCT generation.

Besides conventional MRI modalities, such as T1w and T2w, other sequences were used to generate pCT. Many of the previous studies adopted the ultra-short time echo sequence, which could make the segmentation of bone and air easier, for brain pCT generation (22–26). However, MAEs in these studies ranged from 130 to 165 HU, which are not better than that of this study or the

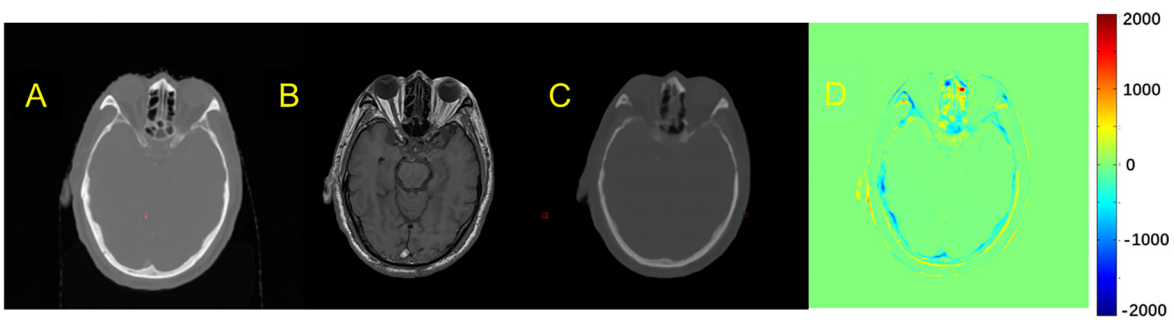


FIGURE 3 | Comparison example of pCT and original CT: **(A)** Original CT images; **(B)** T1-weighted MR images; **(C)** Predicted pCT images; **(D)** Difference between the real CT and predicted pCT values, where MAE is 73.1 HU.

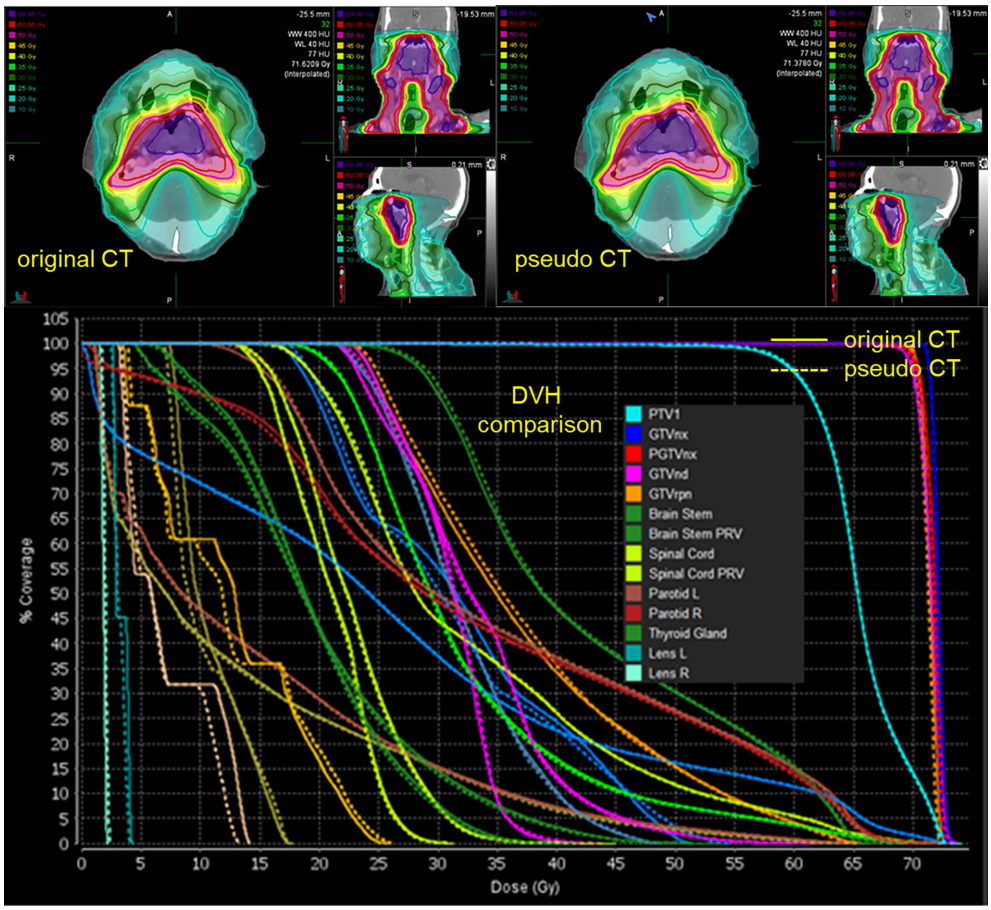


FIGURE 4 | Spatial dose distributions of the original CT (left-up panel) and pCT (down-left panel) with identical beam assignment and their DVH comparison. The solid lines in the DVH correspond to the original CT, and the dotted lines correspond to pCT.

aforementioned conventional modality-based methods. Meanwhile, such dedicated MRI modalities result in additional scanning time, which is not conducive to their clinical applications.

The limitation of this study lies in the relatively low CT number and dose prediction accuracy at the surface of the patient. As mentioned in *Results* section, this may mainly due to the imperfect MR-CT registration and the residual MR distortions, despite the bias correction. A dedicated phantom experiment may be needed to further validate the performance of the proposed MR-pCT generative model, where a perfect MR-CT alignment could be implemented.

This study proved that deep CNN is an important tool to solve the problem of MRI to pCT generation for HN, with high conversion accuracy and efficiency. It can be of great value to the MRI-only radiotherapy community, especially those sites equipped with MR-linacs, by greatly reducing the additional

TABLE 3 | Reference dose values and dose uncertainties for dosimetry metrics.

Dosimetry metrics	**PGTV D_{99Gy} (Gy)**	**PTV $V_{95\%}$ (%)**	**Lens D_{max} (Gy)**
Reference value	69.73 ± 0.44	98.74 ± 0.39	4.22 ± 1.58
Dose uncertainty (relative value)	0.26 ± 0.10 (0.4% ± 0.2%)	0.1 ± 0.1	0.26 ± 0.20 (6.1% ± 4.6%)
Dosimetry metrics	**Spinal Cord D_{max} (Gy)**	**Brain Stem D_{max} (Gy)**	**Parotid V_{30Gy} (%)**
Reference value	32.20 ± 2.61	44.42 ± 6.48	52.76 ± 4.67
Dose uncertainty (relative value)	0.52 ± 0.51 (1.6% ± 1.5%)	0.68 ± 0.34 (1.6% ± 0.9%)	0.20 ± 0.17

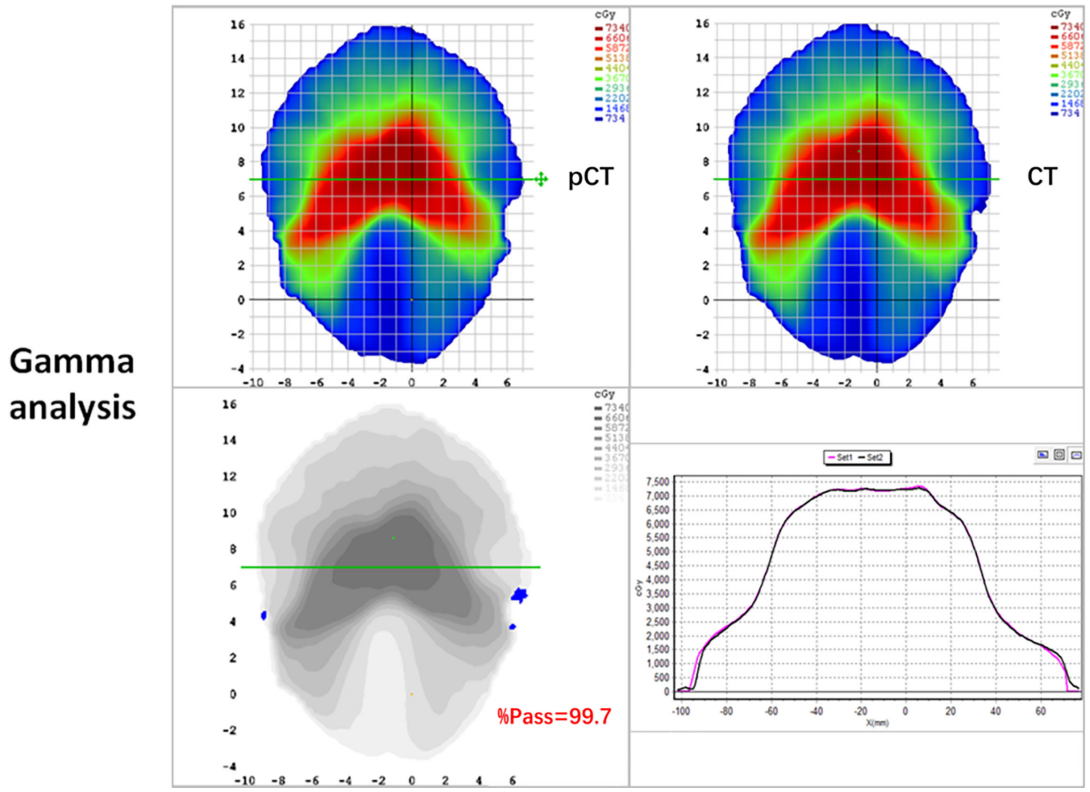


FIGURE 5 | 2D gamma analysis of pCT (up-left panel) and original CT dose distributions (up-right panel). The gamma pass rate of the slice is 99.7% (down-left panel), and the dose profiles are in good agreement in the high dose range (down-right panel).

imaging dose to patients and by ensuring the accuracy of delineation and dose calculation for each fraction.

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

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

Concept and design of study: XC, KM, and JD. Acquisition of data: XC and JL. Analysis and interpretation of data: XM and XC. Drafting the manuscript: XM and XC. Revising the manuscript critically for important intellectual content: XM, XC, YW, KM, and JD. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Natural Science Foundation of China (11975313, 12005302), the Beijing Nova Program (Z201100006820058), the Beijing Municipal Science & Technology Commission (Z181100001918002), and the Beijing Hope Run Special Fund of Cancer Foundation of China (LC2019B06, LC2018A14).

ACKNOWLEDGMENTS

The authors sincerely thank Dr. Weijie Cui, Dr. Junjie Miao and Dr. Wenlong Xia for technical assistance. We also thank the radiation oncologists in our department for delineation.

REFERENCES

1. Ramsey CR, Arwood D, Scaperoth D, Oliver AL. Clinical Application of Digitally-Reconstructed Radiographs Generated From Magnetic Resonance Imaging for Intracranial Lesions. *Int J Radiat Oncol Biol Phys* (1999) 45 (3):797–802. doi: 10.1016/S0360-3016(99)00173-X
2. Lee YK, Bollet M, Charles-Edwards G, Flower MA, Leach MO, McNair H, et al. Radiotherapy Treatment Planning of Prostate Cancer Using Magnetic Resonance Imaging Alone. *Radiother Oncol* (2003) 66(2):203–16. doi: 10.1016/S0167-8140(02)00440-1
3. Andreasen D, Van Leemput K, Hansen RH, Andersen JA, Edmund JM. Patch-Based Generation of a Pseudo CT From Conventional MRI Sequences for MRI-Only Radiotherapy of the Brain. *Med Phys* (2015) 42(4):1596–605. doi: 10.1118/1.4914158
4. Dowling JA, Lambert J, Parker J, Salvado O, Fripp J, Capp A, et al. An Atlas-Based Electron Density Mapping Method for Magnetic Resonance Imaging (MRI)-Alone Treatment Planning and Adaptive MRI-Based Prostate Radiation Therapy. *Int J Radiat Oncol Biol Phys* (2012) 83(1):e5–11. doi: 10.1016/j.ijrobp.2011.11.056
5. Dowling JA, Sun J, Pichler P, Rivest-Henault D, Ghose S, Richardson H, et al. Automatic Substitute Computed Tomography Generation and Contouring for Magnetic Resonance Imaging (MRI)-Alone External Beam Radiation Therapy From Standard MRI Sequences. *Int J Radiat Oncol Biol Phys* (2015) 93(5):1144–53. doi: 10.1016/j.ijrobp.2015.08.045
6. Edmund JM, Andreasen D, Mahmood F, Van Leemput K. Cone Beam Computed Tomography Guided Treatment Delivery and Planning Verification for Magnetic Resonance Imaging Only Radiotherapy of the Brain. *Acta Oncol* (2015) 54(9):1496–500. doi: 10.3109/0284186X.2015.1062546
7. Greer PB, Dowling JA, Lambert JA, Fripp J, Parker J, Denham JW, et al. A Magnetic Resonance Imaging-Based Workflow for Planning Radiation Therapy for Prostate Cancer. *Med J Aust* (2011) 194(4):S24–7. doi: 10.5694/j.1326-5377.2011.tb02939.x
8. Sjolund J, Forsberg D, Andersson M, Knutsson H. Generating Patient Specific Pseudo-CT of the Head From MR Using Atlas-Based Regression. *Phys Med Biol* (2015) 60(2):825–39. doi: 10.1088/0031-9155/60/2/825
9. Kim J, Glide-Hurst C, Doemer A, Wen N, Movsas B, Chetty IJ. Implementation of a Novel Algorithm for Generating Synthetic CT Images From Magnetic Resonance Imaging Data Sets for Prostate Cancer Radiation Therapy. *Int J Radiat Oncol Biol Phys* (2015) 91(1):39–47. doi: 10.1016/j.ijrobp.2014.09.015
10. Koivula L, Wee L, Korhonen J. Feasibility of MRI-Only Treatment Planning for Proton Therapy in Brain and Prostate Cancers: Dose Calculation Accuracy in Substitute CT Images. *Med Phys* (2016) 43(8):4634. doi: 10.1118/1.4958677
11. Korhonen J, Kapanen M, Keyrilainen J, Seppala T, Tenhunen M. A Dual Model HU Conversion From MRI Intensity Values Within and Outside of Bone Segment for MRI-Based Radiotherapy Treatment Planning of Prostate Cancer. *Med Phys* (2014) 41(1):011704. doi: 10.1118/1.4842575
12. Yu H, Caldwell C, Balogh J, Mah K. Toward Magnetic Resonance-Only Simulation: Segmentation of Bone in MR for Radiation Therapy Verification of the Head. *Int J Radiat Oncol Biol Phys* (2014) 89(3):649–57. doi: 10.1016/j.ijrobp.2014.03.028
13. Han X. MR-Based Synthetic CT Generation Using a Deep Convolutional Neural Network Method. *Med Phys* (2017) 44(4):1408–19. doi: 10.1002/mp.12155
14. Largent A, Barateau A, Nunes JC, Mylona E, Castelli J, Lafond C, et al. Comparison of Deep Learning-Based and Patch-Based Methods for Pseudo-CT Generation in MRI-Based Prostate Dose Planning. *Int J Radiat Oncol Biol Phys* (2019) 105(5):1137–50. doi: 10.1016/j.ijrobp.2019.08.049
15. Wang T, Manohar N, Lei Y, Dhabaan A, Shu HK, Liu T, et al. MRI-Based Treatment Planning for Brain Stereotactic Radiosurgery: Dosimetric Validation of a Learning-Based Pseudo-CT Generation Method. *Med Dosim* (2019) 44(3):199–204. doi: 10.1016/j.meddos.2018.06.008
16. Liu F, Yadav P, Baschnagel AM, McMillan AB. MR-Based Treatment Planning in Radiation Therapy Using a Deep Learning Approach. *J Appl Clin Med Phys* (2019) 20(3):105–14. doi: 10.1002/acm2.12554
17. Wang Y, Liu C, Zhang X, Deng W. Synthetic CT Generation Based on T2 Weighted MRI of Nasopharyngeal Carcinoma (NPC) Using a Deep Convolutional Neural Network (DCNN). *Front Oncol* (2019) 9:1333. doi: 10.3389/fonc.2019.01333
18. Isola P, Zhu JY, Zhou TH, Efros AA. (2017). Image-to-Image Translation with Conditional Adversarial Networks. In: *30th Ieee Conference on Computer Vision and Pattern Recognition (Cvpr 2017)*. pp. 5967–76.
19. Hofmann M, Steinke F, Scheel V, Charpiat G, Farquhar J, Aschoff P, et al. MRI-Based Attenuation Correction for PET/MRI: A Novel Approach Combining Pattern Recognition and Atlas Registration. *J Nucl Med* (2008) 49(11):1875–83. doi: 10.2967/jnumed.107.049353
20. Gudur MS, Hara W, Le QT, Wang L, Xing L, Li R. A Unifying Probabilistic Bayesian Approach to Derive Electron Density From MRI for Radiation Therapy Treatment Planning. *Phys Med Biol* (2014) 59(21):6595–606. doi: 10.1088/0031-9155/59/21/6595
21. Alvarez Andres E, Fidon L, Vakalopoulou M, Lerousseau M, Carre A, Sun R, et al. Dosimetry-Driven Quality Measure of Brain Pseudo Computed Tomography Generated From Deep Learning for MRI-Only Radiation Therapy Treatment Planning. *Int J Radiat Oncol Biol Phys* (2020) 108 (3):813–23. doi: 10.1016/j.ijrobp.2020.05.006
22. Johansson A, Garpebring A, Karlsson M, Asklund T, Nyholm T. Improved Quality of Computed Tomography Substitute Derived From Magnetic Resonance (MR) Data by Incorporation of Spatial Information–Potential Application for MR-Only Radiotherapy and Attenuation Correction in Positron Emission Tomography. *Acta Oncol* (2013) 52(7):1369–73. doi: 10.3109/0284186X.2013.819119
23. Rank CM, Hunemohr N, Nagel AM, Rothke MC, Jakel O, Greilich S. MRI-Based Simulation of Treatment Plans for Ion Radiotherapy in the Brain Region. *Radiother Oncol* (2013) 109(3):414–8. doi: 10.1016/j.radonc.2013.10.034
24. Zheng W, Kim JP, Kadbi M, Movsas B, Chetty IJ, Glide-Hurst CK. Magnetic Resonance-Based Automatic Air Segmentation for Generation of Synthetic Computed Tomography Scans in the Head Region. *Int J Radiat Oncol Biol Phys* (2015) 93(3):497–506. doi: 10.1016/j.ijrobp.2015.07.001
25. Price RG, Kim JP, Zheng W, Chetty IJ, Glide-Hurst C. Image Guided Radiation Therapy Using Synthetic Computed Tomography Images in Brain Cancer. *Int J Radiat Oncol Biol Phys* (2016) 95(4):1281–9. doi: 10.1016/j.ijrobp.2016.03.002
26. Hsu SH, Cao Y, Huang K, Feng M, Balter JM. Investigation of a Method for Generating Synthetic CT Models From MRI Scans of the Head and Neck for Radiation Therapy. *Phys Med Biol* (2013) 58(23):8419–35. doi: 10.1088/0031-9155/58/23/8419

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: 14 September 2021
doi: 10.3389/fonc.2021.707464

Beam Angle Optimization for Double-Scattering Proton Delivery Technique Using an Eclipse Application Programming Interface and Convolutional Neural Network

Wonjoong Cheon, Sang Hee Ahn, Seonghoon Jeong, Se Byeong Lee, Dongho Shin, Young Kyung Lim, Jong Hwi Jeong, Sang Hee Youn, Sung Uk Lee, Sung Ho Moon, Tae Hyun Kim and Haksoo Kim *

Proton Therapy Center, National Cancer Center, Goyang-si, South Korea

To automatically identify optimal beam angles for proton therapy configured with the double-scattering delivery technique, a beam angle optimization method based on a convolutional neural network (BAODS-Net) is proposed. Fifty liver plans were used for training in BAODS-Net. To generate a sequence of input data, 25 rays on the eye view of the beam were determined per angle. Each ray collects nine features, including the normalized Hounsfield unit and the position information of eight structures per 2° of gantry angle. The outputs are a set of beam angle ranking scores (S_{beam}) ranging from 0° to 359°, with a step size of 1°. Based on these input and output designs, BAODS-Net consists of eight convolution layers and four fully connected layers. To evaluate the plan qualities of deep-learning, equi-spaced, and clinical plans, we compared the performances of three types of loss functions and performed *K*-fold cross-validation (K = 5). For statistical analysis, the volumes V_{27Gy} and V_{30Gy} as well as the mean, minimum, and maximum doses were calculated for organs-at-risk by using a paired-samples *t*-test. As a result, smooth-L1 loss showed the best optimization performance. At the end of the training procedure, the mean squared errors between the reference and predicted S_{beam} were 0.031, 0.011, and 0.004 for L1, L2, and smooth-L1 loss, respectively. In terms of the plan quality, statistically, $Plan_{BAO}$ has no significant difference from $Plan_{Clinic}$ (P >.05). In our test, a deep-learning based beam angle optimization method for proton double-scattering treatments was developed and verified. Using Eclipse API and BAODS-Net, a plan with clinically acceptable quality was created within 5 min.

Keywords: deep-learning, convolutional neural network, beam angle optimization, proton therapy, double-scattering technique

OPEN ACCESS

Edited by:
Jose Eduardo Villarreal Barajas, Royal Devon and Exeter Hospital, United Kingdom

Reviewed by:
Weiwei Zong, Henry Ford Health System, United States
Giuseppe Magro, National Center of Oncological Hadrontherapy, Italy

***Correspondence:**
Haksoo Kim
haksoo.kim@ncc.re.kr

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

Received: *10 May 2021*
Accepted: *16 August 2021*
Published: *14 September 2021*

Citation:
Cheon W, Ahn SH, Jeong S, Lee SB, Shin D, Lim YK, Jeong JH, Youn SH, Lee SU, Moon SH, Kim TH and Kim H (2021) Beam Angle Optimization for Double-Scattering Proton Delivery Technique Using an Eclipse Application Programming Interface and Convolutional Neural Network. Front. Oncol. 11:707464. doi: 10.3389/fonc.2021.707464

INTRODUCTION

Interest in beam angle optimization (BAO) research has been on the rise recently again. When intensity-modulated radiotherapy (IMRT) emerged as a novel treatment method, BAO research was being actively undertaken. IMRT could achieve high dose conformity while minimizing undesirable dose to organs-at-risk (OARs). However, the conventional BAO process for the IMRT plan is based on trial-and-error searching by a planner; the optimal beam angle is affected by the experience and understanding of the treatment planning system (TPS) of the planner (1). Thus, various studies for BAO have been conducted to reduce the workload of treatment planning and decrease the planning time. These BAO studies incorporated techniques such as simulated annealing (2–9), geometric information scoring (10–19), gradient descent (20–25), genetic algorithms (26–29), and neural networks (30–34). However, the advent of volumetric-modulated arc therapy and the templatization of the radiation treatment plan, including dose prescription and gantry angles, have reduced interest in BAO research for X-ray therapy.

Recently, with increasing interest in proton and heavy ion therapy, which rely on the characteristics of a Bragg peak and a relatively higher radiation biological effect than X-ray therapy, several recent studies on intensity-modulated ion therapy (35) in BAO research have been published (1, 36–39).

The present study was inspired by two previous studies on BAO. In 1999, Hosseini-Ashrafi et al. conducted a study on the BAO of X-ray therapy in which they used an artificial neural network (ANN) (30). In that study, the radiation treatment plans were divided into several templates, and the ANN classified the test data according to the template. The ANN consisted of three layers of a multi-layer perceptron. The input contained 12 pre-calculated features, which were the body contour outline, treatment volume, sensitive organs, and border of tissue inhomogeneity for each case. The output contained three types of binary data for eight classification tasks. The ANN was validated using the leave-one-out method, which showed the feasibility of applying ANNs to the BAO problem.

In 2002, Pugachev et al. published a research paper on BAO for IMRT (4). They proposed beam's eye view dosimetrics (BEVD) to overcome low computation speed, which is a disadvantage of the simulated algorithm. The BEVD score was calculated by using the geometric and dosimetric information of the patient. This score was used for ranking information and as a prescreening tool to optimize the beam orientation by using a simulated annealing-based BAO algorithm. The treatment plans generated with the guidance of the BEVD score were compared with those created with five equiangular-spaced beams. They validated the feasibility of the BEVD score for the BAO problem. The BEVD guidance indicated that the computational efficiency increased by a factor of ~10.

In the current study, we developed a deep-learning based BAO method for the three-ports proton double-scattering (DS) technique using the geometric information of the patient computed tomography (CT) anatomy and Hounsfield unit (HU) data as well as a convolutional neural network (CNN). In particular, the DS technique was used for large-field proton therapy. The proposed method requires only geometric information without a fluence optimization process. The geometric information is automatically extracted using an application programming interface (API) of a TPS (Eclipse, Varian, Palo Alto, CA, USA). A set of beam angle ranking scores (S_{beam}) for all angles is predicted using the deep-learning model. The quality of the treatment plans created with the predicted S_{beam} guidance was statistically compared for equi-spaced and clinical plans. The evaluation was performed using the dose–volume histogram (DVH) parameters.

MATERIALS AND METHODS

Patient Database

Patient data from 50 liver cases, consisting of average intensity projection (AIP) CT images calculated from the four-dimensional (4D) CT images of 40–60% phases, a digital image communication in medicine-radiation therapy (DICOM-RT) structure file, and a DICOM-RT plan file, were used in this study. The patients were originally treated using the proton DS delivery technique configured with three DS fields at the National Cancer Centre in the Republic of Korea (40).

To automatically access information of interesting structures in the TPS, an in-house software was developed using the Eclipse script API. The eight structures of interest included the body, total liver volume (TLV), primary gross tumor volume (PGTV), duodenum, stomach, esophagus, heart, and spleen. The body contour included areas such as immobilizers that should be considered for dose calculation.

Geometric Information and a Set of Beam Angle Ranking Scores for Beam Angle Optimization: Input and Output of the Deep-Learning Model

To train a beam angle optimization network (BAODS-Net), geometric information extracted from the AIP CT images and DICOM-RT structures (an input of BAODS-Net) and S_{beam} generated from RT-plan (an output of BAODS-Net) were used.

The geometric information was collected by the ray tracing method (4). A ray, which is a collector, was determined to penetrate from the body contour to the isocenter, and the path of the ray was tracked in a 3D treatment room coordinate system. The ray collected geometric information by dividing the length of 1,000 mm into 4,000 bins (**Figure 1A**). The collected data included the following: specifically HU from the AIP CT images as a double data type and the anatomy position information of interest of eight structures as a binary data type (**Figure 1B**). Thus, the shape of geometric information extracted from a collector was 9 × 4,000. Geometric information collected by the penetrating ray is useful data to evaluate the best DS field that considers range uncertainties and OAR positions. In this regard, additional 24 parallel rays penetrating the target volume were created. In detail, the positions of the 24 rays were automatically determined at an isometric angle on two ellipses

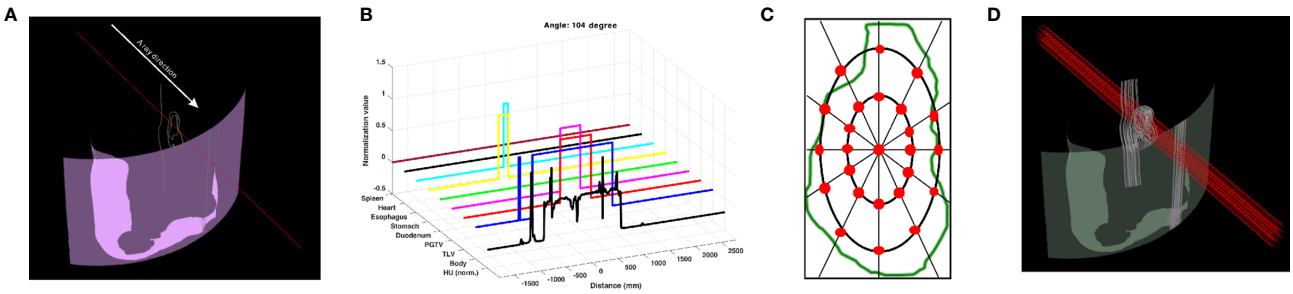


FIGURE 1 | Data collector of patient CT anatomy and Hounsfield unit (HU) data used as input to BAODS-Net. **(A)** Data collector in 3D. **(B)** Extracted patient anatomy and HU data. **(C)** Twenty-five data collectors in the cross-section of the primary gross tumor volume at the isocenter plane. **(D)** Final data collectors consisting of 25 rays per angle.

with different radii placed on the PGTV cross-section in the isocenter plane observed from the beam's eye view (**Figure 1C**). Thus, at a specific gantry angle, 25 data collectors extracted nine geometric information points (**Figure 1D**). In the same manner, the patient CT anatomy and HU data for all directions were collected at angles from 0° to 358° in steps of 2°. Finally, the shape of the geometric information extracted by the 25 data collectors on coplanar was 40,500 × 4,000. The geometric information was reshaped to input 4D tensor (batch, channel, height, and width). For pre-processing, only the HU values were normalized by using Z-score normalization per patient. The Z-score normalization is a standardization method for the fast convergence of deep-learning models. For the case of the anatomy position information of eight structures, normalization was not performed.

A S_{beam} ground-truth for each patient was generated using the gantry angle information in the DICOM-RT plan files with steps of 1°. The S_{beam} comprised one-dimensional (1D) data continuously ranging from 0.0 to 1.0. The gantry angles used in the clinic were assigned a value of 1.0; otherwise, a value of 0.0 was assigned. Then, to induce effective optimization of BAODS-Net, a normalized Gaussian filter was applied. Finally, the size of S_{beam} was 360, and the shape of S_{beam} was reshaped to a batch-considered shape. An example of a reference S_{beam} is shown in **Figure 2**.

K-fold cross-validation (CV) could provide a better indication of how well the BAODS-Net was universalized to unobserved

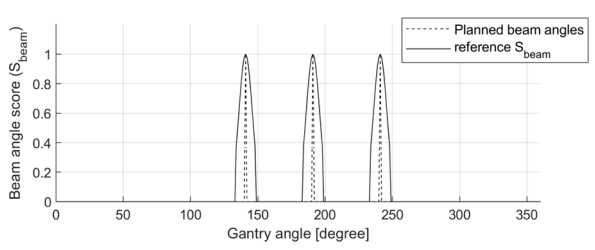


FIGURE 2 | Sample of planned beam angles (140°, 190°, and 240°) and a reference set of beam angle ranking scores (S_{beam}) for BAODS-Net.

data. We performed a patient-wise K-fold CV method (K = 5), and all datasets were divided into five disjointed and identically sized subsets (41).

Double-Scattering Beam Angle Optimization Network

In this paper, the BAODS-Net based on a CNN is proposed as shown in **Figure 3**. It consists of two main stages: a feature extractor and a predictor. The feature extractor was configured with eight convolution layers to extract distinguishable features of the geometric information by applying a convolution layer with various strides. The first convolution layer was designed with dimensions of 1 × 9 for the width and height, respectively. The convolution layer was operated with a stride size of nine and with the same padding option. This is because the nine geometric features extracted by a ray are intended to be integrated into a weighted geometric feature. The second convolution layer was designed with dimensions of 1 × 25 for the width and height, respectively. A stride size of 25 was used to integrate each weighted geometric feature extracted from the 25 rays into one weighted ray representing a specific angle. The remaining part of the feature extractor was designed with six convolution layers (width: 3, height: 3, stride: 1) and max-pooling layers. The output of the feature extractor was flattened and then passed to the predictor. The predictor was designed with four fully connected (FC) layers to continuously predict S_{beam} from 0° to 359°. Although the FC layer is computationally expensive, it can effectively predict S_{beam} with a non-linear activation function because the FC layer has a structure agnostic property.

Batch normalization was applied for fast convergence in the optimization process and to ensure the robustness of the performance (42). Randomly initialized biases were added for all layers. Additionally, the activation function was the leaky rectified linear unit (43), which was used to maintain the contribution of negative data, and the adaptive momentum estimation optimizer was employed (44). The BAODS-Net trained for 5,000 epochs with a learning rate of 0.001 and weight decay of 0.0002, and to compute the running average of the gradient, β_1 was 0.9 and β_2 was 0.999. The output shape and parameters of the layers composing the BAODS-Net are summarized in **Table 1**.

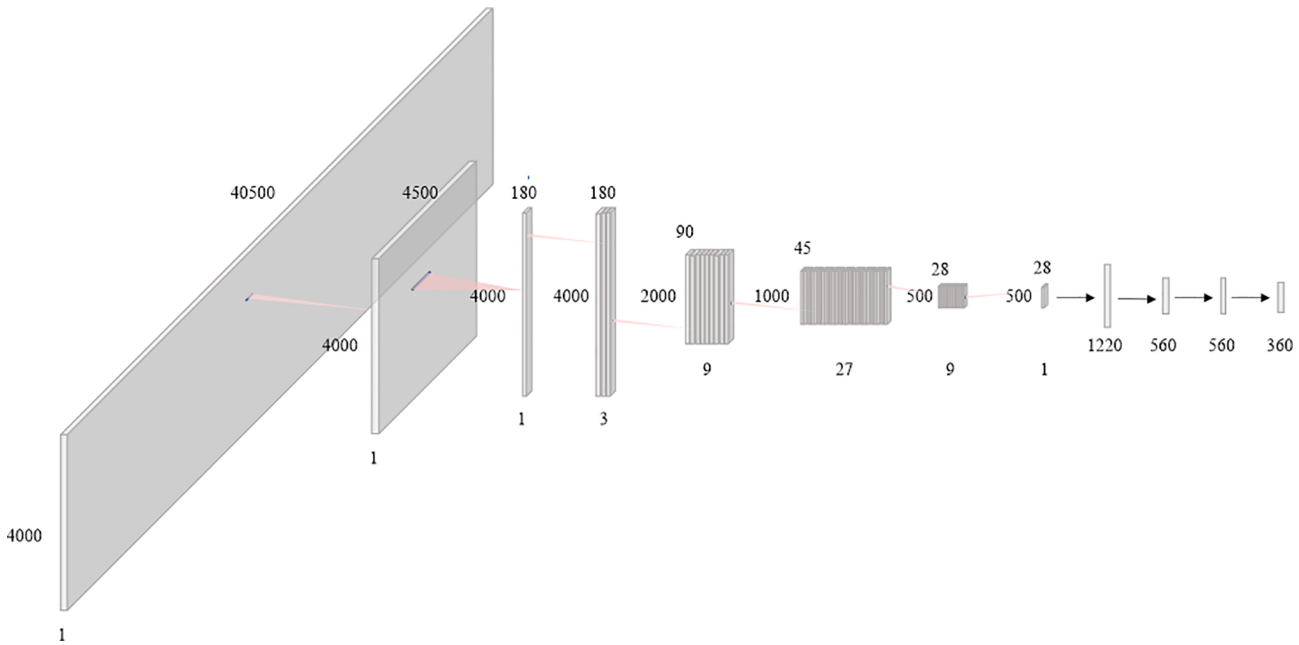


FIGURE 3 | Diagram of the proposed beam angle optimization for proton double-scattering network (BAODS-Net).

TABLE 1 | The architecture of the proposed BAODS-Net.

	Layer	Output Shape (B, C, W, H)	Kerner Size
Feature extractor	Conv2d	(None, 1, 4500, 4000)	(1, 1, 9, 1)
	BatchNorm2d	(None, 1, 4500, 4000)	2
	Conv2d	(None, 1, 180, 4000)	(1, 1, 25, 1)
	BatchNorm2d	(None, 1, 180, 4000)	2
	Conv2d	(None, 3, 178, 3998)	(3, 1, 3, 3)
	BatchNorm2d	(None, 3, 178, 3998)	6
	MaxPool2d	(None, 3, 89, 1999)	0
	Conv2d	(None, 9, 87, 1997)	(9, 3, 3, 3)
	BatchNorm2d	(None, 9, 87, 1997)	18
	MaxPool2d	(None, 9, 43, 998)	0
	Conv2d	(None, 27, 41, 996)	(27, 9, 3, 3)
	BatchNorm2d	(None, 27, 41, 996)	54
	MaxPool2d	(None, 27, 20, 498)	0
	Conv2d	(None, 9, 18, 496)	(9, 27, 3, 3)
	BatchNorm2d	(None, 9, 18, 496)	18
	MaxPool2d	(None, 9, 9, 248)	0
	Conv2d	(None, 3, 7, 246)	(3, 9, 3, 3)
	BatchNorm2d	(None, 3, 7, 246)	6
	Conv2d	(None, 1, 5, 244)	(1, 3, 3, 3)
	BatchNorm2d	(None, 1, 5, 244)	2
Predictor	Linear	(None, 560)	(560, 1220)
	BatchNorm1d	(None, 560)	1,120
	Linear	(None, 560)	(560, 560)
	BatchNorm1d	(None, 560)	1,120
	Linear	(None, 560)	(560, 560)
	BatchNorm1d	(None, 560)	1,120
	Linear	(None, 360)	(360, 560)

B, batch size; C, channel; W, width; H, height; Conv2d, two-dimensional (2D) convolution layer; BatchNorm2d, 2D batch normalization layer; MaxPool2d, 2D Max pooling layer; Linear, fully connected layer; BatchNorm1d, one-dimensional batch normalization layer.

Training and Validation of BAODS-Net

In the training process, BAODS-Net was optimized to predict S_{beam} by using the training data. To find the training loss function that could achieve the best BAODS-Net performance, we compared the performances of the three types of loss functions: L1 loss (Eq. 1), L2 loss (Eq. 2), and smooth-L1 loss (Eq. 3). The prediction accuracy for S_{beam} was calculated by using mean squared error (MSE) between the reference and predicted S_{beam}.

$$\mathbf{L1\ loss}(x, y) = \frac{\mathbf{1}}{n}\sum_i \mathbf{Z}_i, \mathbf{Z}_i = |\mathbf{x_i} - \mathbf{y_i}|, \tag{1}$$

$$\mathbf{L2\ loss}(x, y) = \frac{\mathbf{1}}{n}\sum_i \mathbf{Z}_i, \mathbf{Z_i} = (\mathbf{x}_i - \mathbf{y}_i)^2, \tag{2}$$

$$\mathbf{Smooth\ L1\ loss}(x, y) = \frac{\mathbf{1}}{n}\sum_i \mathbf{Z_i}, \mathbf{Z}_i = \begin{cases} \mathbf{0.5}\frac{(x_i - y_i)^2}{\beta}, \mathbf{if}\,|\mathbf{x_i} - \mathbf{y_i}| < \boldsymbol{\beta} \\ |\mathbf{x}_i - \mathbf{y}_i| - \mathbf{0.5}\boldsymbol{\beta}, \ \textit{otherwise} \end{cases} \tag{3}$$

where (x, y) is the reference S_{beam} and predicted S_{beam}, respectively, and n is the number of samples. The hyper-parameter beta (β) in Eq. 3 was a value for applying additional weight to the loss. β was empirically determined to be 0.5.

In this experiment, the data of the first fold were used. Specifically, the smooth-L1 loss could be interpreted as a combination of L1 and L2 losses.

To improve the BAODS-Net performance and reduce its generalization error, a 1D augmentation technique, which is a 1D data translation ranging from −2 to 2°, was applied to the reference S_{beam}. The augmentation data were randomly generated for each epoch. Through the K-fold CV principle, we independently conducted five different runs for five separate CV datasets to evaluate the BAODS-Net performance.

Plan Creation With BAODS-Net

The procedure for creating a three-ports proton DS plan with the guidance of BAODS-Net ($Plan_{BAO}$) was as follows: (i) the patient CT anatomy and HU data were automatically extracted by using an in-house software based on Eclipse API, (ii) the patient CT anatomy and HU data were fed into the BAODS-Net, and then the BAODS-Net output (S_{beam}) was predicted; (iii) the specific angles in S_{beam} were selected according to a selection rule. The rule preferentially selected the three gantry angles corresponding to the highest score. However, if the interval was less than 30°, the next priority angle was selected; (iv) the collimator and compensator were designed using the default TPS option without manual modification for objective evaluation of BAODS-Net performance, and (v) the field weight was set to one for all fields.

Plan Comparison of BAODS-Net, Equi-Spaced Angle, and Clinical Plan

To validate the liver treatment plan quality for 50 patients, results were obtained by combining the results of the five folds. The DVH parameters were analyzed for $Plan_{BAO}$, the equi-spaced plan [gantry angles were fixed at 0°, 120°, 240° ($Plan_{Equi}$)], and the clinical plan ($Plan_{Clinic}$). Although the equi-spaced plan is rarely applied in clinics for proton beam by a planner, we added to the equi-spaced plan for comparative study (4). The $Plan_{Clinic}$s were created by a qualified planner with 5 years of clinical experience. The evaluation metric is defined below, and the conformity index (CI) was calculated for PGTV (Eq. 4).

$$\mathbf{Conformity\ index} = \frac{\mathbf{TV} \times \mathbf{PIV}}{\mathbf{TV_{PIV}^2}}, \tag{4}$$

where TV is the target volume and PIV is the prescribed isodose volume. The closer PIV is to TV, the closer the CI is to 1. The volumes V_{27Gy} and V_{30Gy} for TLV as well as the mean, minimum, and maximum doses for OARs were calculated. V_{xGy} represents the volume percentage of the whole organ receiving a dose ≥xGy. For statistical analysis, these results were compared with the paired-samples *t*-test. All statistical analyses were implemented using SAS 9.4 software (SAS Institute Inc., Cary, NC, USA), and the statistically significant level was set at $P = .05$.

RESULT

Performance Comparison of L1, L2, and Smooth-L1 Loss Functions

To determine the optimal loss function, BAODS-Net was trained using L1, L2, and smooth-L1 loss. The training time was approximately 120 h by using the training data of the first fold, which corresponded to 5,000 epochs when using an NVIDIA Quadro GV100 graphics processing unit (GPU) (NVIDIA, Santa Clara, CA, USA). The seed number was fixed in the training procedure. The MSE between the predicted and reference S_{beam} was evaluated when L1, L2, and smooth-L1 loss were used for each model training procedure. At 5,000 epochs, the MSEs were 0.031, 0.011, and 0.004 for L1, L2, and smooth-L1 loss, respectively. As a result, the smooth-L1 loss was adopted as a metric for the training loss.

Plan Comparison of BAOBS-Net, Equi-Spaced Angle, and Planner

To evaluate the cases of 50 patients as test data, BAODS-Net was trained and evaluated with five different folds, and the training losses in the five different runs were recorded. At the 5,000th epoch, the mean and standard deviation of MSEs between the predicted and reference S_{beam} for the five folds were 0.0037 and 0.0006, respectively.

The plans created by guidance with BAODS-Net, equi-spaced angle, and the planner method were compared using the DVH parameters. In **Figure 4**, the paired two test cases of $Plan_{BAO}$ and $Plan_{Clinic}$ are visually analyzed, including the reference, predicted S_{beam}, and 2D dose distribution at the isocenter plane. The star markers in **Figures 4A, D** are the gantry angles finally selected for $Plan_{BAO}$.

Data from a total of 37 out of 50 patients were used for comparative evaluation; 13 cases were outside the approved proton range of the TPS for using the equi-spaced angle. The errors occurred in the right posterior oblique (RPO) field of 120°.

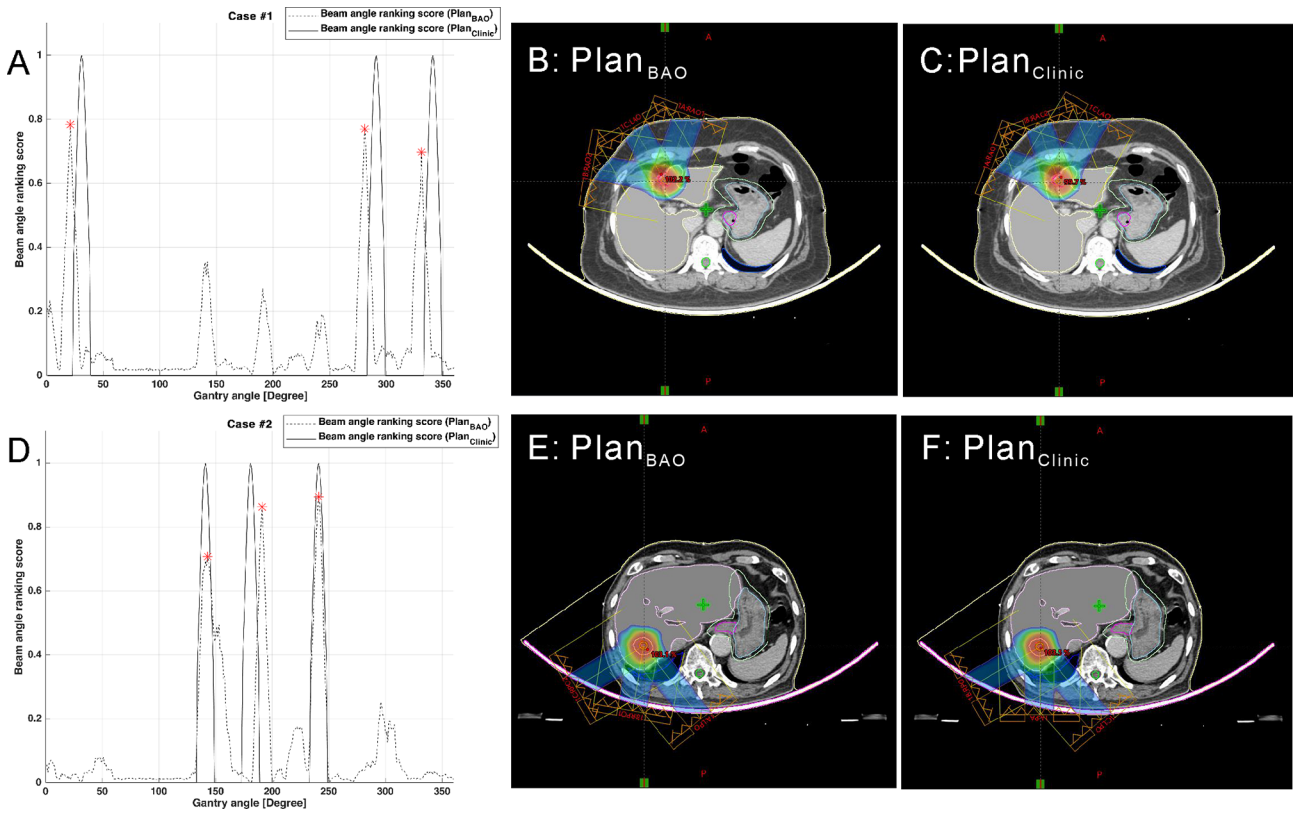


FIGURE 4 | Two test cases of reference and predicted S_{beam} and 2D dose distribution at the isocenter plane: case 1 **(A–C)** and case 2 **(D–F)**.

For the evaluation data, the mean CIs for PGTV were 1.04, 0.99, and 1.17 for BAODS-Net, equi-spaced angle, and planner, respectively. The mean (standard deviation) V_{27Gy} values for TLV were 10.3% (5.4), 10.6% (5.3), and 10.3% (5.3), respectively. The mean (standard deviation) V_{30Gy} values for TLV were 9.7% (5.1), 9.8% (5.0), and 9.7% (5.0), respectively. Although V_{27Gy} and V_{30Gy} of $Plan_{Equi}$ were higher than other plans, the $Plan_{BAO}$ had no statistically significant difference with both $Plan_{Clinic}$ (P = .94) and $Plan_{Equi}$ (P = .11).

For the mean dose of OARs, the results of statistical comparison between $Plan_{BAO}$ and $Plan_{Clinic}$ are summarized in **Table 2**. The results of statistical comparison between $Plan_{BAO}$ and $Plan_{Equi}$ are summarized in **Table 3**. As a statistical result, the mean dose has no significant differences between $Plan_{BAO}$ and $Plan_{Clinic}$ (P >.05), while the mean dose of $Plan_{Equi}$ has a significant difference with $Plan_{BAO}$ (P <.05). These results signified that $Plan_{BAO}$ is superior to $Plan_{Equi}$ and similar to $Plan_{Clinic}$ in OARs. The mean dose is visualized for each structure in **Figure 5** as a boxplot. The central mark (red) indicates the median, and the top and bottom edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers (-) extend to the most extreme data points, while not considering outliers (+). **Table 4** summarizes the average of the mean, minimum, maximum doses of OARs for the three planning methods. As a result, guidance using the BAODS-Net method may engender a plan with a quality similar to that created by the planner. In the case of the equi-spaced plan, the quality is relatively low compared to that of the clinical plan.

DISCUSSION

The conventional procedure for creating a proton DS treatment plan is time-consuming and planner dependent. BAO can be utilized as a logical step for the development of efficient and optimal proton plans, similar to studies finding optimal fields in the static IMRT planning area. To date, there is no clinically applicable commercial software for BAO or for enabling intuitive comprehension by a planner.

In this study, we designed BAODS-Net, a new deep-learning based method of BAO for proton therapy. BAODS-Net is based on a CNN and employs the patient anatomy and HU data from the DICOM-RT structure file and AIP CT images, which are automatically extracted using the Eclipse API. The output is the predicted S_{beam} as angle ranking information, which could be used as *a priori* knowledge to guide the determination of the three gantry angles used for clinical practice.

According to the study results, the proposed method produced clinically acceptable and practical plans. The time required to create a proton DS treatment plan was decreased

TABLE 2 | Comparison of mean dose for organs-at risk-between $Plan_{BAO}$ and $Plan_{Clinic}$.

	Normalized Mean Dose (%)		Mean Difference (95% CI)		*p*-value
	$Plan_{BAO}$ (*N* = 37)	$Plan_{Clinic}$ (*N* = 37)			
Total liver volume					
Mean ± SD	10.7 ± 5.3	10.7 ± 5.2	0.051	(-0.335, 0.438)	0.7891[a]
Median (min–max)	9.2 (3.7–22.9)	9.2 (3.7–21.3)			
Duodenum					
Mean ± SD	1.8 ± 6.9	1.1 ± 4.6	0.678	(-0.115, 1.472)	0.0913[a]
Median (min–max)	0 (0–40.4)	0 (0–27.4)			
Stomach					
Mean ± SD	1.3 ± 3.3	1.0 ± 3.0	0.235	(-0.430, 0.900)	0.4778[a]
Median (min–max)	0 (0–14)	0 (0–14)			
Esophagus					
Mean ± SD	1.6 ± 4.9	1.6 ± 4.1	<0.001	(-0.551, 0.551)	1.0000[a]
Median (min–max)	0 (0–27.6)	0 (0–21.1)			
Heart					
Mean ± SD	0.5 ± 1.0	0.6 ± 1.0	-0.081	(-0.254, 0.092)	0.3473[a]
Median (min–max)	0 (0–5.6)	0 (0–3.9)			
Spleen					
Mean ± SD	0.2 ± 0.8	0.5 ± 1.9	-0.346	(-0.915, 0.223)	0.2252[a]
Median (min–max)	0 (0–4.8)	0 (0–10.3)			

[a]*Paired samples t-test.*

to approximately 5 min; specifically, the predicted S_{beam} was calculated within approximately 0.2 s through BAODS-Net.

This study provides key contributions and is distinguished from recent BAO research in several ways. This study is the first to employ the BAO method of the proton DS delivery technique, and a deep-learning based one-stop solution was developed. By leveraging this solution, a planner can refer to the predicted S_{beam} in the commercial TPS using Eclipse API. In contrast, conventional BAO research required multiple steps to solve the BAO problem, specifically optimizing the fluence map and then computing the dose influence matrices. However, in our research, only the patient CT anatomy and HU data were used for BAO procedures without dosimetric information from candidate beams. Similarly, Barkousaraie et al. (34) proposed a BAO method using the art column generation (CG) method and a CNN. The architecture of their deep-learning model is based on U-net (45), and the model is trained to mimic the result of the CG algorithm by using only extracted features from the patient anatomy. Although this approach also does not directly use dosimetric information for BAO, more time and additional effort are required to obtain the CG algorithm results.

The following factors may have affected the measurement accuracy. The predicted beam could not provide an optimal

TABLE 3 | Comparison of mean dose for organs-at-risk between $Plan_{BAO}$ and $Plan_{Equi}$.

	Normalized Mean Dose (%)		Mean Difference (95% CI)		*p*-value
	$Plan_{BAO}$ (*N* = 37)	$Plan_{Equi}$ (*N* = 37)			
Total liver volume					
Mean ± SD	10.7 ± 5.3	12.9 ± 5.8	-2.195	(-2.981, -1.409)	<.0001[a]
Median (min–max)	9.2 (3.7–22.9)	11.5 (3.8–24.8)			
Duodenum					
Mean ± SD	1.8 ± 6.9	2.3 ± 6.8	-0.565	(-1.056, -0.074)	0.0253[a]
Median (min–max)	0 (0–40.4)	0 (0–39.8)			
Stomach					
Mean ± SD	1.3 ± 3.3	2.4 ± 4.2	-1.132	(-2.069, -0.196)	0.0192[a]
Median (min–max)	0 (0–14)	0 (0–18.5)			
Esophagus					
Mean ± SD	1.6 ± 4.9	6.5 ± 8	-4.868	(-7.270, -2.465)	0.0002[a]
Median (min–max)	0 (0–27.6)	3.2 (0–26.1)			
Heart					
Mean ± SD	0.5 ± 1	0.7 ± 1.2	-0.230	(-0.409, -0.050)	0.0136[a]
Median (min–max)	0 (0–5.6)	0.1 (0–4.9)			
Spleen					
Mean ± SD	0.2 ± 0.8	1.3 ± 3.4	-1.143	(-2.109, -0.178)	0.0216[a]
Median (min–max)	0 (0–4.8)	0 (0–14.4)			

[a]*Paired samples t-test.*

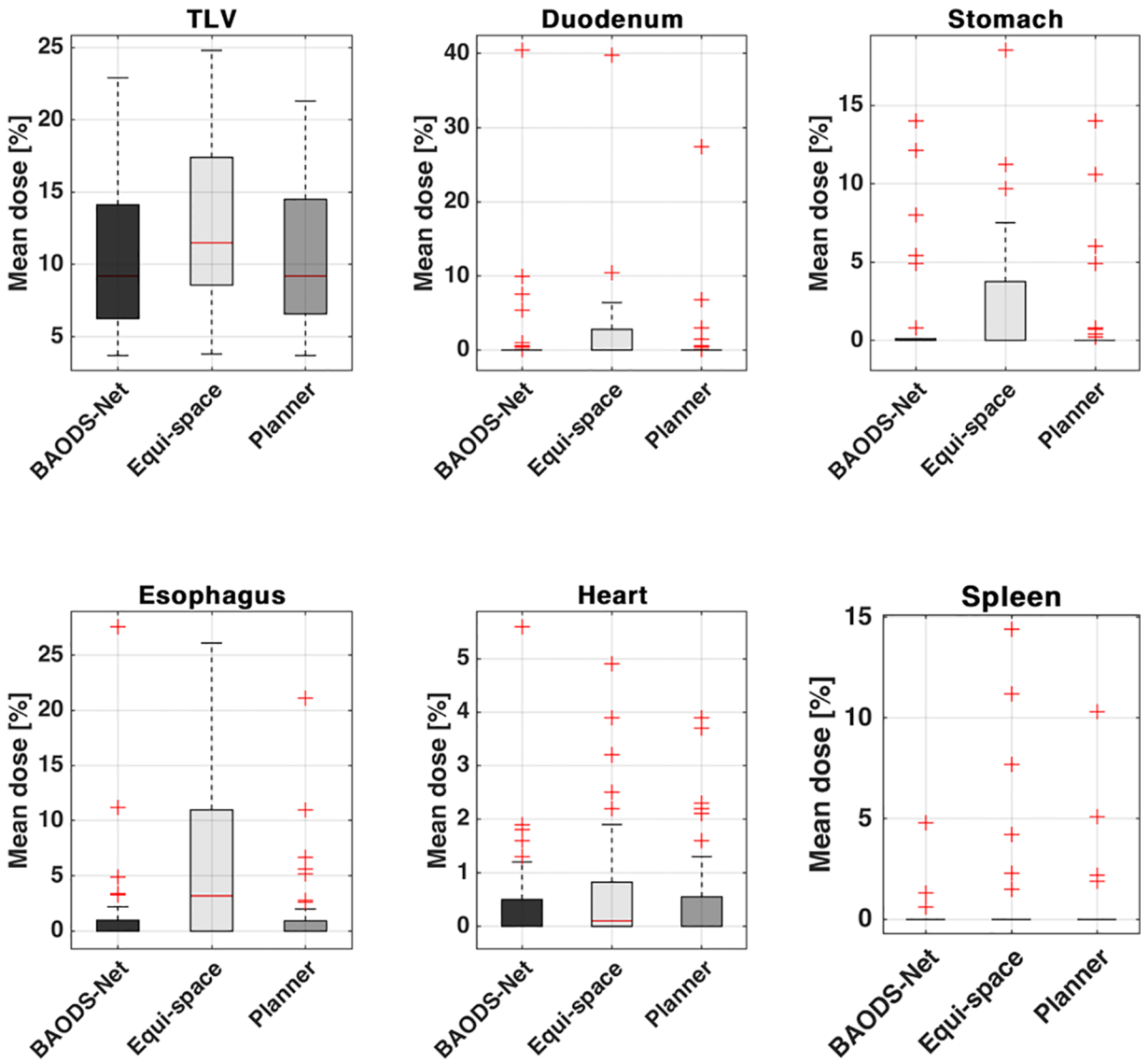


FIGURE 5 | Box plot for the average of the mean dose of each structure for plans guided by BAODS-Net, equi-spaced angle, and the planner method, respectively.

beam configuration because a reference S_{beam} was originated from $Plan_{Clinic}$. In addition, the $Plan_{BAO}$ was generated without manual modification/optimization procedure such as beam weight, collimator design, compensator design, *etc.* In other words, it means that the plan quality of $Plan_{BAO}$ has the scope for improvement. In this study, we considered only coplanar proton DS plans for liver cases. However, if the search space is expanded for a non-coplanar proton DS plan, the proposed method could be applied to a non-coplanar proton DS plan for other diseases. Meanwhile, it should be noted that the field design of $Plan_{Equi}$, specifically anterior–posterior, RPO, and left posterior oblique can be disadvantageous for liver cases. However, according to **Figure 5** and **Tables 2–4**, it can be confirmed that $Plan_{BAO}$ can create a plan of similar quality to that of $Plan_{Clinic}$ by considering the patient CT anatomy and HU data.

CONCLUSION

In this paper, we validated the feasibility of using BAODS-Net for BAO of the three-port proton DS plan. BAODS-Net only used geometric information automatically extracted through the Eclipse API and could successfully predict the S_{beam} for the planning. The results clearly showed its potential for facilitating the three-port proton DS planning. The BAODS-Net dramatically reduced the planning time and brought us one step closer to real-time adaptive proton radiotherapy. Finally, the quality of $Plan_{BAO}$ was statistically verified to be similar to that of $Plan_{Clinic}$ in the mean dose of OARs ($P > .05$)

TABLE 4 | Average of dose–volume histogram parameters of organs-at-risk for the three planning methods.

	Minimum dose (%)	Maximum dose (%)	Mean dose (%)	Conformity index
BODY (BAODS-Net)	0.00	104.78	1.78	0.00
BODY (Equi-spaced angle)	0.00	**106.42**	**2.31**	0.00
BODY (Planner)	0.00	102.24	1.80	0.00
TLV (BAODS-Net)	0.00	104.11	10.72	0.00
TLV (Equi-spaced angle)	0.00	**106.10**	**12.92**	0.00
TLV (Planner)	0.00	104.66	10.67	0.00
PGTV (BAODS-Net)	**95.28**	104.78	101.71	1.04
PGTV (Equi-spaced angle)	93.96	**106.41**	**102.83**	0.99
PGTV (Planner)	92.53	102.62	99.32	**1.17**
Duodenum (BAODS-Net)	0.00	12.91	1.76	0.00
Duodenum (Equi-spaced angle)	0.00	**16.22**	**2.33**	0.00
Duodenum (Planner)	0.00	9.55	1.09	0.00
Stomach (BAODS-Net)	0.00	16.00	1.25	0.00
Stomach (Equi-spaced angle)	0.00	**21.11**	**2.38**	0.00
Stomach (Planner)	0.00	11.76	1.02	0.00
Esophagus (BAODS-Net)	0.00	13.94	1.62	0.00
Esophagus (Equi-spaced angle)	0.00	**23.81**	**6.49**	0.00
Esophagus (Planner)	0.00	12.36	1.62	0.00
Heart (BAODS-Net)	0.00	27.00	0.48	0.00
Heart (Equi-spaced angle)	0.00	**28.04**	**0.71**	0.00
Heart (Planner)	0.00	26.80	0.56	0.00
Spleen (BAODS-Net)	0.00	2.42	0.18	0.00
Spleen (Equi-spaced angle)	0.00	**5.47**	**1.32**	0.00
Spleen (Planner)	0.00	4.94	0.53	0.00

A higher score from the comparison of the three methods is highlighted in bold for organs-at-risk. Bold values stand for having statistical significance. TLV, total liver volume; PGTV, primary gross tumor volume.

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 below: 10.6084/m9.figshare.14034143.

AUTHOR CONTRIBUTIONS

WC and HK contributed to the conception and design of the study. WC, SA, and SJ organized the database. WC, SY, and SL performed the statistical analysis. HK, SM, and TK provided guidance on methodology and the overall project. SBL, DS, YL, and JJ provided lab and technical support. WC wrote the first draft of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This research was funded by the National Research Foundation of the Korean National Cancer Centre Fund, grant number 2110610-1.

REFERENCES

1. Taasti VT, Hong L, Shim JSA, Deasy JO, Zarepisheh M. Automating Proton Treatment Planning With Beam Angle Selection Using Bayesian Optimization. *Med Phys* (2020) 47(8):3286–96. doi: 10.1002/mp.14215
2. Steadham AM, Liu HH, Crane CH, Janjan NA, Rosen II. Optimization of Beam Orientations and Weights for Coplanar Conformal Beams in Treating Pancreatic Cancer. *Med Dosim* (1999) 24:265–71. doi: 10.1016/S0958-3947(99)00028-X
3. Rowbottom CG, Oldham M, Webb S. Constrained Customization of non-Coplanar Beam Orientations in Radiotherapy of Brain Tumours. *Phys Med Biol* (1999) 44:383. doi: 10.1088/0031-9155/44/2/007
4. Pugachev A, Xing L. Incorporating Prior Knowledge Into Beam Orientation Optimization in IMRT. *Int J Radiat Onc Biol Phys* (2002) 54:1565–74. doi: 10.1016/S0360-3016(02)03917-2
5. Morrill SM, Lane RG, Jacobson G, Rosen II. Treatment Planning Optimization Using Constrained Simulated Annealing. *Phys Med Biol* (1991) 36:1341. doi: 10.1088/0031-9155/36/10/004
6. Bortfeld T, Schlegel W. Optimization of Beam Orientations in Radiation Therapy: Some Theoretical Considerations. *Phys Med Biol* (1993) 38:291. doi: 10.1088/0031-9155/38/2/006
7. Lu HM, Kooy HM, Leber ZH, Ledoux RJ. Optimized Beam Planning for Linear Accelerator-Based Stereotactic Radiosurgery. *Int J Radiat Onc Biol Phys* (1997) 39:1183–9. doi: 10.1016/S0360-3016(97)00344-1
8. Pugachev A, Li JG, Boyer AL, Hancock SL, Le QT, Donaldson SS, et al. Role of Beam Orientation Optimization in Intensity-Modulated Radiation Therapy. *Int J Radiat Onc Biol Phys* (2001) 50:551–60. doi: 10.1016/S0360-3016(01)01502-4
9. Pugachev A, Boyer A, Xing L. Beam Orientation Optimization in Intensity-Modulated Radiation Treatment Planning. *Med Phys* (2000) 27:1238–45. doi: 10.1118/1.599001
10. Cho BC, Roa WH, Robinson D, Murray B. The Development of Target-Eye-View Maps for Selection of Coplanar or Noncoplanar Beams in Conformal Radiotherapy Treatment Planning. *Med Phys* (1999) 26:2367–72. doi: 10.1118/1.598751
11. Woudstra E, Storchi P. Constrained Treatment Planning Using Sequential Beam Selection. *Phys Med Biol* (2000) 45:2133. doi: 10.1088/0031-9155/45/8/306

12. Pugachev A, Xing L. Pseudo Beam's-Eye–View as Applied to Beam Orientation Selection in Intensity-Modulated Radiation Therapy. *Int J Radiat Onc Biol Phys* (2001) 51:1361–70. doi: 10.1016/S0360-3016(01)01736-9
13. Meedt G, Alber M, Nüsslin F. Non-Coplanar Beam Direction Optimization for Intensity-Modulated Radiotherapy. *Phys Med Biol* (2003) 48:2999. doi: 10.1088/0031-9155/48/18/304
14. Das S, Cullip T, Tracton G, Chang S, Marks L, Anscher M, et al. Beam Orientation Selection for Intensity-Modulated Radiation Therapy Based on Target Equivalent Uniform Dose Maximization. *Int J Radiat Onc Biol Phys* (2003) 55:215–24. doi: 10.1016/S0360-3016(02)03817-8
15. Woudstra E, Heijmen BJ. Automated Beam Angle and Weight Selection in Radiotherapy Treatment Planning Applied to Pancreas Tumors. *Int J Radiat Onc Biol Phys* (2003) 56:878–88. doi: 10.1016/S0360-3016(03)00266-9
16. Woudstra E, Heijmen BJ, Storchi PR. Automated Selection of Beam Orientations and Segmented Intensity-Modulated Radiotherapy (Imrt) for Treatment of Oesophagus Tumors. *Radiother Onc* (2005) 77:254–61. doi: 10.1016/j.radonc.2005.06.028
17. Engel K, Tabbert E. Fast Simultaneous Angle, Wedge, and Beam Intensity Optimization in Inverse Radiotherapy Planning. *Optimiz Eng* (2005) 6:393–419. doi: 10.1007/s11081-005-2065-3
18. Meyer J, Hummel SM, Cho PS, Austin-Seymour MM, Phillips MH. Automatic Selection of non-Coplanar Beam Directions for Three-Dimensional Conformal Radiotherapy. *Br J Radiol* (2005) 78:316–27. doi: 10.1259/bjr/13015047
19. Ranganathan V, Sathiya Narayanan VK, Bhangle JR, Gupta KK, Basu S, Maiya V, et al. An Algorithm for Fast Beam Angle Selection in Intensity Modulated Radiotherapy. *Med Phys* (2010) 37:6443–52. doi: 10.1118/1.3517866
20. Das SK, Marks LB. Selection of Coplanar or Noncoplanar Beams Using Three-Dimensional Optimization Based on Maximum Beam Separation and Minimized Nontarget Irradiation. *Int J Radiat Onc Biol Phys* (1997) 38:643–5. doi: 10.1016/S0360-3016(97)89489-8
21. Wang X, Zhang X, Dong L, Liu H, Gillin M, Ahamad A, et al. Effectiveness of Noncoplanar IMRT Planning Using a Parallelized Multiresolution Beam Angle Optimization Method for Paranasal Sinus Carcinoma. *Int J Radiat Onc Biol Phys* (2005) 63:594–1. doi: 10.1016/j.ijrobp.2005.06.006
22. Oliver M, Gladwish A, Craig J, Chen J, Wong E. Incorporating Geometric Ray Tracing to Generate Initial Conditions for Intensity Modulated Arc Therapy Optimization. *Med Phys* (2008) 35:3137–50. doi: 10.1118/1.2937650
23. Jia X, Men C, Lou Y, Jiang S. Beam Orientation Optimization for Intensity Modulated Radiation Therapy Using Adaptive L2, 1–Minimization. *Phys Med Biol* (2011) 56:6205. doi: 10.1088/0031-9155/56/19/004
24. Schreibmann E, Xing L. Feasibility Study of Beam Orientation Class-Solutions for Prostate IMRT: Beam Orientation Class-Solutions for Prostate IMRT. *Med Phys* (2004) 31:2863–70. doi: 10.1118/1.1797571
25. Bangert M, Unkelbach J. Accelerated Iterative Beam Angle Selection in IMRT. *Med Phys* (2016) 43:1073–82. doi: 10.1118/1.4940350
26. Ezzell GA. Genetic and Geometric Optimization of Three-Dimensional Radiation Therapy Treatment Planning. *Med Phys* (1996) 23:293–305. doi: 10.1118/1.597660
27. Langer M, Brown R, Morrill S, Lane R, Lee O. A Generic Genetic Algorithm for Generating Beam Weights. *Med Phys* (1996) 23:965–71. doi: 10.1118/1.597858
28. Wu X, Zhu Y, Dai J, Wang Z. Selection and Determination of Beam Weights Based on Genetic Algorithms for Conformal Radiotherapy Treatment Planning. *Phys Med Biol* (2000) 45:2547. doi: 10.1088/0031-9155/45/9/308
29. Li Y, Yao J, Yao D. Automatic Beam Angle Selection in IMRT Planning Using Genetic Algorithm. *Phys Med Biol* (2004) 49:1915. doi: 10.1088/0031-9155/49/10/007
30. Hosseini-Ashrafi M, Bagherebadian H, Yahaqi E. Pre-Optimization of Radiotherapy Treatment Planning: An Artificial Neural Network Classification Aided Technique. *Phys Med Biol* (1999) 44:1513. doi: 10.1088/0031-9155/44/6/306
31. Rowbottom CG, Webb S, Oldham M. Beam-Orientation Customization Using an Artificial Neural Network. *Phys Med Biol* (1999) 44:2251. doi: 10.1088/0031-9155/44/9/312
32. Llacer J, Li S, Agazaryan N, Promberger C, Solberg TD. Non-Coplanar Automatic Beam Orientation Selection in Cranial IMRT: A Practical Methodology. *Phys Med Biol* (2009) 54:1337. doi: 10.1088/0031-9155/54/5/016
33. Skrobala A, Malicki J. Beam Orientation in Stereotactic Radiosurgery Using an Artificial Neural Network. *Radiother Onc* (2014) 111:296–300. doi: 10.1016/j.radonc.2014.03.010
34. Sadeghnejad Barkousaraie A, Ogunmolu O, Jiang S, Nguyen D. A Fast Deep Learning Approach for Beam Orientation Optimization for Prostate Cancer Treated With Intensity-Modulated Radiation Therapy. *Med Phys* (2020) 47:880–97. doi: 10.1002/mp.13986
35. Lomax A. Intensity Modulation Methods for Proton Radiotherapy. *Phys Med Biol* (1999) 44:185. doi: 10.1088/0031-9155/44/1/014
36. Gu W, O'Connor D, Nguyen D, Yu VY, Ruan D, Dong L, et al. Integrated Beam Orientation and Scanning-Spot Optimization in Intensity-Modulated Proton Therapy for Brain and Unilateral Head and Neck Tumors. *Med Phys* (2018) 45:1338–50. doi: 10.1002/mp.12788
37. Gu W, Neph R, Ruan D, Zou W, Dong L, Sheng K. Robust Beam Orientation Optimization for Intensity-Modulated Proton Therapy. *Med Phys* (2019) 46:3356–70. doi: 10.1002/mp.13641
38. Cao W, Lim GJ, Lee A, Li Y, Liu W, Ronald Zhu X, et al. Uncertainty Incorporated Beam Angle Optimization for IMPT Treatment Planning. *Med Phys* (2012) 39:5248–56. doi: 10.1118/1.4737870
39. Gu W, O'Connor D, Nguyen D, Yu VY, Ruan D, Dong L, et al. Integrated Beam Angle and Scanning Spot Optimization for Intensity Modulated Proton Therapy. *Int J Radiat Onc Biol Phys* (2017) 99:S107. doi: 10.1016/j.ijrobp.2017.06.254
40. Lee N, Kim TY, Kang DY, Choi JH, Jeong JH, Shin D, et al. Development of Manual Multi-Leaf Collimator for Proton Therapy in National Cancer Center. *Prog Med Phys* (2015) 26(4):250–57. doi: 10.14316/pmp.2015.26.4.250
41. Nie Y, De Santis L, Carratù M, O'Nils M, Lundgren J, Sommella P. *Deep Melanoma Classification With K-Fold Cross-Validation for Process Optimization*. Bari, Italy: 2020 IEEE Int Sympos Med Measure Appl (MeMeA) (2020) p. 1–6.
42. Ioffe S, Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *Proc 32nd Int Conf Mach Learning PMLR* (2015) 37:448–56.
43. Xu B, Wang N, Chen T, Li M. Empirical Evaluation of Rectified Activations in Convolutional Network. *arXiv:1505.00853* (2015).
44. Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. *arXiv:14126980* (2014).
45. Ronneberger O, Fischer P, Brox T. *U-Net: Convolutional Networks for Biomedical Image Segmentation*. Springer, Cham: Med Image Comput Assist Interv (2015). p. 234–41.

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: 29 September 2021
doi: 10.3389/fonc.2021.719398

Application of Deep Convolution Network to Automated Image Segmentation of Chest CT for Patients With Tumor

Hui Xie[1,2], *Jian-Fang Zhang*[3] *and Qing Li*[2,4*]

[1] Department of Radiation Oncology, Affiliated Hospital (Clinical College) of Xiangnan University, Chenzhou, China, [2] Key Laboratory of Medical Imaging and Artifical Intelligence of Hunan Province, Chenzhou, China, [3] Department of Physical Examination, Beihu Centers for Disease Control and Prevention, Chenzhou, China, [4] School of Medical Imaging and Rehabilitation, Xiangnan University, Chenzhou, China

Objectives: To automate image delineation of tissues and organs in oncological radiotherapy by combining the deep learning methods of fully convolutional network (FCN) and atrous convolution (AC).

Methods: A total of 120 sets of chest CT images of patients were selected, on which radiologists had outlined the structures of normal organs. Of these 120 sets of images, 70 sets (8,512 axial slice images) were used as the training set, 30 sets (5,525 axial slice images) as the validation set, and 20 sets (3,602 axial slice images) as the test set. We selected 5 published FCN models and 1 published Unet model, and then combined FCN with AC algorithms to generate 3 improved deep convolutional networks, namely, dilation fully convolutional networks (D-FCN). The images in the training set were used to fine-tune and train the above 8 networks, respectively. The images in the validation set were used to validate the 8 networks in terms of the automated identification and delineation of organs, in order to obtain the optimal segmentation model of each network. Finally, the images of the test set were used to test the optimal segmentation models, and thus we evaluated the capability of each model of image segmentation by comparing their Dice coefficients between automated and physician delineation.

Results: After being fully tuned and trained with the images in the training set, all the networks in this study performed well in automated image segmentation. Among them, the improved D-FCN 4s network model yielded the best performance in automated segmentation in the testing experiment, with an global Dice of 87.11%, and a Dice of 87.11%, 97.22%, 97.16%, 89.92%, and 70.51% for left lung, right lung, pericardium, trachea, and esophagus, respectively.

Conclusion: We proposed an improved D-FCN. Our results showed that this network model might effectively improve the accuracy of automated segmentation of the images in thoracic radiotherapy, and simultaneously perform automated segmentation of multiple targets.

Keywords: deep learning, convolutional network, medical image segmentation, similarity coefficient, radiotherapy

OPEN ACCESS

Edited by:
Jose Eduardo Villarreal Barajas, Royal Devon and Exeter Hospital, United Kingdom

Reviewed by:
Weiwei Zong, Henry Ford Health System, United States
Jung Sun Yoo, Hong Kong Polytechnic University, Hong Kong, SAR China

***Correspondence:**
Qing Li
xnxyliqing@163.com

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

Received: *02 June 2021*
Accepted: *13 September 2021*
Published: *29 September 2021*

Citation:
Xie H, Zhang J-F and Li Q (2021) Application of Deep Convolution Network to Automated Image Segmentation of Chest CT for Patients With Tumor. Front. Oncol. 11:719398. doi: 10.3389/fonc.2021.719398

INTRODUCTION

As medical imaging technology and computer technology are being increasingly applied in the field of oncology radiotherapy, radiotherapy has now developed to a stage where precision radiotherapy, characterized by image-guided and adaptive radiotherapy, became predominant (1, 2). Precision radiotherapy requires precise delineation of the target area and organs at risk, accompanied by online image-guided therapeutic irradiation, as well as the modification and adjustment of subsequent radiotherapy plans, which ultimately aimed to ensure the delivery of the effective dose to the target while avoiding normal tissues and organs. In current practices of clinical radiation therapy planning, the delineation of the target area and organs at risk usually involves manual work of experienced radiologists and tumor radiotherapy physicists, which is a time- and labor-intensive process. The accuracy and efficiency rely heavily on the clinical experience of physicians and physicists, and it cannot avoid the large variability between delineators. The development of computer-automated processing and artificial intelligence is driving rapid advances in automated and semi-automated delineation algorithms based on various computational image processing techniques, some of which have been put into clinical practices, including segmentation algorithms based on features of image gray level, color and texture, nonlinear diffusion algorithms using level set model, automated segmentation algorithms based on templates, and machine learning algorithms based on manually extracted features (3). However, these semi-automated and automated segmentation algorithms are still immature. Especially when boundaries between organ tissues are not obvious, the performance of automated segmentation is particularly unsatisfactory. The template-based algorithm requires a lot of running time due to the compositions of the template library, while the recognition of image features depending on professional experience is not necessarily ideal. Besides, most of the current algorithms are designed for a single organ or tissue, thereby being incapable of auto-segmenting multiple organs or tissue, which results in the inefficiency of clinical work.

In recent years, artificial intelligence technologies based on deep learning have presented tremendous opportunities for various fields including clinical medicine. Deep convolutional neural network (DCNN), or convolutional neural network (CNN) (4), is widely used in computer image recognition and more and more in the research of automated segmentation of medical images. For example, the U-net DCNN proposed by Olaf et al. (5) was applied to biomedical image recognition to achieve automated segmentation of biological cell images. When the DCNN is applied to medical image segmentation, image features can be extracted layer by layer from low to high through multi-layer convolution operation, and the automatically extracted features are correctly classified through iterative training and learning of calibration datasets, so as to achieve simultaneous segmentation of multi-structure targets (6–8). If we combine the trained DCNN model and graphic processing units (GPU) hardware acceleration, computed tomography (CT) images of tissues or organs experiencing radiotherapy can be segmented rapidly, and the structure of the target area and organs at risk can be accurately delineated automatically, which will promote the further development of precision radiotherapy.

Abbreviations: FCN, fully convolutional network; AC, atrous convolution; D-FCN, dilation fully convolutional networks; DCNN, Deep convolutional neural network; CNN, convolutional neural network; GPU, graphic processing units; CT, computed tomography.

MATERIALS AND METHODS

Patient Datasets and Computer Working Platform

For this experiment, we collected the image data from the image database of clinical radiotherapy cases established at the early phase by the Department of Radiotherapy of Affiliated Hospital of Xiangnan University. Our research team searched the image database according to the disease type and structure, with search items such as lung cancer, left lung, right lung, pericardium, trachea, and esophagus, and eventually obtained the image data of clinical lung cancer cases undergoing radiotherapy. The image data included chest CT scan sequences of desensitized patients and the corresponding files of organ structure contour. With the aid of relevant medical image processing technology that analyzed and extracted the contouring data of the structure of each organ in the images, the organ delineation atlas corresponding to each slice image on the CT image sequence was thus generated.

The experimental data set contained a total of 120 sets of chest CT images. Among them, 70 sets were randomly selected as the training set that included 8,512 axial slice images and organ delineation contour maps; 30 sets were randomly selected as the validation set that included 5,525 axial slice images and organ delineation atlases; 20 sets were randomly selected as the test set that included 3,602 axial slice images and organ delineation atlases. **Figure 1** is one of the examples, in which **Figure 1A** is the axial slice image of the patient and **Figure 1B** is the organ atlases delineated by the physician.

This study was performed on an ultramicro 4028GR-TR computer server. Its hardware system contained two Intel E5-2650V4 models of CPU, 128 GB of memory, 3 TB of SSD hard disk, and 8 GPUs of NVIDIA GeForce 1080Ti model; the software system included Ubuntu Server 16.04 operating system, CUDA8.0 and cuDNN6.0, and the latest Caffe deep learning framework.

Optimization and Improvement of Fully Convolutional Network

The basic mechanism of the FCN proposed by Shelhamer et al. (9) is that FCN extracts image features through convolution, performs feature compression for feature image pooling processing, obtains segmented images as big as the original image through upsampling, and then optimizes output adjustment with the jump structure.

Six published networks were used in our study, including FCN based on the VGG16 (10) algorithm, the DeepLab series

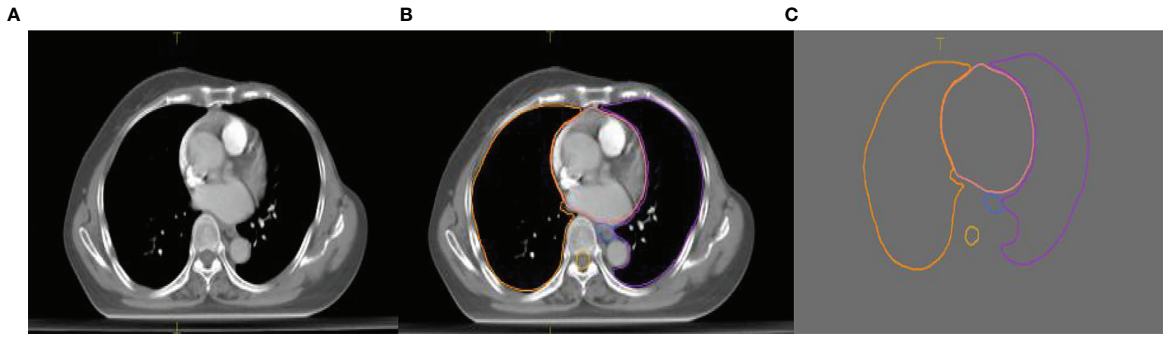


FIGURE 1 | Images of a lung patient in the experimental data. **(A)** is the original axial slice image of CT scan; **(B)** is the axial slice image of CT scan delineated by the physician; **(C)** is the contour of delineated organs extracted after processing.

proposed by Chen et al., and the U-net (5, 9, 11, 12), as well as three dilation fully convolutional networks (D-FCN) modified by our research team through combining the deep learning methods of fully convolutional network and atrous convolution (AC). It's reported that systematic dilation supports exponential expansion of the receptive field without loss of resolution or coverage, which increases the accuracy of state-of-the-art semantic segmentation systems (13). We used the training dataset for tuning and training to obtain the optimal network models for the forementioned chest image segmentation by comparing and comprehensively analyzing the automated segmentation results and manual delineation results of each network training model.

Preprocessing of data: Because the pre-training models selected for this study are based on the results of training with RGB three-channel natural images, and the medical image sets used in this study are single-channel CT images, it is necessary to construct the single-channel medical image into three-channel image in the data input layer. In the present study, we made two copies of the original image data to constitute virtual three-channel medical image data.

Published model training: We selected 5 DCNNs based on the FCN VGG16 algorithm, including FCN 32s, FCN 16s, FCN 8s, DeepLab-largeFOV, and DeepLabv2-VGG16, and 1 U-net model, which are suitable for image segmentation. We also leveraged these models trained on other natural image data sets as pre-trained models. We modified and optimized the pre-trained models, including changing the data input layer to adapt it to the data format of the medical image in our datasets. We added the window adjustment layer by combining the difference between medical images and natural images. In our study, the [-300, 600] window width was divided into three equal parts according to the characteristics of the window values of each structure of chest CT. The equally divided value range was the window width, and the median value was the window value. The window was adjusted for each channel separately. We set the number of characteristic maps of the output layer according to the category of the target that the experiment was designed to segment, and used the images in the training set to perform 500,000 repeated iterations for tuning and training these network models, so as to obtain the optimal training result of each network. In addition, it is necessary in the training process to adjust and optimize the training hyperparameters as actual training situations might change – specifically, learning strategy, initial learning rate, batch size, momentum, weight decay rate, etc., to improve the prediction accuracy of the model.

Training of the modified models: While combining the characteristics of FCN and the idea of atrous convolution, the pool3, pool4, and pool5 of the FCN 32s network, as well as part or all of the subsequent convolutional layers, were modified into dilation convolutional layer, namely, the so-called D-FCN. A total of 3 modified FCN models were thereby generated: D-FCN 4s, D-FCN 8s, and D-FCN 16s (**Figure 2**). Similarly, we employed the same datasets to tune and train the modified D-FCN models with the FCN32s network model as a pre-training model.

Optimal model validation: During the training process, a series of training models were generated with every 5,000 iterations as an observation mirror image. The manually delineated structural contour regions in the 5,525 images of the 30 patients in the validation set were used as the prediction targets to validate the segmentation consistency of the training models that were obtained from the training of the above 9 networks, respectively. We worked out the Dice coefficient by calculating the similarity between the automated segmentation results of the training models and the manual delineation results, and thus drew the Dice curve of the training models under different iterative mirrors of each network. Finally, we found the optimal segmentation model of each network by analyzing the Dice curve.

Automated Image Segmentation Test of Network Models

The manually delineated contour regions of 20 sets of 3,602 axial slice images in the test set were used as the prediction targets. The optimal segmentation models selected above were employed to perform the automated segmentation of the targets so that we could test the effectiveness of each network model and the accuracy of automated segmentation. We calculated the similarity between the automated segmentation results and the manual delineation results in terms of global and individual organ structures, respectively. We compared the Dice coefficients

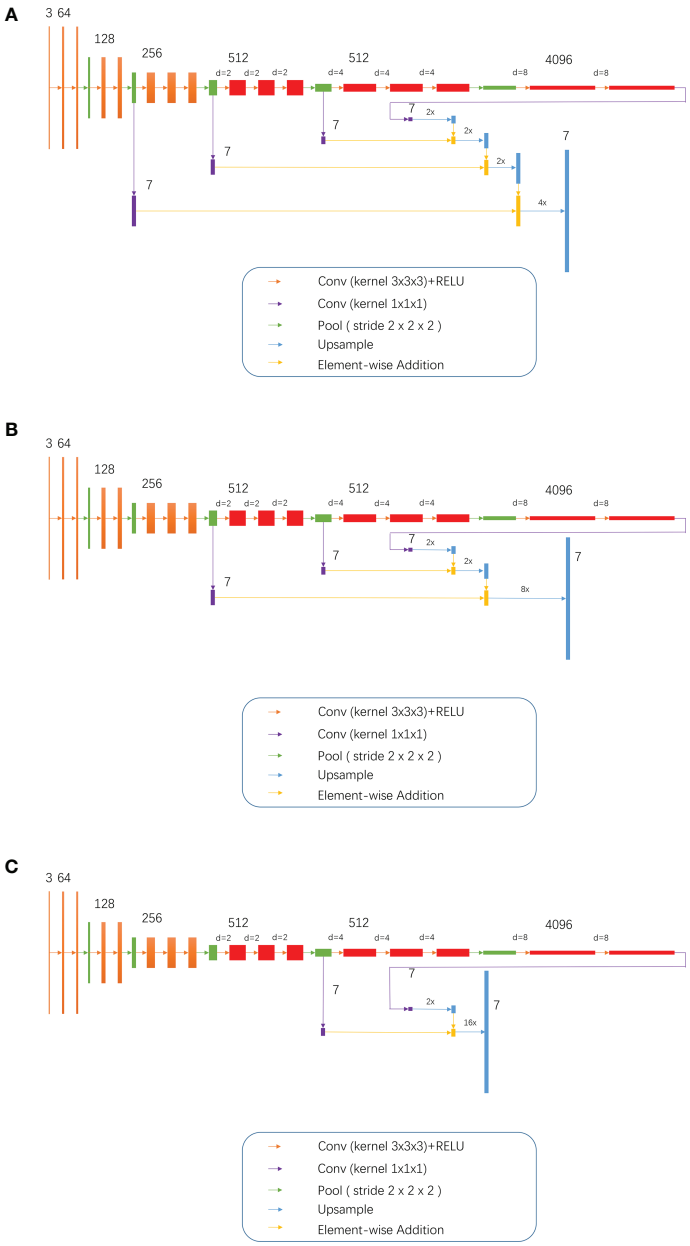


FIGURE 2 | 3, 7, 64, 128, 512 and 4096 meant 3 7, 64, 128, 512 and 4096 image channels, respectively; d meant d-1 dilation were plug in between every two elements of the convolution kernel; 2x and 4x are multiples of upsampling. **(A)** D-FCN 4S; **(B)** D-FCN 8S; **(C)** D-FCN 16S.

and comprehensively evaluated each network model while considering the speed of automated segmentation processing.

Evaluation Indicators

As we all know, intersection-Over-Union (IoU) and Dice coefficient are both important and common indicators for segmentation neural network assessment. The previous report which compared Dice coefficient with IoU, indicated that using Dice could have higher score than IoU (14). Therefore, in this paper, Dice coefficient is used to evaluate the effect of automated segmentation by network models, that is, to evaluate the similarity between the automated image segmentation results and the manual delineation results of physicians. Dice is calculated by:

$$\mathrm{Dice}(X, Y) = \frac{2 * |X \cap Y|}{|X| + |Y|}$$

Where X denotes the set of pixels for the automatically segmented image, Y denotes the set of pixels for the manually delineated image, $|X \cap Y|$ is the intersection of two sets of pixels,

and | X + | Y | is the union set of the both. The range of Dice is [0, 1], and the higher the value of Dice is, the closer the result of automated segmentation is to that of manual delineation. In this paper, we calculated not only the global Dice of all segmented target regions, but also the Dice of individual segmented target region, so as to evaluate the effect of automated segmentation by the model more comprehensively.

RESULTS

In our study, the training set was comprised of 70 sets of 8,512 CT axial slice images of patients undergoing pulmonary radiotherapy, as well as organ atlases manually delineated by radiologists. Nine deep networks, including 6 published networks and 3 networks modified by us, were tuned and trained for automated image segmentation, respectively. 30 sets of 5,525 CT images, as well as manually delineated organ contour atlases, constituted the validation set, and were used to validate the consistency of the models obtained from tuning and training. The optimal segmentation model of each network was determined by Dice analysis. Finally, the effectiveness and accuracy of the optimal segmentation model of each network were tested by a test set containing 20 sets of 3,602 CT images, and the performance of each model in automated image segmentation of radiotherapy localization was comprehensively evaluated.

Training and Optimization Results of Network Model

Figure 3 presents the Dice curves for the training and validation processes of the 9 networks. The Dice value of each network model increased with the number of iterations during the training. The convergence of Dice values showed that the improved D-FCN 4s model constructed in this study

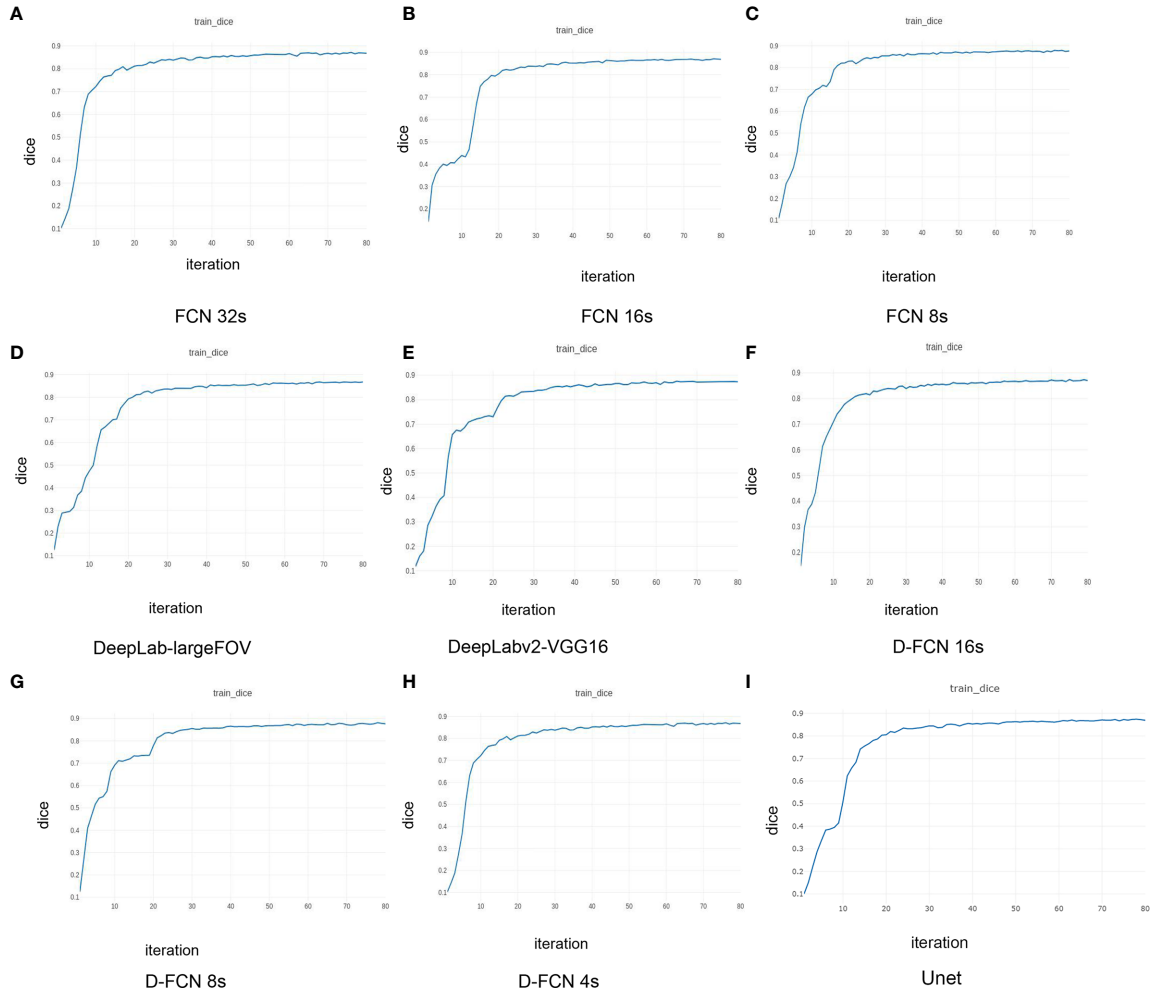


FIGURE 3 | | Dice curves of the training effects of the 8 network models. **(A)** is the Dice curve of the training effect of FCN 32s network; **(B)** is of FCN 16s network; **(C)** is of FCN 8s network; **(D)** is of DeepLab-largeFOV network; **(E)** is of DeepLabv2-VGG16 network; **(F)** is of D-FCN 16s network; **(G)** is of D-FCN 8s network; **(H)** is of D-FCN 4s network; **(I)** is of Unet network.

(**Figure 3H**) had the fastest convergence rate and the best stable convergence rate compared with the other models.

Table 1 shows the statistical results of the iterative operation of automated segmentation of organs for each model, including the optimal Dice score and the number of iterations when reaching the optimal value. All the models in our study presented high global optimal Dice, which suggested that the automated segmentation results were close to the expert delineation results. The D-FCN 4s model proposed in this paper had the highest global Dice (87.11%) compared with the other models, indicating that it had superior performance in automated segmentation to the other models.

Test Results and Analysis of Automated Segmentation

Table 2 shows the test results of automated segmentation of target structures for the 9 models by using the images in the test set. The table lists the global Dice of one-time automated segmentation of 6 target organs of each test case by different models, the optimal Dice of individual organ structure, and the automated segmentation operation time of each model. The comparison between the automated segmentation results of each model for the test set or the validation set both showed that D-FCN4s has a better segmentation effect than or is equivalent to the other network models, regardless of the global Dice or the Dice of the individual structure. Regarding automated segmentation operation time, D-FCN4s was slower in prediction segmentation than the other models because it preserved more image details for the sake of a finer segmentation effect. There was no downsampling operation above the Pool3 layer, and as a result, the resolution of the feature image in the following layers was larger, so the amount of computation increased greatly and the speed of prediction became slower. However, the predicted automated delineation speed of DFCN4s, which took less than 3 minutes on average, was acceptable in the practice of radiotherapy

Figure 4 shows the comparison between the results of automated segmentation delineation of some test cases and manual delineation by radiologists. In this figure, each horizontal line lists a comparison of different test cases. The left-side images were delineated by physicians and the right-side images by the D-FCN4s model automatically. The delineated contours of the both sides are very consistent with each other, especially for some closed esophageal or tracheal contours that are not easy to be distinguished by naked eyes. The trained D-FCN4s show good ability of predictive segmentation.

DISCUSSION

When designing a clinical radiotherapy plan, radiologists are required not only to accurately determine and delineate the

TABLE 1 | Iterative operation results of automated organ segmentation for the 8 network models.

Network Model	Global Dice/%	95% CI		Best epoch (× 10000)
		Lower	Upper	
FCN 32s	86.32	78.25	93.50	77
FCN 16s	86.51	78.68	93.69	80
FCN 8s	86.95	79.23	93.80	78
Deeplab-largeFOV	86.36	77.94	93.67	76
Deeplabv2-VGG16	86.89	79.00	93.93	73
D-FCN 16s	86.47	78.09	93.82	72
D-FCN 8s	87.05	79.11	94.01	78
D-FCN 4s	87.11	79.40	93.95	78
Unet	86.81	78.96	93.75	74

CI, confidence interval.

TABLE 2 | Test results of the optimal segmentation models of the 8 network models.

Network Model	Dice/%							Average time/s
	Global	Lung (L)	Lung (R)	Heart	Esophageal	Trachea	Spinal Cord	
FCN 32s	86.32	97.15	96.75	88.87	69.13	84.36	81.68	32
FCN 16s	86.51	97.2	96.87	89.34	69.78	84.65	81.2	36
FCN 8s	86.95	97.17	97.08	89.17	70.63	85.13	82.52	36
Deeplab-largeFOV	86.36	97.17	96.94	89.28	68.19	84.51	82.08	13
Deeplabv2-VGG16	86.89	97.14	97.14	89.85	69.83	85.15	82.2	35
D-FCN 16s	86.47	97.2	97.02	89.71	68.23	84.75	81.9	30
D-FCN 8s	87.05	97.21	97.01	90.21	69.79	85.4	82.68	55
D-FCN 4s	87.11	97.22	97.16	89.92	70.51	85.05	82.78	173
Unet	86.81	97.17	96.98	89.67	69.95	84.55	82.51	20

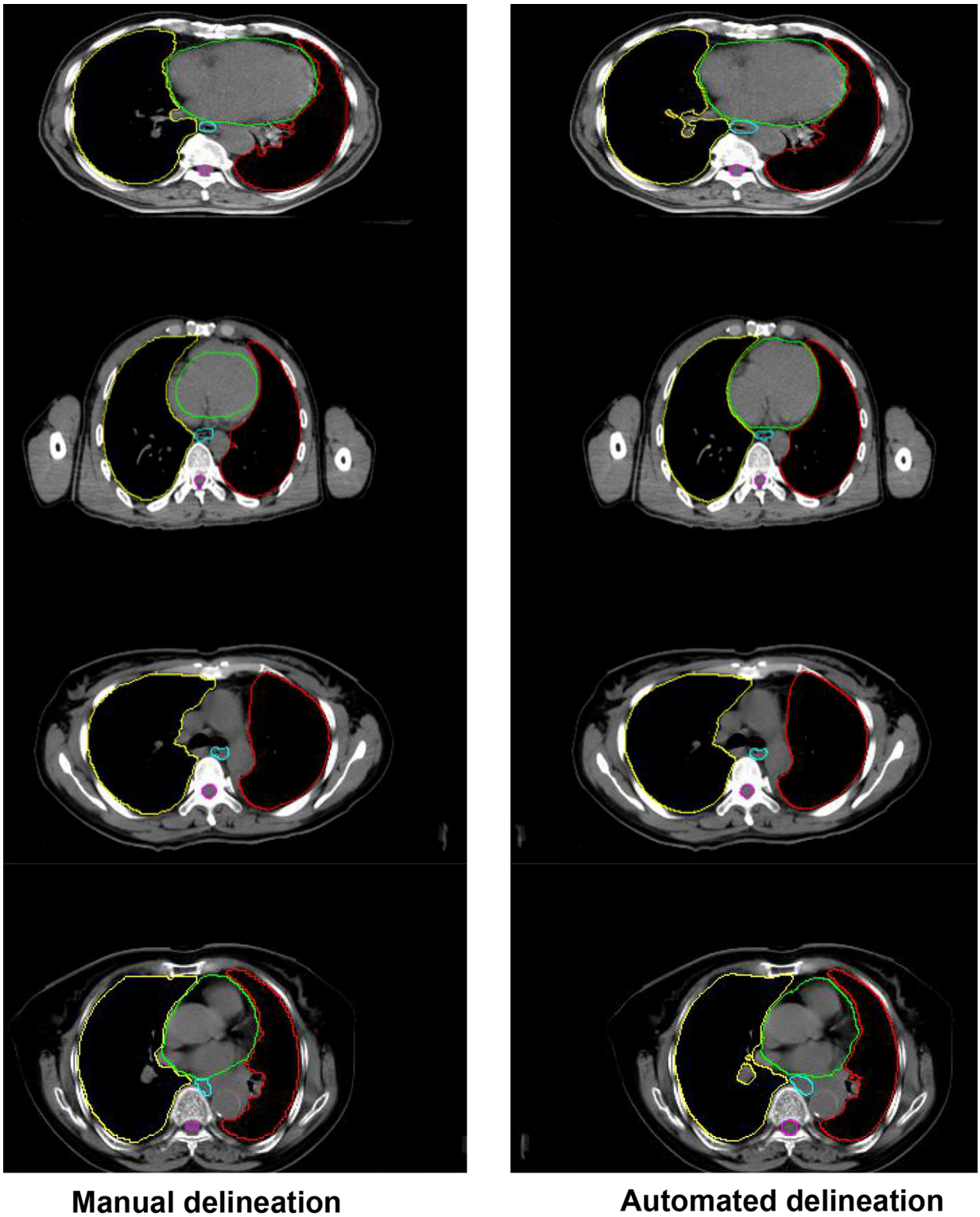


FIGURE 4 | The comparison between the automated segmentation delineation of some test cases and the manual delineation results of the radiologist. In the figure, each horizontal line lists a comparison of different test cases. The left side is delineated by physicians and the right side by the D-FCN4s model automatically.

tumor target area to be treated, but also to delineate the normal tissues and organs at risk that may be potentially irradiated. The accuracy of contouring organs at risk determines the quality of dose optimization in radiotherapy planning (15), thus directly affecting the success of radiotherapy or the incidence of complications (16). However, the accuracy of manual delineation is highly dependent on the clinical experience of radiologists, whose manual work might be inefficient (17–19). Therefore, automated organ delineation methods based on image segmentation have been attracting the tremendous interest of many scholars, who developed many different automated image segmentation and delineation algorithm models. Nonetheless, so far, most of the automated partition and delineation software commonly used in radiotherapy clinically are regional segmentation methods based on regional features such as gray level distribution (20), and the template automated delineation method based on empirical atlas and deformation model (21). The former is not effective for regional segmentation for little variation in gray level distribution, while the latter is sensitive to template quality, and the delineation effect is not good enough to meet the clinical requirements. Relevant studies (22) indicated that the mean Dice values of template-based automated partition delineation software using single and multiple template delineation, in the automated delineation of the geometric accuracy of organs at risk in head and neck radiotherapy, were 0.68 ± 0.20 and 0.74 ± 0.16, respectively (P = 0.01). Peng Yinglin et al. (23) found in the pre-clinical test report of the same automated delineation software that when the image to be delineated is significantly different from the template image, the Dice score of automated organ delineation was only 0.46 ~ 0.89. In recent years, the emerging deep learning method based on neural networks can automatically learn features and perform feature recognition at multiple levels, which has achieved great results in the application of automated recognition and segmentation of medical images. Ilsang Woo et al. (24) used a convolutional network method for automated segmentation of magnetic resonance images, and the Dice/% was greater than 85%, significantly higher than that of the traditional algorithm. Khalifa F et al. (25) employed a multiscale convolutional network method to perform automated segmentation of the kidneys on abdominal CT images with a Dice/% of 97.27%. In this paper, the pre-training and tuning methods of deep learning were leveraged to pre-train and optimize the published DCNN models such as FCN and DeepLab and the Unet model suitable for natural image segmentation, which were further improved by optimizing the FCN network model and combining the atrous convolution method. After the models were fully tuned and trained, the automated segmentation test was performed with another set of image in the same category. The results showed that the performance of the improved network model in automated organ segmentation was better than that of the other published models. The D-FCN 4S model proposed in this study brought about results that were very close to those of manual delineation in most segmentation of organ structures. The test experiment showed that in terms of individual organ structure, this D-FCN 4S model had the highest accuracy in automated segmentation of lung and pericardium, with an Dice of 97% and 89%, respectively. The similarity between automated and manual delineation of trachea and esophagus was relatively low, with an Dice of 70.51% and 85.05%, respectively. Since there is often great disagreement when physicians delineate organs such as the esophagus in the closed state (19, 26–29), using these datasets with great disagreement to train network models might potentially reduce the automated recognition ability of the models. At the same time, individual differences in physician delineation could somewhat reduce the consistency of automated delineation tests (30, 31). Therefore, when using machine learning tools like artificial intelligence automated delineation, it is necessary to label and optimize the data for deep learning models, and the results from automated delineation still need to be confirmed and modified by physicians. In addition, the results of this study revealed that the ability of the model to recognize and segment some small organ structures is relatively poor, and thus we need more efforts for debugging of parameters and iteration deepening when training and optimizing the models. We should seek more appropriate network parameters and iteration endpoints to improve the automated recognition ability and segmentation accuracy of the model. Besides, the cross-validation with smaller bias should be performed in the future studies. These are the issues that need to be addressed in our subsequent studies.

CONCLUSION

This study introduced DCNN based on natural image segmentation into medical image segmentation and proposed an modified D-FCN that could effectively improve the ability of predictive segmentation of target images. Combined with GPU hardware acceleration, further optimization of network parameters and training levels might be expected to achieve rapid segmentation of images of organs at risk in the thoracic radiotherapy planning, thus paving the ground for automated design of radiotherapy plans in the future.

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

HX and QL: substantial contributions to the conception and design of the work. HX, J-FZ, and QL: acquisition, analysis, and interpretation of data for the work. HX: drafting the work.

QL, revising it critically for important intellectual content. HX, J-FZ, and QL, agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by: Science and Technology Funding Project of Chenzhou City, Hunan Province China (No. ZDYF2020165) (QL); Science and Technology Funding Project of Hunan Education Department, China (No.1 8C1023) (HX); Science and Technology Funding Project of Hunan Province, China (No. 2020SK52201) (HX); Key Laboratory of Tumor Precision Medicine, Hunan colleges and Universities Project (2019-379) (QL).

ACKNOWLEDGMENTS

Thanks for the support from the above funding and Shenzhen Yino Intelligence Techonology Co, Ltd.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2021.719398/full#supplementary-material

REFERENCES

1. Gendrin C, Markelj P, Pawiro SA, Spoerk J, Bloch C, Weber C, et al. Validation for 2D/3D Registration. II: The Comparison of Intensity- and Gradient-Based Merit Functions Using a New Gold Standard Data Set. *Med Phys* (2011) 38(3):1491–502. doi: 10.1118/1.3553403
2. Wang X, Zhang X, Qiu C, Yang N. STAT3 Contributes to Radioresistance in Cancer. *Front Oncol* (2020) 10:1120. doi: 10.3389/fonc.2020.01120
3. Lösel PD, van de Kamp T, Jayme A, Ershov A, Faragó T, Pichler O, et al. Introducing Biomedisa as an Open-Source Online Platform for Biomedical Image Segmentation. *Nat Commun* (2020) 11(1):5577. doi: 10.1038/s41467-020-19303-w
4. Pospisil DA, Pasupathy A, Bair W. 'Artiphysiology' Reveals V4-Like Shape Tuning in a Deep Network Trained for Image Classification. *Elife* (2018) 7: e38242. doi: 10.7554/eLife.38242
5. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: N Navab, J Hornegger, W Wells, A Frangi, editors. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015; Lecture Notes in Computer Science*, Cham: Springer. (2015) 9351:234–41. doi: 10.1007/978-3-319-24574-4_28
6. Huang Y, Hu G, Ji C, Xiong H. Glass-Cutting Medical Images *via* a Mechanical Image Segmentation Method Based on Crack Propagation. *Nat Commun* (2020) 11(1):5669. doi: 10.1038/s41467-020-19392-7
7. Smistad E, Falch TL, Bozorgi M, Elster AC, Lindseth F. Medical Image Segmentation on GPUs–A Comprehensive Review. *Med Image Anal* (2015) 20(1):1–18. doi: 10.1016/j.media.2014.10.012
8. Niu S, Chen Q, Sisternes L, Ji ZX, Zhou ZM, Rubin D. Robust Noise Region-Based Active Contour Model *via* Local Similarity Factor for Image Segmentation. *Pattern Recogn* (2017) 61:104–19. doi: 10.1016/j.patcog.2016.07.022
9. Shelhamer E, Long J, Darrell T. Fully Convolutional Networks for Semantic Segmentation. *IEEE Trans Pattern Anal Mach Intell* (2017) 39(4):640–51. doi: 10.1109/TPAMI.2016.2572683
10. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* (2014), 1409–556.
11. Chen LC, Papandreou G, Kokkinos I, Murphy K, Yuille AL. DeepLab: Semantic Image Segmentation With Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Trans Pattern Anal Mach Intell* (2018) 40(4):834–48. doi: 10.1109/TPAMI.2017.2699184
12. Artacho B, Savakis A. Waterfall Atrous Spatial Pooling Architecture for Efficient Semantic Segmentation. *Sensors (Basel)* (2019) 19(24):5361. doi: 10.3390/s19245361
13. Yu F, Koltun V. Multi-Scale Context Aggregation by Dilated Convolutions. Published as a conference paper at ICLR 2016.
14. Available at: https://medium.com/datadriveninvestor/deep-learning-in-medical-imaging-3c1008431aaf.
15. Pei Y, Yang W, Wei S, Cai R, Li J, Guo S, et al. Automated Measurement of Hip-Knee-Ankle Angle on the Unilateral Lower Limb X-Rays Using Deep Learning. *Phys Eng Sci Med* (2021) 44(1):53–62. doi: 10.1007/s13246-020-00951-7
16. 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
17. Liang S, Tang F, Huang X, Yang K, Zhong T, Hu R, et al. Deep-Learning-Based Detection and Segmentation of Organs at Risk in Nasopharyngeal Carcinoma Computed Tomographic Images for Radiotherapy Planning. *Eur Radiol* (2019) 29(4):1961–7. doi: 10.1007/s00330-018-5748-9
18. 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
19. 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
20. Bidola P, Martins de Souza E Silva J, Achterhold K, Munkhbaatar E, Jost PJ, Meinhardt AL, et al. A Step Towards Valid Detection and Quantification of Lung Cancer Volume in Experimental Mice With Contrast Agent-Based X-Ray Microtomography. *Sci Rep* (2019) 9(1):1325. doi: 10.1038/s41598-018-37394-w
21. Lin M, Chen JH, Wang X, Chan S, Chen S, Su MY. Template-Based Automatic Breast Segmentation on MRI by Excluding the Chest Region. *Med Phys* (2013) 40(12):122301. doi: 10.1118/1.4828837
22. Yin XJ, Hu CR, Zhang XC, Lin J, Lin SJ. The Feasibility Study of Atlasbased Auto-Segmentation (ABAS) Software in Head-and-Neck Cancer. *Chi J Radiat Oncol* (2016) 25(11):1233–7. doi: 10.3760/cma.j.issn.1004-4221.2016.11.019
23. Peng YL, You Y, Han F, Hu J, Wang ML, Deng XW. Commissioning of an Atlas-Based Auto- Segmentation Software for Application in Organ Contouring of Radiotherapy Planning. *Chi J Radiat Oncol* (2016) 25 (6):609–14. doi: 10.3760/cma.j.issn.1004-4221.2016.06.014
24. Woo I, Lee A, Jung SC, Lee H, Kim N, Cho SJ, et al. Fully Automatic Segmentation of Acute Ischemic Lesions on Diffusion-Weighted Imaging Using Convolutional Neural Networks: Comparison With Conventional Algorithms. *Korean J Radiol* (2019) 20(8):1275–84. doi: 10.3348/kjr.2018.0615
25. Khalifa F, Soliman A, Elmaghraby A, Gimel'farb G, El-Baz A. 3d Kidney Segmentation From Abdominal Images Using Spatial-Appearance Models. *Comput Math Methods Med* (2017) 2017:9818506. doi: 10.1155/2017/9818506
26. Pu J, Gu S, Liu S, Zhu S, Wilson D, Siegfried JM, et al. CT Based Computerized Identification and Analysis of Human Airways: A Review. *Med Phys* (2012) 39 (5):2603–16. doi: 10.1118/1.4703901
27. Kiraly AP, Higgins WE, McLennan G, Hoffman EA, Reinhardt JM. Three-Dimensional Human Airway Segmentation Methods for Clinical Virtual

Bronchoscopy. *Acad Radiol* (2002) 9(10):1153–68. doi: 10.1016/s1076-6332(03)80517-2
28. Artaechevarria X, Pérez-Martín D, Ceresa M, de Biurrun G, Blanco D, Montuenga LM, et al. Airway Segmentation and Analysis for the Study of Mouse Models of Lung Disease Using Micro-CT. *Phys Med Biol* (2009) 54 (22):7009–24. doi: 10.1088/0031-9155/54/22/017
29. Pu J, Fuhrman C, Good WF, Sciurba FC, Gur D. A Differential Geometric Approach to Automated Segmentation of Human Airway Tree. *IEEE Trans Med Imaging* (2011) 30(2):266–78. doi: 10.1109/TMI.2010.2076300
30. Vinod SK, Jameson MG, Min M, Holloway LC. Uncertainties in Volume Delineation in Radiation Oncology: A Systematic Review and Recommendations for Future Studies. *Radiother Oncol* (2016) 121(2):169–79. doi: 10.1016/j.radonc.2016.09.009
31. Ahn SH, Yeo AU, Kim KH, Kim C, Goh Y, Cho S, et al. Comparative Clinical Evaluation of Atlas and Deep-Learning-Based Auto-Segmentation of Organ Structures in Liver Cancer. *Radiat Oncol* (2019) 14(1):213. doi: 10.1186/s13014-019-1392-z

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

MINI REVIEW
published: 08 October 2021
doi: 10.3389/fonc.2021.753286

Precision Interventional Brachytherapy: A Promising Strategy Toward Treatment of Malignant Tumors

Pan He[1†], Siwen Guan[1†], En Ren[1], Hongwei Chen[1], Hu Chen[1], Yisheng Peng[2], Bin Luo[2], Yongfu Xiong[3], Bo Li[2], Jingdong Li[3*], Jingsong Mao[1,4*] and Gang Liu[1*]*

[1] State Key Laboratory of Molecular Vaccinology and Molecular Diagnostics, Center for Molecular Imaging and Translational Medicine, School of Public Health, Xiamen University, Xiamen, China, [2] Department of General Surgery (Hepatobiliary Surgery), The Affiliated Hospital of Southwest Medical University, Luzhou, China, [3] Institute of Hepato-Biliary-Intestinal Disease, Department of Hepatobiliary Surgery, Affiliated Hospital of North Sichuan Medical College, Nanchong, China, [4] Department of Radiology, Xiang'an Hospital of Xiamen University, Xiamen, China

Precision interventional brachytherapy is a radiotherapy technique that combines radiation therapy medicine with computer network technology, physics, *etc.* It can solve the limitations of conventional brachytherapy. Radioactive drugs and their carriers change with each passing day, and major research institutions and enterprises worldwide have conducted extensive research on them. In addition, the capabilities of interventional robotic systems are also rapidly developing to meet clinical needs for the precise delivery of radiopharmaceuticals in interventional radiotherapy. This study reviews the main radiopharmaceuticals, drug carriers, dispensing and fixation technologies, and interventional robotic precision delivery systems used in precision brachytherapy of malignant tumors. We then discuss the current needs in the field and future development prospects in high-precision interventional brachytherapy.

Keywords: precision brachytherapy, interventional surgery, radiopharmaceuticals, drug delivery, tumor treatment

OPEN ACCESS

Edited by:
Jose Eduardo Villarreal Barajas, Royal Devon and Exeter Hospital, United Kingdom

Reviewed by:
Yuanzeng Min, University of Science and Technology of China, China
Peng Mi, Sichuan University, China
Xin Pang, First Affiliated Hospital of Zhengzhou University, China

***Correspondence:**
Bo Li
liboer2002@126.com
Jingdong Li
Li-Jingdong@hotmail.com
Jingsong Mao
maojingsong163@163.com
Gang Liu
gangliu.cmitm@xmu.edu.cn

[†]*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: *04 August 2021*
Accepted: *23 September 2021*
Published: *08 October 2021*

Citation:
He P, Guan S, Ren E, Chen H, Chen H, Peng Y, Luo B, Xiong Y, Li B, Li J, Mao J and Liu G (2021) Precision Interventional Brachytherapy: A Promising Strategy Toward Treatment of Malignant Tumors. Front. Oncol. 11:753286. doi: 10.3389/fonc.2021.753286

INTRODUCTION

Malignant tumors (MT) are among the most severe diseases threatening human health conditions in the 21st century. MT are also the focus of substantial research worldwide (1–3). The primary treatment modalities for MT include surgical treatment, radiotherapy, chemotherapy, immunotherapy, as well as the newly developed photothermal, photodynamic, and sound-dynamic therapies (4–7). Of these, comprehensive surgical treatment mainly based on surgical treatment plays a crucial role in the treatment of MT (8). However, it is often difficult, if not impossible, to diagnose MT early. Accordingly, by the time tumors are found, patients are in the middle and late stage of the disease, and the rate of surgical resection and radical cure is thus low (9–11). Therefore, tools to improve the therapeutic effect of existing treatments on patients with MT is an urgent problem to be solved and a hot research topic at the moment.

Radiotherapy is one of the three most-common MT treatment modalities, along with surgery and chemotherapy. It uses ionizing radiation to kill tumor cells and shrink tumors (12, 13). Versus

surgery and chemotherapy, radiotherapy uses colorless, odorless, invisible, and non-invasive radiation to kill tumor cells. It is widely used in the radical curative treatment or palliative treatment of primary MT and metastatic tumors. Approximately 70% of patients with malignancy need radiation therapy at various stages of their treatment, of which 70% are radical radiotherapy. Radiotherapy accounts for about 40% of cured malignancies (14). The goal of radiotherapy is to maximize the radiation dose to the lesion (target) area for a long time and kill tumor cells while preventing or protecting surrounding normal tissues or organs from unnecessary radiation exposure, thus providing the required special protection to some vital organs such as the brainstem, spinal cord, kidney, gonads, *etc.* (15).

Traditional radiotherapy techniques, such as Co-60 teletherapy with poor precision and limited radiotherapy effectiveness, only achieve the primary stage of the radical curative treatment of the tumor, and also cause temporary or permanent damage to normal tissues and organs (16, 17). Interventional medicine has progressed particularly rapidly and led to the development and use of interventional radiotherapy techniques, in which radiopharmaceuticals are directly injected into the lesions through intubation and injection to enrich the concentration of the drugs in the lesions. This enables precise and targeted delivery of radiopharmaceuticals and overcomes the deficiency of traditional Co-60 teletherapy (18–20). However, there are still a series of problems to be solved in using radiopharmaceuticals in interventional radiotherapies, such as the selection of radiopharmaceuticals and their carriers, the uniform distribution and long-term fixation. In addition, interventional internal radiation therapy uses padding and manual implantation of radioactive drugs by doctors with their bare hands. This can prevent accurate calculations and evaluation of the injection pressure. Surgeons are thus exposed to radiation hazards. There is also an increase in patients with radioactive leaks and absorbed dose by non-target organs when the implantation operation time is too long and the injection pressure is too high. In addition, most implantation of radiopharmaceuticals is performed under computed tomography (CT) guidance. Although CT offers high resolution, it has problems such as being unable to be used in real-time dynamic navigation, with repeat punctures or offer precise delivery.

Precision interventional brachytherapy (PIBT) is a gradually developed radiotherapy technique that combines radiotherapy medicine with computer network technology, physics, *etc.* to address the limitations of conventional internal radiotherapy (21). For example, the most advanced radiotherapy equipment in vascular interventional robotic surgery is accurate to millimeters with very low side effects. This increases the accuracy of radiotherapy and thus application in clinical practice. Interventional doctors use catheters, guide wires, and other interventional devices to eliminate the heavy burden of lead protective aprons and reduce radiation exposure. Robot-assisted percutaneous coronary intervention (PCI) surgery can reduce radiation exposure by 97% (22). In addition, dispensing and fixation technologies for radiopharmaceuticals and drug carriers and precise delivery systems are active research topics. This paper reviews the main radiopharmaceuticals (^{131}I, ^{125}I, ^{177}Lu, etc.), drug carriers (Lipiodol, Microspheres, Hydrogels, etc.), dispensing and fixation technologies (SHIFT, Medrad, etc.), and interventional robotic precision delivery systems used in malignant tumor PIBT (**Figure 1**). It also discusses the current needs of the field and future development prospects.

COMMONLY USED RADIONUCLIDES

As early as 1901, Pierre and Marie Curie used small radium tubes for the first time to treat malignant tumors marking the birth of endoradiotherapy technology (23, 24). In 1970, Felix Mick developed a low-energy ^{125}I particle source containing iodine particles encapsulated in capsules and placed in a titanium tube. The ^{125}I was subsequently used for endoradiotherapy of prostate, liver, and lung cancer; its efficacy was clinically proven and widely recognized over the following decades (25–27). Versus conventional external radiotherapy, permanent ^{125}I seed implantation has its unique advantages. The first is that the release of x-rays, γ-rays, and other types of radiation is from the inside of the tumor tissue, causing DNA damage to tumor cells (28). As a result, the irradiation route does not need to pass through normal tissues to reach the target area. The dose distribution follows the inverse square law with an increase as distance decreases. Thus, the surrounding normal tissues are well protected, and the incidence of complications is low. Second, the local dose is high. The intensity of the implanted radioactive source is small, and the effective irradiation radius is short; thus, a higher radiation dose can be applied to the tumor target area (29).

The commonly used nuclides in nuclear medicine mainly include ^{125}I, ^{103}Pd, ^{169}Yb, ^{198}Au, ^{131}Cs, ^{137}Cs, ^{192}Ir, ^{60}Co, *etc.* (30). New nuclides such as ^{241}Am, ^{152}Cf, ^{26}Ra, and ^{145}Sm have recently attracted considerable attention. They have been used in clinical practice, but the most commonly used nuclides are ^{125}I and ^{103}PD, which have become essential to traditional external radiotherapy (31). The most common models of brachytherapy source models seeds are Pharma Seed BT-125-1 or BT-125-2 (Syncor Pharmaceuticals Inc, Golden, CO, USA), ADVANTAGE^{TM}Pd-103 IAPd-103A (IsoAid LLC, Port Richey, FL, USA), Prospera I-125-Med363 (North American Scientific, Inc., Chatsworth, CA, USA), Best® I-125 (Best Medical International, Inc., Springfield, VA, USA), and Type 6711 ^{125}I particles (HTA Co., Ltd., Beijing, China). The diameter of ^{125}I particles is 0.8 mm, and the length is 4.5 mm; the wall thickness of the enveloping titanium tube is 0.05 m (the source core is φ 0.5mm×3.0mm to adsorb ^{125}I silver rod, which is suitable for killing tumor cells with slow growth). Although radionuclide particle therapy for tumors has high safety and achieves sound therapeutic effects, there are still some constraints, such as selecting radionuclide particles for the precise treatment of tumors with different proliferation rates to obtain the maximum killing effect. Second, there are complications and

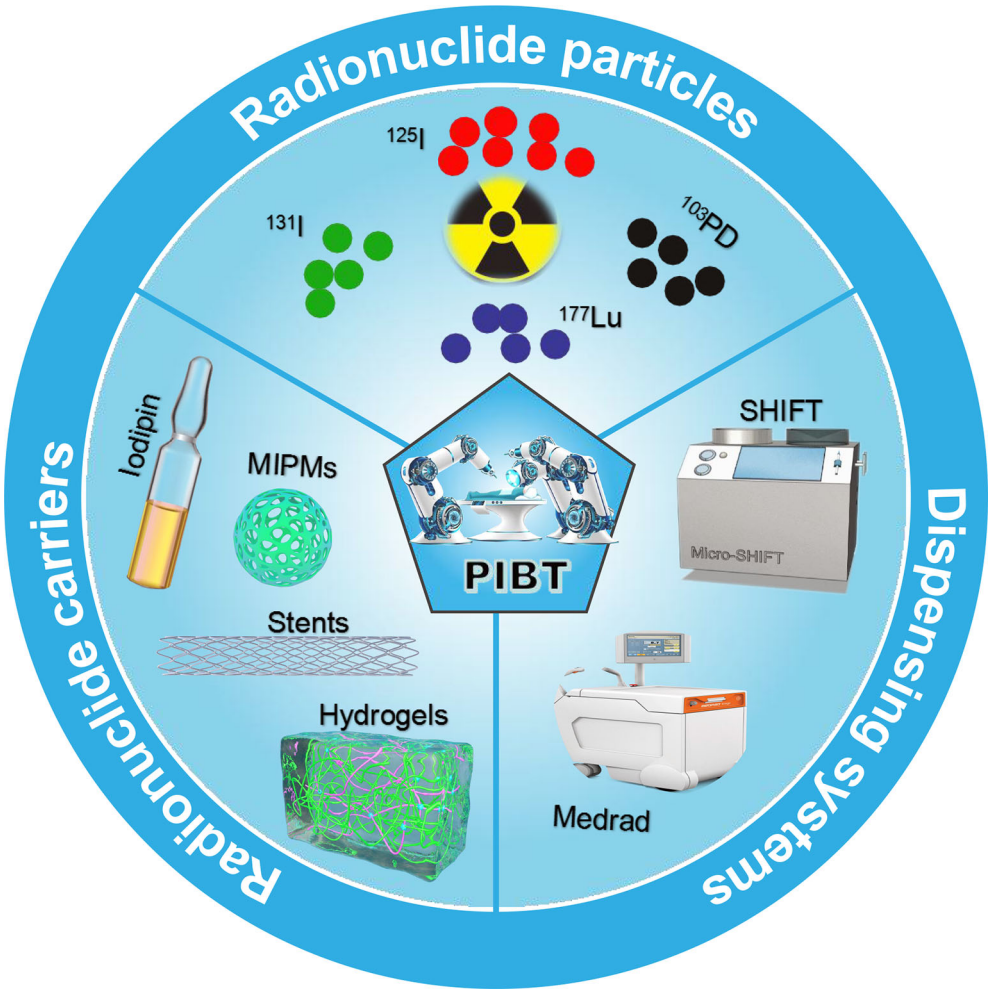


FIGURE 1 | Schematic showing radionuclide particles, radionuclide carriers, and dispensing systems commonly used in interventional brachytherapy.

adverse reactions after particle implantation. Finally, there is a need to study further the efficacy evaluation methods of particle implantation combined with external radiotherapy.

CARRIERS OF RADIONUCLIDES FOR INTERVENTIONAL BRACHYTHERAPY

Iodized Oil

Lipiodol is the most commonly used carrier for radionuclide drugs because it is easy to inject and selectively deposited. For example, ^{131}I-labeled lipiodol has been proven to be clinically effective and is commercially available (32). However, ^{131}I suffers from high-energy gamma photon emission (364 keV, 81%) (33), and the radioactivity yield of ^{131}I-labeled lipiodol is also poor. Due to its suitable decay properties (T½ =6.73 days, Eβ (Max) =0.49 MeV, Eγ=208 KeV [11%]), the half-life of ^{177}Lu is comparable to that of ^{131}I without significant decay loss. The relatively low abundance of low-energy gamma photons can be used for simultaneous scintillation imaging and dosimetry studies without a significant additional dose burden to the patient. Thus, it is a feasible substitute of ^{131}I in lipiodol for liver cancer radiotherapy (34, 35). However, water-soluble nuclide particles are difficult to disperse stably in lipiodol for a long time. For instance, Suresh et al. treated a rat orthotopic liver tumor model with ^{177}Lu-labeled lipophilic 8-hydroxy-quinoline mixed with lipiodol by a traditional method. They found that it was prone to radioactive leakage and deposited in bone tissue (36). This limited the clinical application of lipiodol with ^{177}Lu-labeled. Therefore, it is an important direction and hot topic for future research to develop efficient, simple, and stable lipiodol/nuclide preparations and obtain stable and long-term interventional radiotherapy.

Microspheres

Recently, some progress has been made in developing methods for the preparation of interventional radioactive microsphere

embolization materials (37). For example, Arranja et al. (38) dispersed solid acetylacetone holmium microspheres (HO2 (Acac) 3-MS) in NaH_2PO_4 or NaOH solutions and incubated at room temperature for 2 h to obtain two new inorganic microspheres. They then exchanged them with phosphate or hydroxyl ions through acetylacetone to obtain Ho(OH) 3-MS and Ho(OH) 4-MS. After preparing HopO4-MS and Ho(OH) 3MS, the stable isotope ^{166}Ho was partially converted into radioactive ^{166}Ho by neutron activation, and high activity radioactive microspheres were obtained. Zielhuis et al. (39) used elemental holmium combined with the carboxylic acid group of alginate polymer through electrostatic action to obtain alginate microspheres loaded with holmium. Finally, ^{166}Ho was added into calcium-hardened alginate microspheres to obtain microspheres with high radiochemical stability (94% after 48 h incubation in human serum). Ma et al. (40) performed ^{131}I labeling using gelatin microspheres as carriers, and through a study in a New Zealand rabbit liver model, found that the nuclides were aggregated in the liver in the form of microspheres after ^{131}I-GMSS administration. In addition, radioactivity was detected 48 days after injection of ^{131}IGMS, and the microspheres were degraded to different extents 24, 32, and 48 days after the injection of ^{131}I-GMSS. Although these microspheres offer high activity and degradability, they are primarily limited to basic research at the animal level, and few radioactive microspheres can be applied in human clinical practice.

The most commonly used clinical radiation microspheres are radioactive ^{90}Y microspheres. They can be injected into tumor lesions through digital subtraction angiography (DSA) super-selection, and the β-rays emitted by them can be used to kill tumors and perform endoradiotherapy (41–43). Therapheres and SiR-spheres are available in the market. They are safe and efficacious for the treatment of TM. Several side effects are associated with trans-arterial procedures (44, 45) including dissociation of cargo and formation of ectopic embolism. Furthermore, since ^{90}Y only emits beta rays and cannot be detected by single-photon emission computed tomography/positron emission tomography (SPECT/PET) imaging, it is difficult to obtain the drug distribution behavior *in vivo* by imaging techniques. There is a blind spot of the correlation between therapeutic effect and nuclide quantification. Thus, it is challenging to make unified clinical recommendations. ^{90}Y microsphere treatment is economically expensive, and thus widespread use of this treatment is quite limited.

Others

Other radiopharmaceutical carriers mainly include scaffolds, hydrogels, *etc.*, such as the intensity-modulated radiation-acrylic repositioning stent for the treatment of head and neck cancer reported by Lee Vsk et al. Retrospective cohort studies of patients with maxillary sinus, nasal, or oral cancer have revealed that acrylic repositioning stents do not alter radiotherapy outcomes and are highly stable (46). Zhu et al. (47) developed a biliary stent loaded with ^{125}I radioactive particles, and used a comparative clinical study of 23 patients to show that ^{125}I seeds in the biliary stent not only improved the patency of the patients' biliary tract, but also prolonged the patients' survival time.

Hydrogels have become a hot research topic in recent years due to their excellent biocompatibility, biodegradability, and outstanding clinical application value (48, 49). Hydrogel carriers have also played an essential role in the study of cancer brachytherapy. For instance, Schaal et al. (50) used radionuclide ^{131}I to label a thermal micelle composed of an elastin-like polypeptide (ELP) to form an *in situ* hydrogel brachytherapy of prostate cancer. The study was performed on a human PC-3M-Luc-C6 prostate tumor model and human BXPC3-Luc2 pancreatic tumor model and found that the ELP pool retained 52% and more than 70% of radioactivity for 60 days in prostate and pancreatic tumors, respectively. Furthermore, after 72 h, there was no significant accumulation of radioactivity in the tissue outside the target (≦0.1%ID); the median survival time of the two groups of nude mice was significantly extended.

Puente et al. (51) used an injectable chitosan hydrogel capable of releasing a chemotherapy drug (temozolomide, TMZ) while retaining a radioactive isotope preparation (iodine, ^{131}I) as the carrier of intracavity local radiotherapy and chemotherapy for the intraperitoneal therapy of brain gliomas. Some studies have shown that injectable chemical-radio-hydrogel implants can potentially improve local control and overall prognosis of invasive, poor-prognosis brain tumors. Although some studies suggest that hydrogels have many advantages, they are still limited to basic research at the animal level. These materials have not been clinically translated due to defects in drug delivery and biological behavior.

RADIOPHARMACEUTICAL DISPENSING AND FIXATION TECHNOLOGY

The precise implementation of interventional brachytherapy is a significant clinical problem. The community needs to improve the dispensing efficiency and drug stability while also reducing the radiation injury to medical staff. This is an inevitable trend for nuclear medicine: Replacing manual operation with intelligent equipment (52, 53). The central dispensing systems are UG-05 (Japan), Medrad (United States), IRIS (Italy). However, such equipment has a single type of dispensing and is expensive.

Zhang et al. successfully developed a PET molecular imaging probe microfluidic modular integrated synthesis system for the above problems. The system uses a modular microfluidic chip strategy to synthesize different positron emission tomography (PET) molecular image probes on an instrument and achieve the chemical purity and radioactive chemical purity of online controller drugs (54). The precision instrument equipment not only dramatically expands individualization and increases the accuracy of medical PET clinical applications, but also plays a vital role in the research and development of related nuclear drugs and radiation protection applications. However, this equipment is not suitable for stable mixing of radiopharmaceuticals and interventional embolic agents (such as lipiodol) commonly used in the clinic. Given this, Liu 's team (55, 56) from Xiamen

University developed a green, chemically free, super-stable homogeneous lipiodol formulation technology (SHIFT) (**Figure 2**). This technology makes the reactor reach the state of supercritical fluid by adjusting the temperature and pressure in the reactor. One can then adjust the physical parameters such as temperature and pressure to adjust the intermolecular force of the drug. This technology not only improves the solubility of drug molecules in lipiodol, but also achieves a homogeneous and stable state for several months to offer long-term fixation of radioactive drugs and interventional embolization agents.

PRECISION DELIVERY SYSTEM OF INTERVENTIONAL SURGERY ROBOTS

Interventional radiation therapy could be completed by an intelligent operating system. In fact, the manual operation based on experience is expected to be replaced by artificial intelligence. In order to solve the fundamental problems facing interventional radiation therapy such as intelligent precise delivery, current research mainly focuses on operator design choices, forced sensing information feedback, master-slave control methods, artificial intelligence algorithms, and the application of medical image analysis. Precision interventional robotic systems mainly include vascular interventional robot systems and particle-implantation robotic systems.

Vascular Interventional Robotic System

The Hansen Sensei® robotic system for percutaneous coronary intervention and percutaneous radiofrequency ablation was launched by Hansen Medical Inc. (Mountain View, CA, USA). It facilitates the entry and exit of the electrode conduit through the contact rolling of the friction wheel. It offers a circumferential rotation of the front end of the conduit by rotating the clamping device at the end of the conduit (57). For peripheral vascular interventional (PVI), a guidewire and a catheter should be used for drug injection. To address this, Hansen Medical Inc. updated its Magellan® robotic system by adding a set of friction band components in contact with each other. They use relative friction and rolling to achieve feed rotation of the guidewire. Versus the Artisan catheter, the Magellan system is an intelligent catheter that is more refined in diameter and has better angulation with tip force feedback. The major problem with this system is that the operation requires the use of a specific catheter, and the cost of a single operation is high (58).

The CorPath® GRX robot launched by Corindus Vascular Robotics Inc. (Waltham, MA, USA) is currently the only robot platform globally that can be used for PCI and PVI treatment at the same time. This robotic system manipulates the guidewire to complete the rotation and twist action of the feed by rolling and rotating the holding chamber through multiple sets of friction wheels. It then performs the rotation and feeds the guidewire through a gear transmission mechanism with position movement controlled by the manipulator's arm. However, the current

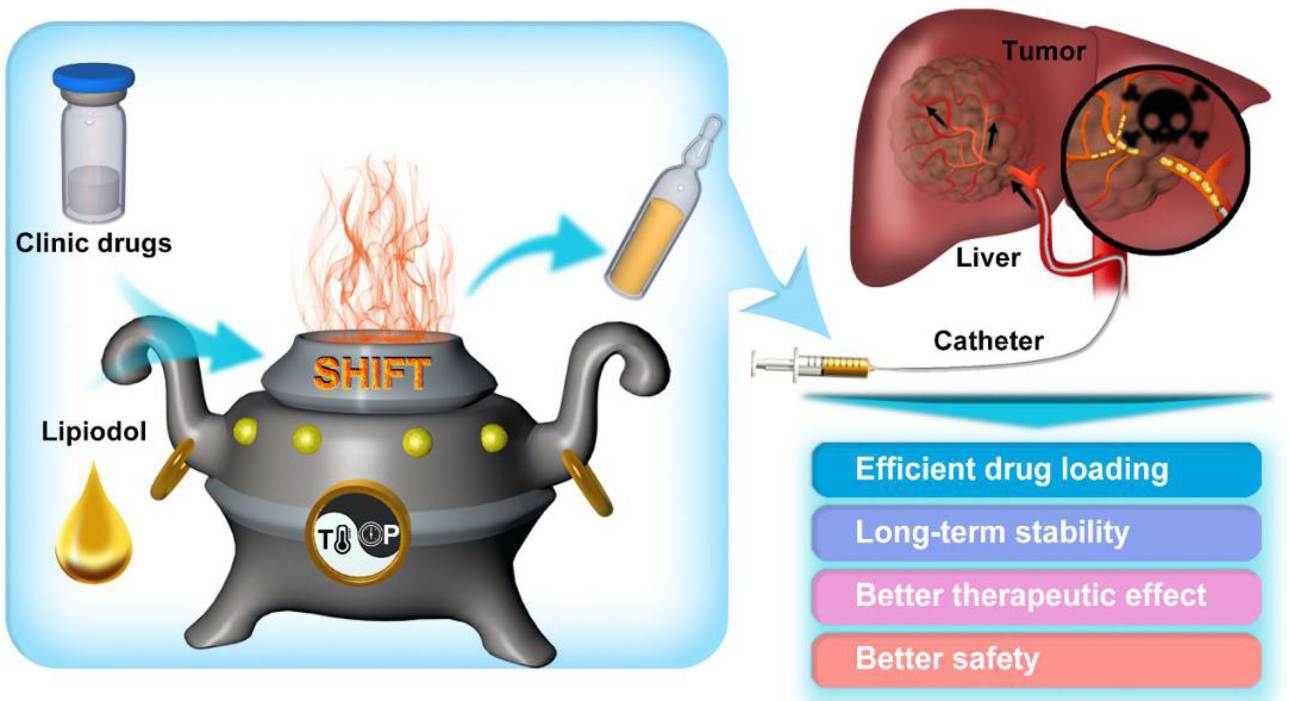


FIGURE 2 | Schematic illustration of super-stable homogeneous lipiodol formulation technology (SHIFT) as a revolutionary strategy for transhepatic arterial chemotherapy and embolization (TACE). The clinic drugs and lipiodol are introduced to develop formulations with SHIFT at a controlled temperature and pressure overcoming current challenges in the hepatocellular carcinoma (HCC) treatment with TACE. (Reprinted with permission from Liu et al. (56). Copyright Elsevier B.V).

CorPath® system still lacks the main end control mode. It can only facilitate the remote end speed of the guidewire tube delivery through the handle (22, 59).

In 2019, the French company Robocath SAS(Rouen France) launched the R-One® robotic system for remote cardiovascular interventional therapy. The system can be used for remote delivery of coronary stents in PCI procedures. The robot is designed with a hinged open-close holding pod similar to that of CorPath®. However, its main-end controller can only control the execution end speed and lacks force feedback design (60).

Yang et al. conducted in-depth research on artificial intelligence. Rafi-Tari found potential operation skills of interventional surgery through an artificial intelligence framework that made the operation of the surgical robot smoother and more stable. They eventually completed the surgical task in an experiment (61, 62). Chi et al. proposed using artificial intelligence to enable the interventional surgery robot to learn from the demonstration of the operation by experts to complete the operation independently or explore autonomously within the vascular model to try to accomplish the surgical goals. The experimental results showed that artificial intelligence could achieve this goal and a more accurate and smoother operation process than manual operation (63–66).

Particle Implantation Robot System

The Elekta-Nucletron FIRST system from Elekta-Nucletron AB (Stockholm, Sweden) includes an integrated real-time particle therapy system (FIRST™) with robot-assisted needle recovery and particle pushing devices. The system includes a computer-controlled three-dimensional (3D) transrectal ultrasound system, an integrated puncture and particle delivery device, and an integrated treatment planning system. The surgical robot was certified by the US FDA and Health Canada in 2001 and by the European Community (EC or CE) in 2002 for use only in treating prostate cancer with particle implantation (67).

The MIRA-V system (68, 69) was developed at the University of Western Ontario, Canada and is an ultrasound-guided minimally invasive robot-assisted particle implantation system for the lung. The robot carries lung dose planning software upgraded from the Prostate Particle Implantation Planning System to improve the accuracy of the execution plan (70). It also has an optical camera and a 5DOF electromagnetic tracer sensor that can monitor the position of the puncture needle tip. However, the system is still in the laboratory stage and has not been reported in clinical application. Recently, a multi-organ particle implantation surgical robot (Para-Brachyrob system) was developed by the Research Center for Industrial Robots Simulation and Testing (CESTER), Technical University of Cluj-Napoca (Cluj-Napoca, Romania) for high-dose-rate brachytherapy (71). It is still in the experimental stage and has not yet received US FDA or CE approval.

DISCUSSION

To summarize, research on precision interventional brachytherapy of malignant tumors has led to significant advances in the types, functions, choices, and quality of radionuclides and their carriers in radiotherapy. However, most nuclides and carriers with excellent performance are still in the basic research and animal study stage. Future efforts include optimizing the performance of existing nuclides and carriers, stabilizing the nuclides in the lesion area for a long time, improving the efficiency and safety of their use, and clinical applications.

In addition, as the key to the accurate delivery of interventional radiotherapy for malignant tumors, the interventional surgical robot has initially achieved image guidance at the technical level and realized preoperative planning, puncture, and drug configuration automatically or semi-automatically. However, clinical applications, to date, have been limited: (1) Indications are narrow. Most interventional surgical robots operate on patients with specific tumors, and surgical robots involving other malignant tumors are still in the laboratory stage. (2) The image guidance system is single-mode. Existing equipment mainly uses ultrasound, computed tomography (CT), or magnetic resonance (MR) imaging to guide implantation. There is no surgical robot with a multi-mode imaging system to guide the implantation process; thus, it is impossible to map the radiopharmaceutical distribution and pressure-gated feedback in real-time accurately. (3) The robotic system is not intelligent enough. It has not achieved the master-slave robot macro/micro composite drive or an operating system with multi-channel control feedbacks such as vision, force, and touch (haptics).

AUTHOR CONTRIBUTIONS

PH and SG drafted the article. ER, HWC, HC, YP, BLuo, and YX devised the concept for the mini-review along with BLi, JL, JM, and GL. All authors revised and contributed to the final version of the manuscript. PH, SG, BLi, JL, JM, and GL contributed with critical revision of the manuscript throughout the process. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Major State Basic Research Development Program of China (2017YFA0205201), the National Natural Science Foundation of China (81925019, 81603015, 81871404, 82003147, and U1705281), the Fundamental Research Funds for the Central Universities (20720190088 and 20720200019), the Nuclear Medicine and Molecular Imaging Key Laboratory of Sichuan province open project (nos.HYX20003), and the Chengdu Gaoxin Medical Association 2020 Annual Cancer Intervention Special Research Fund (2020S02).

REFERENCES

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* (2021) 71:209–49. doi: 10.3322/caac.21660
2. Fan W, Yung B, Huang P, Chen X. Nanotechnology for Multimodal Synergistic Cancer Therapy. *Chem Rev* (2017) 117:13566–638. doi: 10.1021/acs.chemrev.7b00258
3. Chu C, Lin H, Liu H, Wang X, Wang J, Zhang P, et al. Tumor Microenvironment-Triggered Supramolecular System as an In Situ Nanotheranostic Generator for Cancer Phototherapy. *Adv Mater* (2017) 29:10.1002/adma.20160592. doi: 10.1002/adma.201605928
4. Wyld L, Audisio RA, Poston GJ. The Evolution of Cancer Surgery and Future Perspectives. *Nat Rev Clin Oncol* (2015) 12:115–24. doi: 10.1038/nrclinonc.2014.191
5. Vatner RE, Cooper BT, Vanpouille-Box C, Demaria S, Formenti SC. Combinations of Immunotherapy and Radiation in Cancer Therapy. *Front Oncol* (2014) 4:325. doi: 10.3389/fonc.2014.00325
6. Liu Y, Bhattarai P, Dai Z, Chen X. Photothermal Therapy and Photoacoustic Imaging via Nanotheranostics in Fighting Cancer. *Chem Soc Rev* (2019) 48:2053–108. doi: 10.1039/c8cs00618k
7. Zhang Q, Bao C, Cai X, Jin L, Sun L, Lang Y, et al. Sonodynamic Therapy-Assisted Immunotherapy: A Novel Modality for Cancer Treatment. *Cancer Sci* (2018) 109:1330–45. doi: 10.1111/cas.13578
8. Yang Y, Yuan G, Zhan C, Huang Y, Zhao M, Yang X, et al. Benefits of Surgery in the Multimodality Treatment of Stage IIB-IIIC Small Cell Lung Cancer. *J Cancer* (2019) 10:5404–12. doi: 10.7150/jca.31202
9. Schwartsburd P. A View on Pathogenesis of ≪Vicious Cancer Progression Cycle≫. *Front Oncol* (2020) 10:690. doi: 10.3389/fonc.2020.00690
10. Quail DF, Joyce JA. Microenvironmental Regulation of Tumor Progression and Metastasis. *Nat Med* (2013) 19:1423–37. doi: 10.1038/nm.3394
11. Klein CA. Cancer Progression and the Invisible Phase of Metastatic Colonization. *Nat Rev Cancer* (2020) 20:681–94. doi: 10.1038/s41568-020-00300-6
12. Todd I. Fundamentals of Radiation Therapy and Cancer Chemotherapy. *Br J Cancer* (1975) 31:700. doi: 10.1038/bjc.1975.121
13. Martin OA, Martin RF. Cancer Radiotherapy: Understanding the Price of Tumor Eradication. *Front Cell Dev Biol* (2020) 8:261. doi: 10.3389/fcell.2020.00261
14. Lang JY. Review, Rellection and Prospect of Radiotherapy in China in the Past 30 Years. *J Canc Control Treat* (2017) 30:1–4.
15. Hellevik T, Martinez-Zubiaurre I. Radiotherapy and the Tumor Stroma: The Importance of Dose and Fractionation. *Front Oncol* (2014) 4:1. doi: 10.3389/fonc.2014.00001
16. Adamus-Górka M, Mavroidis P, Lind BK, Brahme A. Comparison of Dose Response Models for Predicting Normal Tissue Complications From Cancer Radiotherapy: Application in Rat Spinal Cord. *Cancers (Basel)* (2011) 3:2421–43. doi: 10.3390/cancers3022421
17. Huang CC, Chao PJ, Guo SS, Wang CJ, Luo HL, Su YL, et al. Developing a Multivariable Normal Tissue Complication Probability Model to Predict Late Rectal Bleeding Following Intensity-Modulated Radiation Therapy. *J Cancer* (2019) 10:2588–93. doi: 10.7150/jca.29606
18. Bibok A, Doros A. Role of Interventional Radiological Procedures in the Treatment of Liver Cancer. *Magy Onkol* (2018) 62:45–52.
19. Piron L, Cassinotto C, Guiu B. Prise En Charge Des Tumeurs Malignes Du Foie En Radiologie Interventionnelle [Interventional Radiology of Liver Tumors]. *Presse Med* (2019) 48:1156–68. doi: 10.1016/j.lpm.2019.10.010
20. Boyvat F. Interventional Radiologic Treatment of Hepatocellular Carcinoma. *Exp Clin Transpl* (2017) 15:25–30. doi: 10.6002/ect.TOND16
21. Yang WC, Hsu FM, Yang PC. Precision Radiotherapy for non-Small Cell Lung Cancer. *J BioMed Sci* (2020) 27:82. doi: 10.1186/s12929-020-00676-5
22. Granada JF, Delgado JA, Uribe MP, Fernandez A, Blanco G, Leon MB, et al. First-In-Human Evaluation of a Novel Robotic-Assisted Coronary Angioplasty System. *JACC Cardiovasc Interv* (2011) 4:460–5. doi: 10.1016/j.jcin.2010.12.007
23. Skowronek J. Current Status of Brachytherapy in Cancer Treatment-Short Overview. *J Contemp Brachyther* (2017) 9:581–9. doi: 10.5114/jcb.2017.72607
24. Dobrzynska MM. Maria Sklodowska-Curie, Her Life and Work-the 150 Anniversary of Her Birthday. *Rocz Panstw Zakl Hig* (2017) 68:309–12.
25. D'Addessi A, Racioppi M, Giustacchini M, Alcini A, Alcini E. ^{125}I Seeds Implantation Plus Pelvic Lymphadenectomy in the Management of Localized Prostate Cancer. *Minerva Urol Nefrol* (1995) 47:105–11.
26. Peng S, Yang QX, Zhang T, Lu MJ, Yang G, Liu ZY, et al. Lobaplatin-TACE Combined With Radioactive ^{125}I Seed Implantation for Treatment of Primary Hepatocellular Carcinoma. *Asian Pac J Cancer Prev* (2014) 15:5155–60. doi: 10.7314/apjcp.2014.15.13.5155
27. Cheng J, Ma S, Yang G, Wang L, Hou W. The Mechanism of Computed Tomography-Guided ^{125}i Particle in Treating Lung Cancer. *Med Sci Monit* (2017) 23:292–9. doi: 10.12659/msm.898526
28. Martin RF, Bradley TR, Hodgson GS. Cytotoxicity of an ^{125}I-Labeled DNA-Binding Compound That Induces Double-Stranded DNA Breaks. *Cancer Res* (1979) 39(8):3244–7.
29. Peschel RE, Colberg JW. Surgery, Brachytherapy, and External-Beam Radiotherapy for Early Prostate Cancer. *Lancet Oncol* (2003) 4:233–41. doi: 10.1016/s1470-2045(03)01035-0
30. Keisari Y, Kelson I. The Potentiation of Anti-Tumor Immunity by Tumor Abolition With Alpha Particles, Protons, or Carbon Ion Radiation and Its Enforcement by Combination With Immunoadjuvants or Inhibitors of Immune Suppressor Cells and Checkpoint Molecules. *Cells* (2021) 10:228. doi: 10.3390/cells10020228
31. Marwaha G, Macklis R, Singh AD, Wilkinson A. Brachytherapy. *Dev Ophthalmol* (2013) 52:29–35. doi: 10.1159/000351053
32. Ahmadzadehfar H, Biersack HJ, Ezziddin S. Radioembolization of Liver Tumors With Yttrium-90 Microspheres. *Semin Nucl Med* (2010) 40:105. doi: 10.1053/j.semnuclmed
33. De Ruyck K, Lambert B, Bacher K, Gemmel F, De Vos F, Vral A, et al. 188Re-HDD/lipiodol Therapy for Hepatocellular Carcinoma: An Activity Escalation Study. *Eur J Nucl Med Mol Imaging* (2004) 33:344.
34. Pillai MR, Chakraborty S, Das T, Venkatesh M, Ramamoorthy N. Production Logistics of ^{177}Lu for Radionuclide Therapy. *Appl Radiat Isot* (2003) 59:109. doi: 10.1016/s0969-8043(03)00158-1
35. Chakraborty S, Das T, Sarma HD, Venkatesh M, Banerjee S. Preparation and Preliminary Studies on ^{177}Lu-Labeled Hydroxyapatite Particles for Possible Use in the Therapy of Liver Cancer. *Nucl MedBiol* (2008) 35:589. doi: 10.1016/j.nucmedbio.2008.03.003
36. Subramanian S, Das T, Chakraborty S, Sarma HD, Banerjee S, Samuel G, et al. Preparation of ^{177}Lu-Labeled Oxine in Lipiodol as a Possible Agent for Therapy of Hepatocellular Carcinoma: A Preliminary Animal Study. *Cancer Biother Radiopharm* (2010) 25:539–43. doi: 10.1089/cbr.2010.0792
37. Sinha VR, Goyel V, Trehan A. Radioactive Microspheres in Therapeutics. *Pharmazie* (2004) 59:419–26.
38. Arranja AG, Hennink WE, Chassagne C, Denkova AG, Nijsen JFW. Preparation and Characterization of Inorganic Radioactive Holmium-166 Microspheres for Internal Radionuclide Therapy. *Mater Sci Eng C Mater Biol Appl* (2020) 106:110244. doi: 10.1016/j.msec.2019.110244
39. Zielhuis SW, Seppenwoolde JH, Bakker CJ, Jahnz U, Zonnenberg BA, van het Schip AD, et al. Characterization of Holmium Loaded Alginate Microspheres for Multimodality Imaging and Therapeutic Applications. *J BioMed Mater Res A* (2007) 82:892–8. doi: 10.1002/jbm.a.31183
40. Ma Y, Wan Y, Luo DH, Duan LG, Li L, Xia CQ, et al. Direct In Vivo Injection of ^{131}I-GMS and its Distribution and Excretion in Rabbit. *World J Gastroenterol* (2010) 16:2120–8. doi: 10.3748/wjg.v16.i17.2120
41. Subbiah V, Murthy R, Anderson PM. [90Y]Yttrium Microspheres Radioembolo-Therapy in Desmoplastic Small Round Cell Tumor Hepatic Metastases. *J Clin Oncol* (2011) 29:e292–4. doi: 10.1200/JCO.2010.32.4673
42. Brown KT, Do RK, Gonen M, Covey AM, Getrajdman GI, Sofocleous CT, et al. Randomized Trial of Hepatic Artery Embolization for Hepatocellular Carcinoma Using Doxorubicin-Eluting Microspheres Compared With Embolization With Microspheres Alone. *J Clin Oncol* (2016) 34:2046–53. doi: 10.1200/JCO.2015.64.0821
43. Mikell JK, Mahvash A, Siman W, Mourtada F, Kappadath SC. Comparing Voxel-Based Absorbed Dosimetry Methods in Tumors, Liver, Lung, and at the Liver-Lung Interface for (90)Y Microsphere Selective Internal Radiation Therapy. *EJNMMI Phys* (2015) 2:16. doi: 10.1186/s40658-015-0119-y
44. Helmberger T, Golfieri R, Pech M, Pfammatter T, Arnold D, Cianni R, et al. Clinical Application of Trans-Arterial Radioembolization in Hepatic

Malignancies in Europe: First Results From the Prospective Multicentre Observational Study CIRSE Registry for SIR-Spheres Therapy (CIRT). *Cardiovasc Intervent Radiol* (2021) 44:21–35. doi: 10.1007/s00270-020-02642-y
45. Riaz A, Awais R, Salem R. Side Effects of Yttrium-90 Radioembolization. *Front Oncol* (2014) 4:198. doi: 10.3389/fonc.2014.00198
46. Lee VSK, Nguyen CT, Wu J. The Fabrication of an Acrylic Repositioning Stent for Use During Intensity Modulated Radiation Therapy: A Feasibility Study. *J Prosthodont* (2019) 28:643–8. doi: 10.1111/jopr.13074
47. Zhu HD, Guo JH, Huang M, Ji JS, Xu H, Lu J, et al. Irradiation Stents vs. Conventional Metal Stents for Unresectable Malignant Biliary Obstruction: A Multicenter Trial. *J Hepatol* (2018) 68:970–7. doi: 10.1016/j.jhep.2017.12.028
48. Singhal A, Sinha N, Kumari P, Purkayastha M. Synthesis and Applications of Hydrogels in Cancer Therapy. *Anticancer Agents Med Chem* (2020) 20:1431–46. doi: 10.2174/1871521409666200120094048
49. Cai Y, Zheng C, Xiong F, Ran W, Zhai Y, Zhu HH, et al. Recent Progress in the Design and Application of Supramolecular Peptide Hydrogels in Cancer Therapy. *Adv Healthc Mater* (2021) 10:e2001239. doi: 10.1002/adhm.202001239
50. Schaal JL, Li X, Mastria E, Bhattacharyya J, Zalutsky MR, Chilkoti A, et al. Injectable Polypeptide Micelles That Form Radiation Crosslinked Hydrogels in Situ for Intratumoral Radiotherapy. *J Control Release* (2016) 228:58–66. doi: 10.1016/j.jconrel.2016.02.040
51. Puente P, Fettig N, Luderer MJ, Jin A, Shah S, Muz B, et al. Injectable Hydrogels for Localized Chemotherapy and Radiotherapy in Brain Tumors. *J Pharm Sci* (2018) 107:922–33. doi: 10.1016/j.xphs.2017.10.042
52. Wang C, Zhu X, Hong JC, Zheng D. Artificial Intelligence in Radiotherapy Treatment Planning: Present and Future. *Technol Cancer Res Treat* (2019) 18:1533033819873922. doi: 10.1177/1533033819873922
53. Akmal JS, Salmi M, Mäkitie A, Björkstrand R, Partanen J. Implementation of Industrial Additive Manufacturing: Intelligent Implants and Drug Delivery Systems. *J Funct Biomater* (2018) 9:41. doi: 10.3390/jfb9030041
54. Lei M, Pan JZ, Xu GM, Du PZ, Tian M, Zhang H. Automated Microfluidic Chip System for Radiosynthesis of PET Imaging Probes. *J Zhejiang Univ Sci B* (2019) 20:865–7. doi: 10.1631/jzus.B1900535
55. Chen H, Cheng HW, Wu WY, Li DF, Mao JS, Chu CC, et al. The Blooming Intersection of Transcatheter Hepatic Artery Chemoembolization and Nanomedicine. *Chin Chem Lett* (2020) 31:1375–81. doi: 10.1016/j.cclet.2020.03.024
56. Cheng HW, Yang XM, Liu G. Superstable Homogeneous Iodinated Formulation Technology: Revolutionizing Transcatheter Arterial Chemoembolization. *Sci Bull* (2020) 65:1685–7. doi: 10.1016/j.scib.2020.06.029
57. Saliba W, Cummings JE, Oh S, Zhang Y, Mazgalev TN, Schweikert RA, et al. Novel Robotic Catheter Remote Control System: Feasibility and Safety of Transseptal Puncture and Endocardial Catheter Navigation. *J Cardiovasc Electrophysiol* (2006) 17:1102–5. doi: 10.1111/j.1540-8167.2006.00556.x
58. Riga CV, Bicknell CD, Rolls A, Cheshire NJ, Hamady MS. Robot-Assisted Fenestrated Endovascular Aneurysm Repair (FEVAR) Using the Magellan System. *J Vasc Interventional Radiol* (2013) 24:191–6. doi: 10.1016/j.jvir.2012.10.006
59. Al Nooryani A, Aboushokka W. Rotate-On-Retract Procedural Automation for Robotic-Assisted Percutaneous Coronary Intervention: First Clinical Experience. *Case Rep Cardiol* (2018) 2018:6086034. doi: 10.1155/2018/6086034
60. Buzurovic I, Podder TK, Yu Y. Prediction Control for Brachytherapy Robotic System. *J Robot* (2010) 2010:10. doi: 10.1155/2010/581840
61. Rafii-Tari H, Liu J, Payne CJ, Bicknell C, Yang GZ. Hierarchical HMM Based Learning of Navigation Primitives for Cooperative Robotic Endovascular Catheterization. *Med Image Comput Comput Assist Interv* (2014) 17:496–503. doi: 10.1007/978-3-319-10404-1_62
62. Rafii-Tari H, Liu J, Lee SL, Bicknell C, Yang GZ. Learning-Based Modeling of Endovascular Navigation for Collaborative Robotic Catheterization. *Med Image Comput Comput Assist Interv* (2013) 16:369–77. doi: 10.1007/978-3-642-40763-5_46
63. Chi W, Dagnino G, Kwok TM. Collaborative Robot-Assisted Endovascular Catheterization With Generative Adversarial Imitation Learning; Proceedings of the 2020 IEEE International Conference on Robotics and Automation (ICRA). *IEEE* (2020). doi: 10.1109/ICRA40945.2020.9196912
64. Chi W, Liu J, Rafii-Tari H, Riga C, Bicknell C, Yang GZ, et al. Learning-Based Endovascular Navigation Through the Use of Non-Rigid Registration for Collaborative Robotic Catheterization. *Int J Comput Assist Radiol Surg* (2018) 13:855–64. doi: 10.1007/s11548-018-1743-5
65. Chi W, Liu J, Abdelaziz ME. Trajectory Optimization of Robot-Assisted Endovascular Catheterization With Reinforcement Learning; Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). *F IEEE* (2018).
66. Rafii-tari H, Payne CJ, Yang G-Z. Current and Emerging Robot-Assisted Endovascular Catheterization Technologies: A Review. *Ann Biomed Eng* (2014) 42:697–715. doi: 10.1007/s10439-013-0946-8
67. Van Gellekom MP, Moerland MA, Wijrdeman HK, Battermann JJ. Quality of Permanent Prostate Implants Using Automated Delivery With Seedselectron Versus Manual Insertion of RAPID Strands. *Radiother Oncol* (2004) 73:49–56. doi: 10.1016/j.radonc.2004.08.004
68. Hungr N, Troccaz J, Zemiti N, Tripodi N. Design of an Ultrasound-Guided Robotic Brachytherapy Needle-Insertion System. *Conf Proc IEEE Eng Med Biol Soc* (2009) 2009:250–3. doi: 10.1109/IEMBS.2009.5333801
69. Podder TK, Beaulieu L, Caldwell B, Cormack RA, Crass JB, Dicker AP, et al. AAPM and GEC-ESTRO Guidelines for Image-Guided Robotic Brachytherapy: Report of Task Group 192. *Med Phys* (2014) 41:101501. doi: 10.1118/1.4895013
70. Lin AW, Trejos AL, Mohan S, Bassan H, Kashigar A, Patel RV, et al. Electromagnetic Navigation Improves Minimally Invasive Robot-Assisted Lung Brachytherapy. *Comput Aided Surg* (2018) 13:114–23. doi: 10.3109/10929080801969725
71. Popescu T, Kacsó AC, Pisla D, Kacsó G. Brachytherapy Next Generation: Robotic Systems. *J Contemp Brachyther* (2015) 7:510–4. doi: 10.5114/jcb.2015.56769

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: 12 October 2021
doi: 10.3389/fonc.2021.706034

Variations in Head and Neck Treatment Plan Quality Assessment Among Radiation Oncologists and Medical Physicists in a Single Radiotherapy Department

Elisabetta Cagni[1,2], Andrea Botti[1†], Linda Rossi[3†], Cinzia Iotti[4], Mauro Iori[1], Salvatore Cozzi[4], Marco Galaverni[4], Ala Rosca[4], Roberto Sghedoni[1], Giorgia Timon[4], Emiliano Spezi[2] and Ben Heijmen[3]*

[1] Medical Physics Unit, Azienda Unità Sanitaria Locale Istituto di Ricovero e Cura a Carattere Scientifico (USL-IRCCS) di Reggio Emilia, Reggio Emilia, Italy, [2] School of Engineering, Cardiff University, Cardiff, United Kingdom, [3] Department of Radiation Oncology, Erasmus MC Cancer Institute, Rotterdam, Netherlands, [4] Radiotherapy Unit, Azienda USL-IRCCS di Reggio Emilia, Reggio Emilia, Italy

OPEN ACCESS

Edited by:
Jose Eduardo Villarreal Barajas, Royal Devon and Exeter Hospital, United Kingdom

Reviewed by:
Samantha Warren, Royal Cornwall Hospital Trust, United Kingdom
Gage Redler, Moffitt Cancer Center, United States

***Correspondence:**
Elisabetta Cagni
elisabetta.cagni@ausl.re.it

[†]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: *06 May 2021*
Accepted: *30 August 2021*
Published: *12 October 2021*

Citation:
Cagni E, Botti A, Rossi L, Iotti C, Iori M, Cozzi S, Galaverni M, Rosca A, Sghedoni R, Timon G, Spezi E and Heijmen B (2021) Variations in Head and Neck Treatment Plan Quality Assessment Among Radiation Oncologists and Medical Physicists in a Single Radiotherapy Department. Front. Oncol. 11:706034. doi: 10.3389/fonc.2021.706034

Background: Agreement between planners and treating radiation oncologists (ROs) on plan quality criteria is essential for consistent planning. Differences between ROs and planning medical physicists (MPs) in perceived quality of head and neck cancer plans were assessed.

Materials and Methods: Five ROs and four MPs scored 65 plans for in total 15 patients. For each patient, the clinical (CLIN) plan and two or four alternative plans, generated with automated multi-criteria optimization (MCO), were included. There was always one MCO plan aiming at maximally adhering to clinical plan requirements, while the other MCO plans had a lower aimed quality. Scores were given as follows: 1–7 and 1–2, not acceptable; 3–5, acceptable if further planning would not resolve perceived weaknesses; and 6–7, straightway acceptable. One MP and one RO repeated plan scoring for intra-observer variation assessment.

Results: For the 36 unique observer pairs, the median percentage of plans for which the two observers agreed on a plan score (100% = 65 plans) was 27.7% [6.2, 40.0]. In the repeat scoring, agreements between first and second scoring were 52.3% and 40.0%, respectively. With a binary division between unacceptable (scores 1 and 2) and acceptable (3–7) plans, the median inter-observer agreement percentage was 78.5% [63.1, 86.2], while intra-observer agreements were 96.9% and 86.2%. There were no differences in observed agreements between RO–RO, MP–MP, and RO–MP pairs. Agreements for the highest-quality, automatically generated MCO plans were higher than for the CLIN plans.

Conclusions: Inter-observer differences in plan quality scores were substantial and could result in inconsistencies in generated treatment plans. Agreements among ROs were not

better than between ROs and MPs, despite large differences in training and clinical role. High-quality automatically generated plans showed the best score agreements.

Keywords: radiotherapy treatment planning, head and neck cancer, subjective plan quality assessment, inter-observer and intra-observer variation, automated treatment planning

1 INTRODUCTION

Advanced radiotherapy delivery approaches such as intensity-modulated radiation therapy (IMRT) and volumetric modulated arc therapy (VMAT) have substantially increased opportunities for sparing organs at risk (OARs) with proven clinical impact (1–5). Ideally, for each individual patient, the applied treatment plan maximally exploits the full potential of the applied delivery technique. Currently, most treatment plans are generated with interactive trial-and-error planning ("manual planning"). It is well-known that plan quality in manual planning may be suboptimal, e.g., depending on experience and ambition of the planner, and on allotted planning time (6, 7). In recent years, several systems for automated plan generation have been developed, often resulting in enhanced plan quality compared with manual planning (8–13).

In both manual and automated planning, human evaluation and judgment of treatment plans are crucial. Normally, plans are produced by medical physicists (MPs) or dosimetrists and presented to treating radiation oncologists (ROs) for approval. During manual plan generation, planners usually develop a range of (intermediate) plans, but generally only a single plan or sometimes two competing plans are discussed with the RO. Prior to approval, the RO may request for adaptation of presented plans. A necessary assumption for this workflow to work well is that (unknown) disparity between planners and ROs on characteristics of good/optimal plans is absent or minor. In case of large disparity, a plan with high quality from the planner's point of view may be presented to the RO, while a different plan with lower quality according to the planner, but clearly more attractive to the RO if she/he would have been aware of it, is intentionally not generated or presented. In such cases, there is no guarantee that plan modifications are requested and, if requested, to what extent the adapted plans would satisfy the needs of the RO.

In this study, we have systematically investigated differences between five ROs and four planning MPs, all working in a single radiotherapy department, in perceived quality of head and neck (HN) cancer plans. With the use of automated planning, multiple plans were generated per patient. Plan quality was scored using visual analogue scales.

2 MATERIAL AND METHODS

2.1 Patients and Clinical Treatment Plans

Planning CT data, contoured structures, and the clinical (CLIN) plan of 15 arbitrarily selected oropharyngeal HN cancer patients, recently treated with radiotherapy at Azienda USL-IRCCS Hospital (AUSL) of Reggio Emilia (Italy), were included in this study. Following American Joint Committee on Cancer (AJCC) TNM staging (14), six patients were classified as T2N2, three as T1N2, three as T2N1, and three as T4N2. Bilateral neck was irradiated in all patients. A simultaneous integrated boost (SIB) technique was used for all patients, delivering the prescribed doses in 33 daily fractions. Total doses for PTVhigh, PTVmedium, and PTVlow were 69.96, 59.4, and 54 Gy, respectively (15–17). For each planning target volume (PTV), the goal was to deliver 100% of the prescribed dose to 95% of the volume. All plans were normalized so that exactly 95% of PTVhigh received the prescription dose. Sizes of the involved PTVs were as follows: 178.5 ± 97.3 cm^3 [63.3, 409.6], 208.4 ± 105.7 cm^3 [39.8, 431.7], and 184.8 ± 51.0 cm^3 [95.2, 248.7] for PTVhigh, PTVmedium, and PTVlow, respectively. OARs considered in planning were spinal cord, brainstem, left and right parotid, esophagus, oral cavity, larynx, mandible, pharyngeal constrictor muscles, and submandibular glands (17). Plans were generated using the following priorities for achieving planning objectives: 1) sparing of brainstem, optic chiasm, and spinal cord (so higher priority than PTV coverage); 2) achievement of PTV dose objectives in the order PTVhigh, PTVmedium, and PTVlow; 3) parotid gland sparing; and 4) sparing of other OARs and healthy tissues. The clinical planning protocol was largely in line with international protocols, such as RTOG (18–21) and JAVELIN protocols (22).

Patients were treated with 3-arc 6-MV VMAT delivered with a TrueBeam STx linac (Varian Medical Systems, Palo Alto, USA) (10 patients) or using TomoTherapy (Accuray Inc., Sunnyvale, USA) (five patients). Clinical planning was performed with the Eclipse Treatment Planning System (TPS) v. 13 (Varian Medical Systems, Palo Alto, USA) or Tomoplan v. 3-4 (Accuray Inc., Sunnyvale, USA).

2.2 Global Study Design

Apart from the CLIN plan, two (for five patients) or four (for 10 patients) additional VMAT plans were evaluated in this study, resulting in a total of 65 evaluable plans. The extra plans had variable plan quality and were generated with automated planning (details in Section 2.5). Each of the 65 available plans was evaluated by five departmental ROs (three with more than 5 years of experience in HN radiotherapy and two with less than 1 year of experience) and four MPs (all with more than 5 years of experience), resulting in a total of 585 subjective plan evaluations. These involved ROs and MPs represented all involved staff in HN treatment in our department at the time of the study.

For each patient, every observer independently gave a score to each of the 3 or 5 available plans in a single session (details in Section 2.3). Scoring was blinded; i.e., observers did not know

how the plans were generated. Apart from giving a quality score to each plan, observers were also asked what change they considered most desirable for improvement of the plan (without knowing whether this would be feasible or not); see also Section 2.3.

To assess intra-observer variability in quality scoring, one RO and one MP performed the entire scoring process for 65 plans a second time, with a delay of at least a month. Previous results were blinded.

2.3 Plan Scoring Procedure

For each patient, all available dose distributions were simultaneously imported into the Eclipse TPS and linked to a virtual plan without any mention of the original delivery approach (VMAT or TomoTherapy), plan geometry, machine parameters, etc. With all plans simultaneously open, the observer gave a separate 1–7 score to each plan, following the routine procedure for plan evaluation (inspection of 3D dose distribution, dose–volume histogram (DVH) data, etc.), with higher scores pointing at perceived higher quality: 1–2, unacceptable (plan category 1); 3–5, acceptable if further planning would not have resulted in a better plan (this planning was not performed in this study) (plan category 2); and 6–7, acceptable, no further planning needed (plan category 3). A 7-point scale was chosen because of good performance in psychometric literature (23–25). In the remainder of this paper, the 1–7 scores are denoted "raw" scores, while plan categories 1–3 define the more intuitive "category" scores. The applied division of the raw scores in categories was made before the start of subjective plan scoring. As is visible in **Supplementary Figure S1**, this division was also explicitly shown to the observers while giving scores to plans. For the analyses, another scoring system was introduced as well, the so-called "binary" scoring system: raw scores 1 and 2 were grouped as binary score 0 (plan is unacceptable), and raw scores 3–7 were given binary score 1 (plan is in principle acceptable).

To express the most urgent need for plan improvement, the observers could choose from A) PTVs (coverage, conformity, and homogeneity), B) OAR group 1 (spinal cord, brainstem, and optical system), C) OAR group 2 (parotids, mandible, oral cavity, larynx, and esophagus), D) unspecified normal tissue, or E) none. See also **Supplementary Figure S1**.

2.4 Evaluation of Inter-Observer Differences in Plan Scoring

With nine observers, there were in total 36 unique combinations of two observers, here designated as "pairs". To analyze inter-observer differences in perceived plan quality, for all these observer pairs, percentages of agreement and disagreement in the scores given to the 65 evaluated plans were established. Analyses were partially based on raw scores, category scores, and binary scores. Observed percentages of agreement in RO–RO pairs and MP–MP pairs were compared with percentages of agreement in RO–MP pairs. Suggested most desired plan improvements were used to generate for each observer separately a frequency analysis of provided suggestions for the 65 evaluated plans.

2.5 Automatically Generated MCOa and MCOx Plans

Autoplans were generated with the Erasmus-iCycle system for fully automated multi-criteria optimization (MCO) (10, 26). Plan optimization in Erasmus-iCycle is based on so-called wish-lists (WLs), containing hard planning constraints and planning objectives with goal values and assigned priorities. A dedicated WL is needed for every treatment site. In essence, the WL defines an optimization protocol for automated multi-criteria generation of a single Pareto-optimal treatment plan for each patient. The aim in WL creation is to maximally ensure the highest clinical quality of the generated Pareto-optimal plans, in line with the clinical planning protocol and tradition [Appendix 10]. Also, in this study, such a WL was created with input of all ROs and MPs involved in the study (WLa). In the remainder of the paper, plans generated with WLa are denoted as "MCOa." These MCOa plans consisted of 23 equi-angular IMRT beams, with high similarity to VMAT and avoiding time for segmentation (27–29). With WLa as a starting point, 20 alternative WLs, "WLx" (x = b, c, d, …), were created for generation of "MCOx" plans. The WLx were derived from WLa by randomly varying the priorities of PTVmedium and PTVlow objectives and of the OARs. For generation of an MCOx plan for a patient, one of the 20 WLx was randomly selected; and in addition, the number of beams was randomly varied between 10 and 23. As for WLa, the 20 WLx enforced adherence to the hard planning constraints for brainstem, optic chiasm, and spinal cord, as in clinical planning (above). At the same time, the WLx allowed generation of MCOx plans with a spread in dosimetric differences compared with the corresponding MCOa plans. For patients 1–10, the CLIN plan was supplemented with the MCOa plan and three MCOx plans (in total five evaluable plans). For patients 11–15, apart from the CLIN and MCOa plan, there was one additional MCOx plan used in this study (three evaluable plans in total). The switch from five to three plans is discussed in Section 4. For putting the subjective scoring of plan quality by observers in context, dosimetric characteristics of CLIN, MCOa, and MCOx plans were analyzed by mutual comparisons of dosimetric plan parameters and DVHs.

2.6 Statistical Analysis

Shapiro's test and Student's t-test were used to assess the normality of distributions and statistical significance of dosimetric differences between plans generated with different planning approaches, i.e., CLIN, MCOa, and MCOx. Wilcoxon's two-sided signed-rank tests were used to assess statistical significance of mean score differences between CLIN, MCOa, and MCOx. Differences were considered significant if $p < 0.05$.

To assess statistical significance (0.05 level) of observed percentages of agreement for the 65 plan scores of the two observers in an observer pair, binomial distributions were used to calculate probabilities of percentage agreements in case of complete uncorrelated (random) choices of the two observers in a pair. To this end, success probabilities p of 1/7, 1/3, and 1/2 were used for raw, category, and binary scores, respectively.

The percentages of agreement in plan scores between the two observers in observer pairs were also analyzed with Cohen's coefficient (K) (30). The relative strength of agreement between the two observers in a pair is dependent on the calculated K-value. Landis and Koch (31) have proposed the following classification: $K < 0$, agreement "poor"; $0 \le K \le 0.2$, agreement "slight"; $0.2 < K \le 0.4$, agreement "fair"; $0.4 < K \le 0.6$, agreement "moderate"; $0.6 < K \le 0.8$, agreement "substantial"; and $0.8 < K \le 1$, agreement "almost perfect". For binary scoring, the resulting number of samples for unapproved status was not enough to achieve significant confidence limits in Cohen's coefficients for many evaluators (32). Therefore, Cohen's analyses were only performed for raw and category scores.

One-way ANOVA tests were performed to assess statistical significance of differences in percentages of agreement between subgroups of observers, 1) only RO–RO, 2) only MP–MP, and 3) only RO–MP pairs, after having assessed the normality of the distribution with the Kolmogorov–Smirnov test. Bartlett's test was used to test the homogeneity of variance. When ANOVA assumptions were not met, the Kruskal–Wallis rank sum test was used as non-parametric alternative to one-way ANOVA. The Wilcoxon's signed-rank test was used to test agreement differences between CLIN and MCOa plans, expert and no-expert ROs, and three and five evaluated plans per patient.

3 RESULTS

3.1 Differences Between Evaluated Clinical, MCOa, and MCOx Plans in Dosimetry

In panels a) and c) of **Supplementary Figure S2**, median DVHs for the CLIN, MCOa, and MCOx plans are presented, showing for each dose the corresponding median volume in the considered plans. For individual patients, the DVH differences between the CLIN, MCOa, and MCOx plans were pairwise quantified by generating differential DVHs: volume differences as a function of dose. Median volume differences and 10th and 90th percentiles are presented in panels b) and d) of **Supplementary Figure S2**. The 10th and 90th percentile curves point at large inter-patient variations in DVH differences between CLIN, MCOa, and MCOx plans. **Supplementary Table S1** shows how the DVH differences translate in differences in dosimetric plan parameters. Only few of the differences between CLIN, MCOa, and MCOx plan parameters were statistically different, while ranges were very broad. This is in line with the observations in **Supplementary Figure S2**. **Supplementary Figure S3** presents for each of the 15 study patients separately an overview of the dosimetric differences between the included three to five treatment plans.

3.2 Scoring for an Example Patient

To introduce the type of scoring data obtained for each patient, **Figure 1** shows the raw scores of the nine observers for the CLIN, MCOa, and MCOx plans of study patient 15, a patient showing large scoring variations. The majority of observers (6/9) selected MCOa as the best plan, while MCOx was selected most as the worst plan (5/9). This ranking of MCOa and MCOx is in line with the applied WLs for generation of these plans (Section 2.5). However, for all three plans, there were large inter-observer differences in raw scores (2–5 for MCOx and 2–6 for CLIN and MCOa). RO4 scored the clinically delivered CLIN plan as unacceptable, while for MP1, this plan was acceptable without further planning attempts. For RO3, MCOa was unacceptable, while for MP2, it could be delivered straightaway. **Figure 1** also shows large inter-observer differences in score ranges. As demonstrated in the group analyses below, large scoring variations were observed for all patients and the vast majority of plans.

3.3 Radiation Oncologist Experience in Head and Neck Radiotherapy and Scoring

As mentioned in Section 2.2, three participating ROs had more than 5-year experience in HN radiotherapy, while the other two had less than 1-year experience for this tumor site. When considering the raw, category, and binary scores of all 65 plans, median values for all five ROs/only three expert ROs were 28.5%/36.9% ($p = 0.5$), 56.2%/61.6% ($p = 1.0$), and 75.4%/75.4% ($p = 0.7$), respectively. Based on these observations, it was decided that in further group analyses, the five ROs in this study were considered as a single group.

3.4 Differences Between Clinical, MCOa, and MCOx Plans in Observer Scores

Table 1 reports differences between CLIN, MCOa, and MCOx in subjective scores, complimentary to the dosimetric differences in **Supplementary Table S1**. The automatically generated MCOa plans outperformed the clinically delivered CLIN plans, but for the binary scores, this was not statistically significant. Score differences were overall the largest between MCOa and MCOx and with the smallest p-values, with the former showing the highest scores, as to be expected from the respective WLs used for automated plan generation (Section 2.5).

3.5 Inter-Observer Variability in Plan Quality Scores

In line with the observations for patient 15 (above), for the majority of plans, inter-observer variations in assigned scores were large (**Figure 2**). For the 65 evaluated plans, the average standard deviation (SD) for the nine raw observer scores was 1.06 [0.33, 1.56] (**Figure 2A**). For 29 of the 65 plans, all category scores (1, 2, and 3) were present in the nine scores (**Figure 2B**). For 15/65 plans, there was at least one observer that scored category 3 (acceptable without further planning attempts), while at the same time), there were also observers that considered the plan unacceptable (category 1). Considering all 65 plans, the median percentage of plans declared unacceptable by an observer was 18.8% ± 8.6% [6.2%, 35.4%]. For CLIN, MCOa, and MCOx plans separately, these percentages were 14.8% ± 9.9% [0.0, 33.3], 4.4% ± 4.7% [0.0, 13.3], and 26.7% ± 12.3% [8.6, 48.6], respectively. Kruskal–Wallis rank tests resulted in a statistically significant difference, with $p = 0.005$. Wilcoxon's signed-rank test

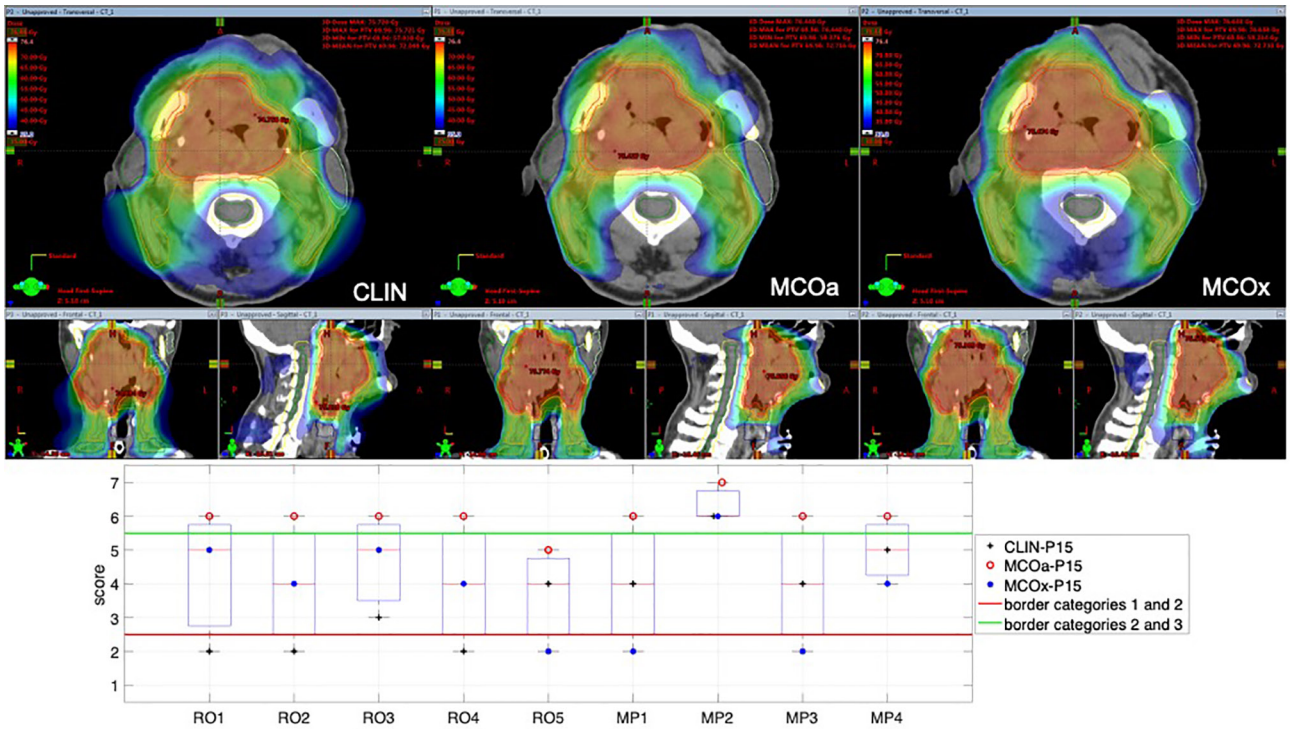


FIGURE 1 | Differences in subjective plan scores among the nine observers in the study, illustrated with an example patient (patient 15). Top panels: dose distributions of the evaluated CLIN, MCOa, and MCOx plans in axial, sagittal, and frontal planes. Bottom panel: corresponding subjective plan quality scores for the CLIN, MCOa, and MCOx plans of patient 15 for each of the nine observers, five ROs (RO1–RO5), and four MPs (MP1–MP4). Plans below the horizontal red line are considered unacceptable (raw scores 1 and 2, category 1). Above the green line are plans that can straightaway be delivered without any attempt to further improve the plan (raw scores 6 and 7, category 3). In the middle are the plans that are acceptable if further planning would not result in significant improvements (raw scores 3–5, category 2). CLIN, clinical; RO, radiation oncologist; MP, medical physicist.

showed a statistically significant difference between MCOa and MCOx ($p = 0.005$), while for CLIN *vs.* MCOa, $p = 0.1$, and for CLIN *vs.* MCOx, $p = 0.2$.

Figures 3A–C show unique pairs of two observers, the percentages of plans for which they agreed in a plan score. Considering all 36 unique observer pairs in this study, the median percentage of agreement in raw plan scores was 27.7% [6.2, 40.0] ("all" boxplot in **Figure 3A**). In case of complete randomness in the scoring of two observers in a pair, an agreement percentage of 14.3% would be expected (horizontal solid line in gray zone). For category (**Figure 3B**) and binary scores (**Figure 3C**), these median percentages were 58.5% [35.4, 73.8] (33.3% expected in case of randomness) and 78.5% [63.1, 86.2] (50% in case of randomness), respectively. The vast majority of percentages of agreement in **Figures 3A–C** are outside the gray zones, meaning that they are statistically significantly different from the corresponding expected values for random scoring, indicated by the horizontal solid lines. With one-way ANOVA p-values of 0.3, 0.6, and 0.4, there were no differences between the observer pair subgroups RO–RO, MP–MP, and RO–MP in the agreement distributions in **Figures 3A–C**, respectively.

Cohen's coefficient analyses for raw scores (**Figure 3D**) resulted in median K-values [range] of 0.46 [0.12, 0.68] when considering all observer pairs, 0.47 [0.17, 0.56] for ROs, 0.51 [0.33, 0.64] for MPs, and 0.46 [0.12, 0.68] for RO–MP. Following the labelling by Landis and Koch (M&M), the overall agreement is "moderate." More in detail, considering all 36 observer pairs, 11% (N = 4) resulted in slight agreement, 25% (N = 9) in fair agreement, 47% (N = 17) in moderate agreement, and 17% (N = 6) in substantial agreement. For category score analyses (**Figure 3E**), Cohen's median K-values [range] were 0.40 [0.03, 0.66] for all, 0.35 [0.04, 0.53] for ROs, 0.44 [0.37, 0.54] for MPs, and 0.39 [0.03-0.66] for RO–MP pairs. The overall agreement, in Landis and Koch scale, resulted in "fair"; 19% (N = 7) resulted in slight agreement, 31% (N = 11) in fair agreement, 47% (N = 17) in moderate agreement, and 3% (N = 1) in substantial agreement.

Figures 3F–H present scoring agreements for CLIN and MCOa plans separately, showing substantially better agreements for the automatically generated MCOa: when considering all 36 observer pairs, agreement percentages for CLIN/MCOa were 20.0%/33.3% ($p < 0.001$), 46.7%/60.0% ($p = 0.005$), and 80.0%/93.3% ($p < 0.001$) for raw, category, and binary scores, respectively.

3.6 Intra-Observer Variation in Plan Quality Scores

For the RO and MP involved in the intra-observer analyses, agreement percentages for the 65 initial raw plan scores and the

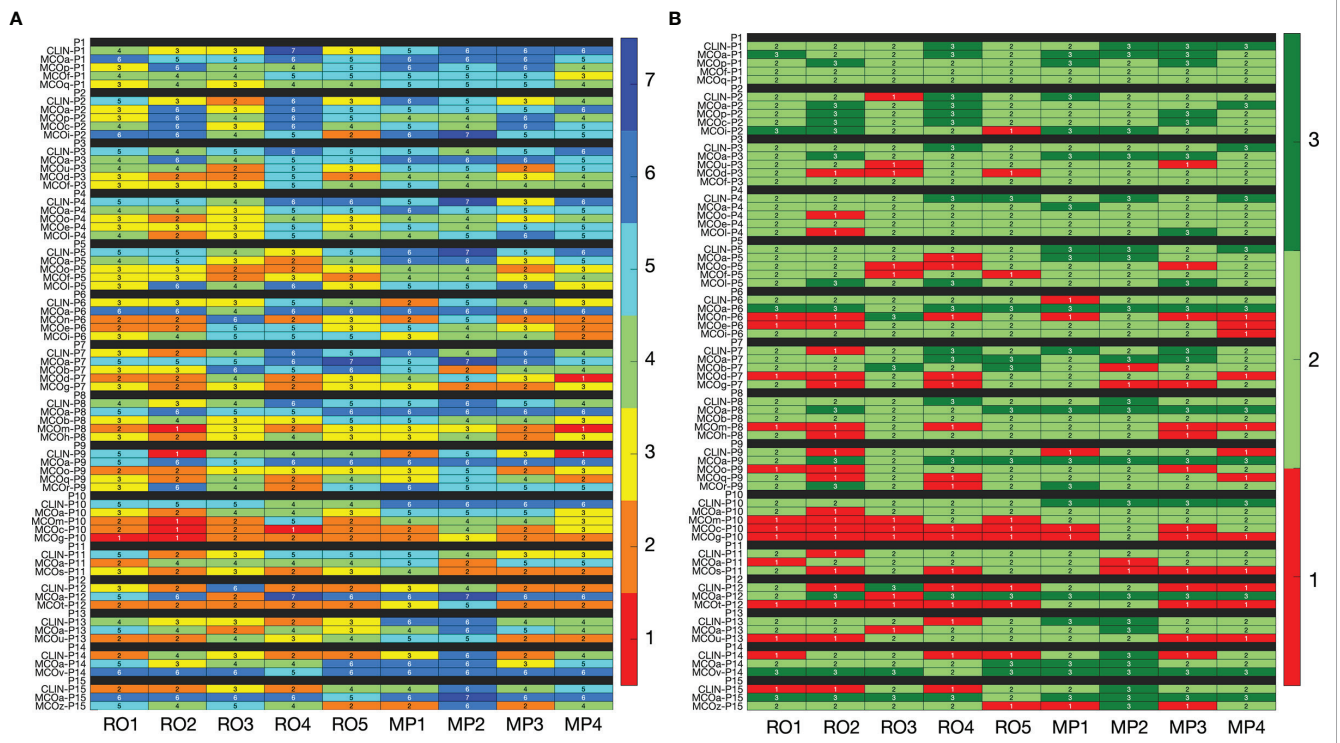


FIGURE 2 | **(A)** Heatmap visualization for raw plan quality scores (1–7, with 7 indicating the highest quality) of the nine observers (x-axis) for all 65 included plans (y-axis). **(B)** Heatmap visualization for category scores derived from the raw scores. In panel B, the color red indicates that the plan is considered unacceptable (category 1, raw scores 1 and 2), while light and dark green (category 2 with raw scores 3–5, and category 3 with raw scores 6 and 7) indicate that the plan is in principle acceptable. In the binary scoring system, red has binary score 0, while both light and dark green have binary score 1. See Section 3.5 for interpretation.

65 repeat raw scores were 40.0%/52.3% for RO/MP (N = 65). This is substantially higher than the expected percentage for random scoring (14.3%) and the median percentage of inter-observer score agreement of 27.7%; see **Figure 3A**. The repeat category agreements for the RO/MP were 70.8%/89.2% (N = 65) with corresponding expected random agreements and median inter-observer agreements of 33.3% and 58.5% (**Figure 3B**), respectively. For binary scoring, the RO/MP agreements were 86.2%/96.2%, with expected random and median inter-observer agreements of 50% and 78.5% (**Figure 3C**), respectively.

3.7 Suggested Plan Improvements

Large variability between observers was also observed in the suggestions for plan improvement. **Figure 4** shows the variability between observers for each of the possible options for improvements. Overall, the most chosen options were PTV conformity and dose reductions in OAR of group 2 (**Supplementary Figure S1**), parotids, esophagus, mandible, oral cavity, and larynx, with median percentages of 24.6% [0.0, 38.5] and 21.5% [13.8, 47.7], respectively. In the intra-observer evaluations, the participating RO and MP showed agreement percentages in the request for plan improvement of 28% and 46%, respectively.

4 DISCUSSION

In most centers, treatment plans are made by radiation therapy technologists (RTTs), dosimetrists, or MPs and evaluated for final approval by the treating ROs. The process, often denoted as manual planning or trial-and-error planning, may have several iterations in which the planner adjusts in-between plans, based on feedback by the RO. Limited common understanding or agreement between planners and ROs on how good plans should look like can result in suboptimal dose distributions, even with iteration loops. In this study, we have systematically investigated differences between five ROs and four planning MPs of a single radiotherapy department in perceived quality of oropharynx cancer plans. To the best of our knowledge, this is the first study that systematically investigates variations in subjective plan quality assessment among ROs and MPs working in a single department.

Even in our relatively small center with ROs and MPs working closely together based on the center's planning protocol (which is in line with international protocols, see M&M), large variations in subjective plan scores were observed. Considering all 36 unique observer pairs, the median percentage of plans for which they disagreed on clinical acceptability was 21.5% (**Figure 3C**), with minimum/maximum disagreements between pairs of 13.8%/36.9%. Based on Landis and Koch's labelling of Cohen's kappa values, the

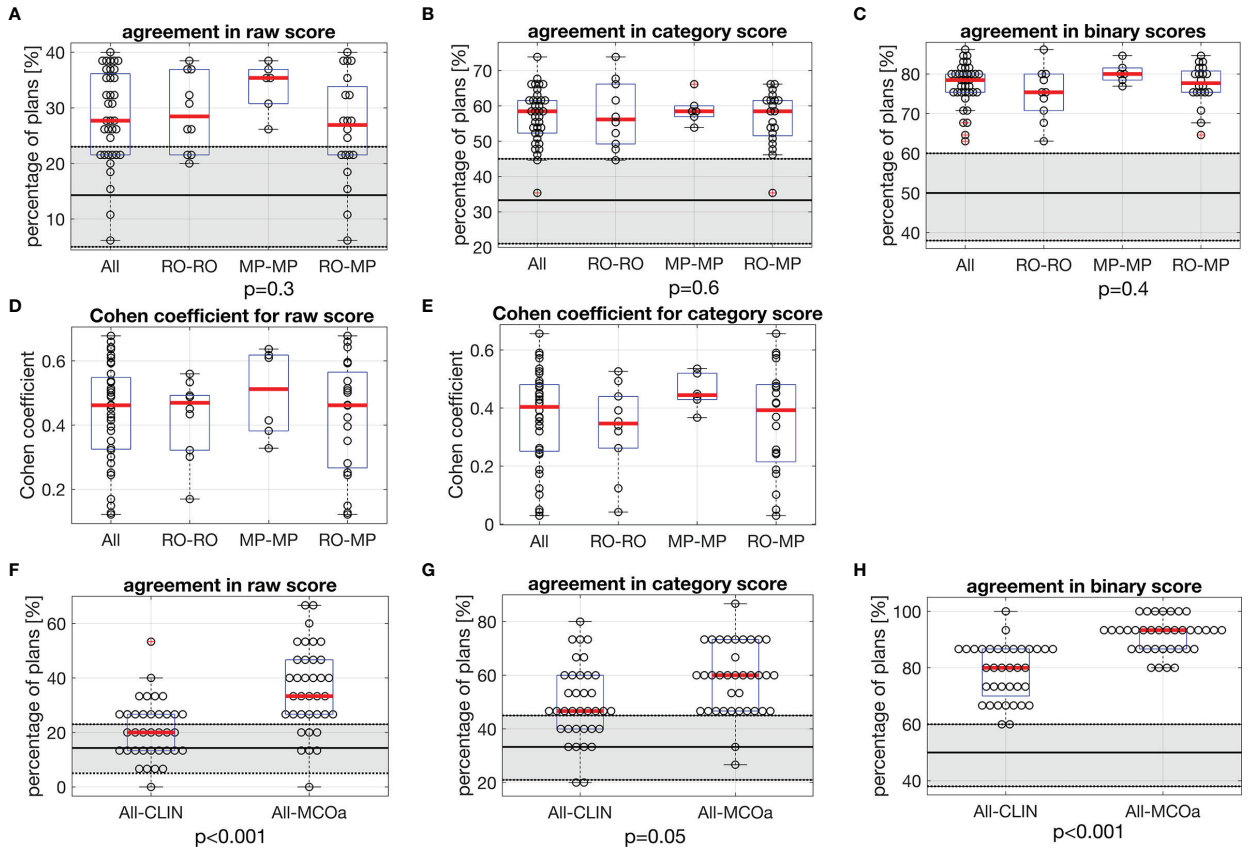


FIGURE 3 | In each panel, horizontal red lines in the boxplots show median values, while the edges of the boxes are the 25th and 75th percentiles. The whiskers extend to the most extreme data points not considered outliers, and the outliers are plotted individually using the "+" symbol. **(A–C)** Each marker shows for one of the unique 36 observer pairs in this study the percentage of 65 evaluated plans for which they agree in **(A)** raw score, **(B)** category score, and **(C)** binary score. In each panel, the first boxplot includes the data for all 36 observer pairs (All). For the other three boxplots, the data are split according to subgroups of observer pairs: RO–RO, pairs consist of two ROs; MP–MP, pairs consist of two MPs; RO–MP, pairs consist of one RO and one MP. Gray zones show expected distributions of agreement percentages in case of random, uncorrelated scoring by the observers in a pair, with the expected value denoted by the solid black line and the 95% confidence interval shown by the dotted borders. For observed agreements outside the gray zones, the difference with the expected score for random scoring is statistically significant. The p-values relate to ANOVA tests between subgroups RO–RO, MP–MP, and RO–MP of all observer pairs. **(D, E)** Corresponding Cohen's coefficients for raw and category scores. **(F–H)** Comparisons between observed agreement percentages for CLIN plans for all observer pairs, compared with observed agreement percentages for MCOa plans (generated with the optimal wish-list) for all observer pairs. The p-values were established with Wilcoxon's signed-rank tests. Gray zones: as in panels **(A–C)**. RO, radiation oncologist; MP, medical physicist; CLIN, clinical.

overall agreements in raw and category scores were "moderate" and "fair," respectively, but large variations between observer pairs were observed, going from "slight agreement" to "substantial agreement."

As shown in **Supplementary Figures S2B, S2D, S3** and **Supplementary Table S1**, dosimetric differences between the CLIN, MCOa, and MCOx plans could be substantial. As demonstrated in **Figure 2A**, for many observer–patient combinations, these dosimetric variations resulted in large variations in the three or five plan scores. On the other hand, different observers did often substantially disagree on the score of the same patient plan (see rows in **Figure 2A**). As can be observed in **Supplementary Figure S3**, dosimetric differences between patient plans, both positive and negative, were mostly not restricted to one parameter or one structure. Probably, different observers often appreciated the mixes in dosimetric pluses and minuses rather differently, contributing to the large disagreements between observers in assigned scores. This would be in line with the large inter-observer variations in suggested plan improvements (Section 3.7).

Figures 3A–C show that agreement percentages for RO–RO, MP–MP, and RO–MP pairs were similar (no statistically significant differences). This implicates that despite large differences in training and clinical roles of ROs and MPs, there were no enhanced rates of score mismatches in RO–MP pairs compared with RO–RO pairs.

Possibly, renewed, broad departmental discussions on plan requirements, aiming at a broadly shared and precisely defined view on plan quality, could improve the current large inter-observer variation in plan quality assessments. Probably also automated planning could result in improvements: as visible in **Figures 3F–H**, scoring agreements were better for the MCOa plans than for the CLIN plans, possibly related to more

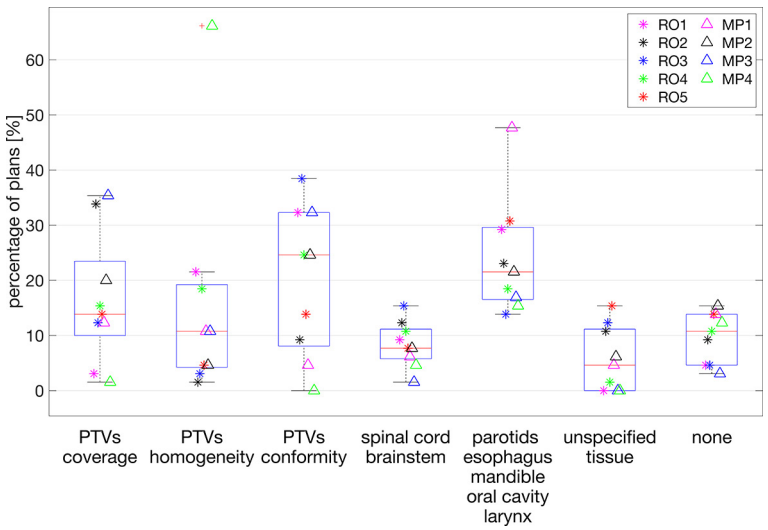


FIGURE 4 | Percentages of plans (y-axis) for which plan improvement options along the x-axis were requested. Each marker indicates a selected observer the percentage of plans for which the corresponding option for plan improvement was selected. For each observer, the presented percentages add up to 100%.

consistent automated generation of the MCOa plans. Apart from the better agreement between observers, MCOa scores were overall also higher than CLIN scores (**Table 1**), and MCOa plans were less frequently considered unacceptable than CLIN plans (4.4% *vs.* 14.8%, p = 0.1, Section 3.5). Enhanced plan quality with automated planning compared with manual planning has been observed previously [see, e.g., (8–12)], but to our knowledge, this is the first study showing also reduced inter-observer variations in subjective plan scores for the autoplans compared with corresponding manual plans. Other studies have pointed at the use of numerical plan quality assessment tools to enhance treatment plan quality (33).

In this study, clinical information about the patients was not available when doing the plan assessments, while it was available when the CLIN plan was made. This could in some cases have influenced scoring of the CLIN plan. On the other hand, all CLIN plans obeyed all hard clinical constraints for targets and OARs.

For some study patients, the CLIN plan was generated for TomoTherapy delivery, while the competitive MCOa and MCOx plans simulated VMAT (*Materials and Methods* section). Although observers were not informed on the delivery mode of presented plans, and all observers were aware that plan quality assessment was the study topic, it cannot be excluded that an observer could have identified TomoTherapy plans, which could possibly have influenced the scoring.

Although the observers were asked to give an absolute score (1–7) to each plan, the scoring of all three or five plans of a patient in a single session could have influenced the scores for the individual plans. For example, a plan could be perceived as unacceptable in the presence of a very good alternative plan, while when scored separately, the former plan could possibly have been acceptable for the observer. Such a mechanism could maybe in part explain the observation that 14.8% (median percentage for the nine observers, Section 3.5) of the CLIN plans was scored

TABLE 1 | Median differences in raw, category, and binary scores assigned by the five ROs, four MPs, and all nine observers combined (All).

***Raw* scores**	**MCOa – CLIN**				**MCOx – CLIN**				**MCOx – MCOa**			
	Diff	**Min**	**Max**	**p**	**Diff**	**Min**	**Max**	**p**	**Diff**	**Min**	**Max**	**p**
All	0.9	−1.4	2.8	**0.01**	−0.6	−2.9	2.8	**0.05**	−1.5	−1.2	3.2	**<0.001**
ROs	0.9	−1.4	3.2	**0.02**	−0.5	−2.7	3.2	0.15	−1.3	−1.4	3.2	**0.001**
MPs	0.9	−1.5	3.5	**0.02**	−0.8	−3.1	2.3	0.06	−1.6	−1.0	3.3	**<0.001**
***Category* scores**	**Diff**	**Min**	**Max**	**p**	**Diff**	**Min**	**Max**	**p**	**Diff**	**Min**	**Max**	**p**
All	0.3	−0.6	1.1	**0.02**	−0.3	−1.3	1.2	**<0.001**	−0.6	−0.6	1.4	**<0.001**
ROs	0.3	−0.4	1.4	**0.02**	−0.2	−1.0	1.4	**0.01**	−0.5	−0.6	1.4	**<0.001**
MPs	0.3	−1.0	1.5	0.1	−0.4	−1.8	1.0	<**0.001**	−0.7	−0.5	1.8	**<0.001**
***Binary* scores**	**Diff**	**Min**	**Max**	**p**	**Diff**	**Min**	**Max**	**p**	**Diff**	**Min**	**Max**	**p**
All	0.1	−0.1	0.4	0.1	−0.2	−0.9	0.4	**0.004**	−0.2	−0.1	0.8	<**0.001**
ROs	0.1	−0.2	0.6	0.2	−0.2	−1.0	0.6	**0.004**	−0.3	−0.2	0.8	**<0.001**
MPs	0.1	−0.3	0.5	0.1	−0.1	−0.8	0.5	**0.01**	−0.2	0.0	0.8	**<0.001**

Significant p-values are reported in bold.
RO, radiation oncologist; MP, medical physicist.

unacceptable, while all CLIN plans fulfilled the clinical hard constraints on PTV coverage, spinal cord Dmax, etc. It could maybe also explain the large difference between MCOa and MCOx in unacceptability rate (4.4% *vs*. 27.7% p = 0.005, Section 3.5), while also the intentionally suboptimal MCOx plans were generated while obeying all hard constraints (PTV, spinal cord, etc.). These observations point at a weak point of current manual planning: evaluating a plan is extremely difficult if there are no alternative plans.

In this study, we started off with five evaluable treatment plans per patient for the first 10 patients and then switched to three plans per patient to reduce the workload for the observers to a more acceptable level. For raw scores, the median of the percentage of score agreements for the 36 unique observer pairs was the highest for three plans per patient (33.3% *vs*. 26.0%), which was borderline significant (p = 0.07). For category scores, three plans resulted in lower agreement percentages (53.3% *vs*. 60%, p = 0.06), which was also seen for binary scores (66.7% *vs*. 81.0%, p < 0.001). Probably, the involved patient numbers are too small to draw strong conclusions on three versus five.

The 70.8% and 89.2% agreements in repeated category scoring and 86.2% and 96.2% in repeated binary scoring (Section 3.6) point at an option for high-accuracy score prediction for single observers with machine learning. Future application of such tools could possibly contribute to enhanced plan quality consistency. This is a topic of ongoing research.

In this study, we considered oropharynx cases with three dose levels and many OARs. The complexity of these cases could have contributed to the observed large and frequent disparities in observer scores. Possibly, for less complex tumor sites, agreement in plan scores could be better, which is a topic for further research.

We believe that this is the first study that quantitatively evaluates variations in subjective assessments of the same treatment plans by various observers (ROs and MPs) in the same department. Our study is very different from, but complementary to, other studies that demonstrate that different planners can generate very different plans for the same patient, even with very detailed, quantitative instructions on how the plan should look like (6). In the latter studies, plan quality differences are usually attributed to differences between planners in planning skills, dedication, and ambition, and in time spent on planning. On the contrary, in our study, all observers evaluate the same plans, and we test how well these plans fit the observer-specific ideas on how good plans should look like.

The results of the current study could stimulate similar studies in other departments as they seem to point at an important weak link in radiotherapy planning. It is commonly recognized that variations between ROs in delineated targets are a major concern in clinical radiotherapy. This study suggests that large inter-observer variations in plan quality assessments (even in a single department) could be another Achilles heel for successful treatment.

5 CONCLUSIONS

Inter-observer differences in treatment plan quality assessments in radiotherapy can be substantial and could hamper consistent preparation of high-quality plans, even in a single radiotherapy department. Agreements between ROs and MPs in plan assessments were similar to agreements among ROs only, despite large differences between ROs and MPs in training and clinical roles. Automatically generated plans (MCOa) showed the highest median scores and best inter-observer score agreements, pointing at a potential for automated planning to improve clinical practice.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Comitato Etico di Area Vasta Emilia Nord—n. 391/2018/OSS/IRCCSRE. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

EC, BH, LR, AB, and ES contributed to the development of the study design and applied methodology. LR, EC, and AB developed the wish-lists for automated planning, in collaboration with CI, RS, MI, AR, GT, SC, and MG. CI, RS, MI, AR, GT, SC, MG, AB, and EC performed the subjective plan evaluations. EC, AB, LR, and BH performed the data analysis, with AB responsible for the statistical analyses. BH and EC supervised all work. All authors contributed to the article and approved the submitted version.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2021.706034/full#supplementary-material

REFERENCES

1. Nutting CM, Morden JP, Harrington KJ, Guerrero Urbano T, Bhide SA, Clark C, et al. Parotid-Sparing Intensity Modulated Versus Conventional Radiotherapy in Head and Neck Cancer (PARSPORT): A Phase 3 Multicentre Randomised Controlled Trial. *Lancet Oncol* (2011) 12:127–36. doi: 10.1016/S1470-2045(10)70290-4
2. Mukesh MB, Barnett GC, Wilkinson JS, Moody AM, Wilson C, Dorling L, et al. Randomized Controlled Trial of Intensity-Modulated Radiotherapy for Early Breast Cancer: 5-Year Results Confirm Superior Overall Cosmesis. *J Clin Oncol* (2013) 31:4488–95. doi: 10.1200/JCO.2013.49.7842
3. Chun SG, Hu C, Choy H, Komaki RU, Timmerman RD, Schild SE, et al. Impact of Intensity-Modulated Radiation Therapy Technique for Locally Advanced Non-Small-Cell Lung Cancer: A Secondary Analysis of the NRG

Oncology RTOG 0617 Randomized Clinical Trial. *J Clin Oncol* (2017) 35:56–62. doi: 10.1200/JCO.2016.69.1378
4. Viani GA, Viana BS, Martin JE, Rossi BT, Zuliani G, Stefano EJ. Intensity-Modulated Radiotherapy Reduces Toxicity With Similar Biochemical Control Compared With 3-Dimensional Conformal Radiotherapy for Prostate Cancer: A Randomized Clinical Trial. *Cancers* (2016) 122:2004–1. doi: 10.1002/cncr.29983
5. Staffurth JRadiotherapy Development B. A Review of the Clinical Evidence for Intensity-Modulated Radiotherapy. *Clin Oncol* (2010) 22:643–57. doi: 10.1016/j.clon.2010.06.013
6. Nelms BE, Robinson G, Markham J, Velasco K, Boyd S, Narayan S, et al. Variation in External Beam Treatment Plan Quality: An Inter-Institutional Study of Planners and Planning Systems. *Pract Radiat Oncol* (2012) 2(4):296–305. doi: 10.1016/j.prro.2011.11.012
7. Berry SL, Boczkowski A, Ma R, Mechalakos J, Hunt M. Interobserver Variability in Radiation Therapy Plan Output: Results of a Single-Institution Study. *Pract Radiat Onco* (2016) l6:442–9. doi: 10.1016/j.prro.2016.04.005
8. Hansen CR, Nielsen M, Bertelsen AS, Hazell I, Holtved E, Zukauskaite R, et al. Automatic Treatment Planning Facilitates Fast Generation of High-Quality Treatment Plans for Esophageal Cancer. *Acta Oncol* (2017) 56:1495–500. doi: 10.1080/0284186X.2017.1349928
9. Rossi L, Sharfo AW, Aluwini S, Dirkx M, Breedveld S, Heijmen B. First Fully Automated Planning Solution for Robotic Radiosurgery - Comparison With Automatically Planned Volumetric Arc Therapy for Prostate Cancer. *Acta Oncol* (2018) 57:1490–8. doi: 10.1080/0284186X.2018.1479068
10. Heijmen B, Voet P, Fransen D, Penninkhof J, Milder M, Akhiat H, et al. Fully Automated, Multi-Criterial Planning for Volumetric Modulated Arc Therapy - An International Multi-Center Validation for Prostate Cancer. *Radiother Oncol* (2018) 128:343–8. doi: 10.1016/j.radonc.2018.06.023
11. Marrazzo L, Meattini I, Arilli C, Calusi S, Casati M, Talamonti C, et al. Auto-Planning for VMAT Accelerated Partial Breast Irradiation. *Radiother Oncol* (2019) 132:85–92. doi: 10.1016/j.radonc.2018.11.006
12. Hussein M, Heijmen BJ, Verellen D, Nisbet A. Automation in Intensity-Modulated Radiotherapy Treatment Planning - A Review of Recent Innovations. *Br J Radiol* (2018) 91(1092):20180270. doi: 10.1259/bjr.20180270
13. Cagni E, Botti A, Wang Y, Iori M, Petit SF, Heijmen BJ. Pareto-Optimal Plans as Ground Truth for Validation of a Commercial System for Knowledge-Based DVH-Prediction. *Phys Medica* (2018) 55:98–106. doi: 10.1016/j.ejmp.2018.11.002
14. Edge SB, Byrd DR, Compton CC, Fritz AG, Greene FL, Trotti A. *AJCC Cancer Staging Manual. 7th ed.* New York: Springer (2010).
15. Wang X, Eisbruch A. IMRT for Head and Neck Cancer: Reducing Xerostomia and Dysphagia. *J Radiat Res* (2016) 57:i69–75. doi: 10.1093/jrr/rrw047
16. Lee NY, Riaz N, Lu JJ. *Target Volume Delineation for Conformal and Intensity-Modulated Radiation Therapy*. Springer (2015). Chapters 1–2.
17. Wolden SL, Chen WC, Pfister DG, Kraus DH, Berry SL, Zelefsky MJ. Intensity-Modulated Radiation Therapy (IMRT) for Nasopharynx Cancer: Update of the Memorial Sloan-Kettering Experience. *Int J Radiat Oncol Biol Phys* (2006) 64:57–62. doi: 10.1016/j.ijrobp.2005.03.057
18. Gregoire V, Evans M, Le QT, Bourhis J, Budach V, Chen A, Eisbruch A, et al. Delineation of the Primary Tumour Clinical Target Volumes (CTV-P) in Laryngeal, Hypopharyngeal, Oropharyngeal and Oral Cavity Squamous Cell Carcinoma: AIRO, CACA, DAHANCA, EORTC, GEORCC, GORTEC, HKNPCSG, HNCIG, IAG-KHT, LPRHHT, NCIC CTG, NCRI, NRG Oncology, PHNS, SBRT, SOMERA, SRO, SSHNO, TROG Consensus Guidelines. *Radiother Oncol* (2018) 126:3–24. doi: 10.1016/j.radonc.2017.10.016
19. Radiation Therapy Oncology Group (RTOG) Protocol. *Loco-Regionally Advanced Oropharyngeal Carcinoma Trial 0022 - Phase I/II Study of Conformal and Intensity Modulated Irradiation for Oropharyngeal Cancer*. (2001). Available at: http://www.rtog.org/ClinicalTrials.
20. Radiation Therapy Oncology Group (RTOG) Protocol. *Loco-Regionally Advanced Nasopharyngeal Carcinoma Trial 0225- A Phase II Study of Intensity Modulated Radiation Therapy (IMRT) +/Chemotherapy for Nasopharyngeal Cancer*. (2003). Available at: http://www.rtog.org/ClinicalTrials.
21. Radiation Therapy Oncology Group (RTOG) Protocol. *Loco-Regionally Advanced Nasopharyngeal Carcinoma: Radiation Therapy Oncology Group (RTOG) Trial 0615 - A Phase II Study of Concurrent Chemoradiotherapy Using Three-Dimensional Conformal Radiotherapy (3D-CRT) or Intensity-Moldulated Radiation Therapy (IMRT) + Bevacizumab (BV) for Locally or Regionally Advanced Nasopharyngeal Cancer*. (2006). Available at: http://www.rtog.org/ClinicalTrials.
22. Yao Y, Lee NY. JAVELIN Head and Neck 100: A Phase III Trial of Avelumab and Chemoradiation for Locally Advanced Head and Neck Cancer. *Fut Oncol* (2019) 15:687–94. doi: 10.2217/fon-2018-0405
23. Taherdoost H. What Is the Best Response Scale for Survey and Questionnaire Design; Review of Different Lengths of Rating Scale/Attitude Scale/Likert Scale. *Int J Acad Res Manage (IJARM)* (2019) 8:⟨hal–02557308⟩.
24. Preston CC, Colman AM. Optimal Number of Response Categories in Rating Scales: Reliability, Validity, Discriminating Power, and Respondent Preferences. *Acta Psychol (Amst)* (2000) 104(1):1–15. doi: 10.1016/s0001-6918(99)00050-5
25. Colman AM, Norris CE. Comparing Rating Scales of Different Lengths: Equivalence of Scores From 5-Point and 7-Point Scales. *Psychol Rep* (1997) 80:355–62. doi: 10.2466/pr0.1997.80.2.355
26. Breedveld S, Storchi PR, Voet PW, Heijmen BJ. Icycle: Integrated, Multicriterial Beam Angle, and Profile Optimization for Generation of Coplanar and Noncoplanar IMRT Plans. *Med Phys* (2012) 39:951–63. doi: 10.1118/1.3676689
27. Sharfo AW, Breedveld S, Voet PW, Heijkoop ST, Mens JM, Hoogeman MS, et al. Validation of Fully Automated VMAT Plan Generation for Library-Based Plan-Of-the-Day Cervical Cancer Radiotherapy. *PloS One* (2016) 11: e0169202. doi: 10.1371/journal.pone.0169202
28. Della Gala G, Dirkx MLP, Hoekstra N, Fransen D, Lanconelli N, van de Pol M, et al. Fully Automated VMAT Treatment Planning for Advanced-Stage NSCLC Patients. *Strahlenther Onkol* (2017) 193:402–9. doi: 10.1007/s00066-017-1121-1
29. Buschmann M, Sharfo AWM, Penninkhof J, Seppenwoolde Y, Goldner G, Georg D, et al. Automated Volumetric Modulated Arc Therapy Planning for Whole Pelvic Prostate Radiotherapy. *Strahlenther Onkol* (2018) 194:333–42. doi: 10.1007/s00066-017-1246-2
30. Cohen J. A Coefficient of Agreement for Nominal Scales. *Educ Psychol Measurement* (1960) 20:37–46. doi: 10.1177/001316446002000104
31. Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data. *Biometrics* (1977) 33:159–74. doi: 10.2307/2529310
32. Watson PF, Petrie A. Method Agreement Analysis: A Review of Correct Methodology. *Theriogenology* (2010) 1167–79. doi: 10.1016/j.theriogenology.2010.01.003
33. Yibing W, Heijmen BJ, Petit SF. Prospective Clinical Validation of Independent DVH Prediction for Plan QA in Automatic Treatment Planning for Prostate Cancer Patients. *Radiother Oncol* (2017) 125:500–6. doi: 10.1016/j.radonc.2017.09.021

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

frontiers in Oncology

REVIEW
published: 14 October 2021
doi: 10.3389/fonc.2021.732196

Current Advances and Challenges in Radiomics of Brain Tumors

Zhenjie Yi[1,2,3], *Lifu Long*[2], *Yu Zeng*[1,3]* *and Zhixiong Liu*[1,3]*

[1] Department of Neurosurgery, Xiangya Hospital, Central South University, Changsha, China, [2] XiangYa School of Medicine, Central South University, Changsha, China, [3] National Clinical Research Center for Geriatric Disorders, Xiangya Hospital, Central South University, Changsha, China

Imaging diagnosis is crucial for early detection and monitoring of brain tumors. Radiomics enable the extraction of a large mass of quantitative features from complex clinical imaging arrays, and then transform them into high-dimensional data which can subsequently be mined to find their relevance with the tumor's histological features, which reflect underlying genetic mutations and malignancy, along with grade, progression, therapeutic effect, or even overall survival (OS). Compared to traditional brain imaging, radiomics provides quantitative information linked to meaningful biologic characteristics and application of deep learning which sheds light on the full automation of imaging diagnosis. Recent studies have shown that radiomics' application is broad in identifying primary tumor, differential diagnosis, grading, evaluation of mutation status and aggression, prediction of treatment response and recurrence in pituitary tumors, gliomas, and brain metastases. In this descriptive review, besides establishing a general understanding among protocols, results, and clinical significance of these studies, we further discuss the current limitations along with future development of radiomics.

Keywords: radiomics, radiogenomics, glioma, pituitary tumor, brain metastases

OPEN ACCESS

Edited by:
Jose Eduardo Villarreal Barajas, Royal Devon and Exeter Hospital, United Kingdom

Reviewed by:
Kevin Martell, University of Calgary, Canada
Yinyan Wang, Capital Medical University, China

***Correspondence:**
Yu Zeng
zengyu@csu.edu.cn
Zhixiong Liu
zhixiongliu@csu.edu.cn

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

Received: *28 June 2021*
Accepted: *23 September 2021*
Published: *14 October 2021*

Citation:
Yi Z, Long L, Zeng Y and Liu Z (2021) Current Advances and Challenges in Radiomics of Brain Tumors. Front. Oncol. 11:732196. doi: 10.3389/fonc.2021.732196

INTRODUCTION

Brain and other CNS tumors, including gliomas, pituitary tumors, and others such as brain metasteses, mainly occur in lung cancer and breast cancer patients. These tumors stand out for their high diversity and heterogeneity, along with dismal prognosis, ranking them among the top 10 causes of cancer deaths, accounting for a significant proportion of the deaths in men less than 40

Abbreviations: AUC, area under the curve; BM, brain metastases; CAD, computer-aided diagnosis and detection; CADe, computer-aided detection; CBV, cerebral blood volume; CCC, concordance correlation coefficient; CE-T1WI, contrast-enhanced T1-weighted MRI; CNN, convolutional neural network; DSC, dynamic susceptibility contrast; DTI, diffusion tensor imaging; DWI, diffusion-weighted imaging; EGFR, epidermal growth factor receptor; FLAIR, fluid-attenuated inversion recovery; GBM, glioblastoma; HGG, high-grade glioma; IDH, isocitrate dehydrogenase; LASSO, least absolute shrinkage and selection operator; LGG, low-grade glioma; MGMT, O-6-methylguanine-DNA methyltransferase; MRI, magnetic resonance imaging; mRMR, minimum redundancy maximum relevance algorithm; MRS, magnetic resonance spectroscopy; NFPA, non-functioning pituitary adenoma; OS, overall survival; PA, pituitary adenoma; PCA, principal component analysis; PCNSL, primary central nervous system lymphoma; PD, progressive disease; PFS, progression-free survival; PsP, Pseudoprogression; PWI, perfusion-weighted imaging; RECIST, Response Evaluation Criteria in Solid Tumours; RF, random forest; RFE, recursive feature elimination; ROC, receiver operating characteristic curve; RQS, radiomics quality score; SAM, significance analysis of microarrays; SVM, support vector machine; TRC, treatment-related changes.

years and women less than 20 years in the United States in 2018 (1), and it is estimated that they will cause 18,600 deaths in 2021.

Clinical radiology is a routinely performed examination for patients who are suspicious of brain or other CNS tumors; recently more and more sophisticated analytic methods have sprung up which supplement traditional imaging techniques. Based on the imaging techniques and incorporated with computer vision and machine learning (2), radiomics was born. Radiomics first appeared in Philippe's review in 2012 (3), initially as an extended technique of computer-aided diagnosis and detection (CAD) systems. The term radiomics refers to the refining of a large mass of quantitative features from complex clinical imaging arrays, then transforming them into high-dimensional data which can subsequently be mined to find their relevance with the tumor's histological features, which reflect underlying genetic mutations and malignancy, along with grade, progression, therapeutic effect, or even overall survival (OS) (4). Deep learning is a branch of machine learning, and machine learning is the necessary path to realize artificial intelligence (AI). The concept of deep learning originated from the study of neural networks that simulate the human brain. In recent years, we have seen a blossoming in AI development, with more intelligent algorithms such as deep learning bringing the possibility of realizing fully automatic image capturing and reading processes. Compared to traditional manual radiology practice requiring trained physicians to deal with large quantities of information, which is labor intensive, subjective, and qualitative, radiomics is able to use AI methods to provide automatic, objective, and quantitative data with high efficiency. In this review, we focus on radiomics and extended imaging techniques. The general clinical applications of these noninvasive methods are shown (**Figure 1**).

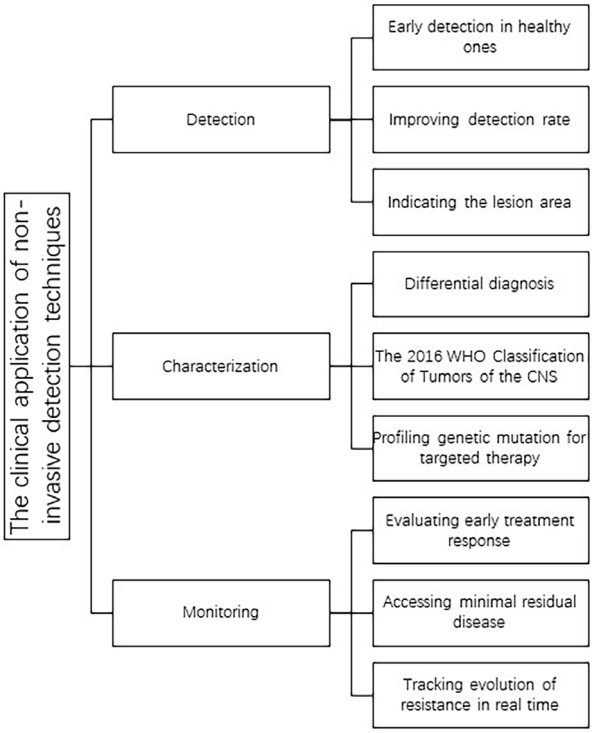


FIGURE 1 | The clinical application of the noninvasive detection techniques.

Clinical Application

Early accurate diagnosis and classification are crucial in prolonging the patient's survival time. The application of radiomics has been initiated in clinical oncology early diagnosis, since its ability to analyze the combination of numerous quantitative features provides the possibility to unravel the underlying pathophysiology that is hard to be perceived by radiologists' eyes and avoid subjective misreading. The general workflow of radiomics involves several discrete steps: imaging, segmentation, feature extraction, feature selection, machine learning, and validation (5). Segmentation is using a series of algorithms to delineate regions of interest (ROIs), which refers to the tumor and its surrounding abnormality from other tissue, and then further subdividing the legion by its intra-heterogeneity to facilitate the next steps. Accurate segmentation is a key step from image processing to quantitative analysis, acting as a prerequisite for the subsequent diagnostic tasks like defining the location, extent of radiation, and tumor feature extraction, and is the most challenging step due to high heterogeneity and irregularity of brain tumors. Feature extraction refers to the use of existing features to calculate a feature set with a higher degree of abstraction, and also refers to an algorithm for calculating a certain feature. Then the extracted features will go through feature selection, which aims at reducing dimensionality and the difficulty of learning tasks to improve the efficiency of the model. The top features will be integrated with clinical results and/or pathological results, together as input into machine learning methods to build prediction or classification models.

The main clinical tasks of radiomics lie in three parts: detection, characterization, and monitoring. By the help of computer-aided detection (CADe), the suspicious area of the image can be highlighted and some features indicating early cancer lesions can be detected, which reduce observational oversights and improve the speed of interpretation (6, 7). Most radiomics models are served for characterization, including diagnostic tasks (differential diagnosis, malignancy, WHO CNS classification, specific genetic mutation status, and treatment effect) and predictive/prognostic tasks (treatment effect, OS/PFS, complication, and tumor recurrence). Monitoring is also significant within clinical practice in evaluating the progression of tumor and the effect of treatment. Traditional protocols to assess the tumor progression, recommended by institutions like Response Evaluation Criteria in Solid Tumors (RECIST) and WHO, are usually defined by the size of tumor (8), which omits much geometry and material information detected by advanced radiological instruments and also oversimplifies the indicators on tumor burden. The emergence of AI-monitoring may help radiologists to establish more sophisticated quantitative protocols towards tumor burden evaluation.

MRI acts as a key part and is usually the first choice in radiological diagnosis of brain tumors. First, MRI has an

TABLE 1 | Radiomic.

Study	No. of patients involved (Training/ Test)	General purpose	CT or MRI contrast(s)	Segmentation method	Nomber of features (extraction/ selection)	Feature selection method	Classification method	Performance (Training/ Test)
Yae Won Park et al. (30)	141/36	To predict the DA response in prolactinoma patients	T2WI	^^3D Slicer software	107/n.a.	/	soft voting (RF, light gradient boosting machine, extra-trees, quadratic discrimination analysis, and linear discrimination analysis)	#0.81/0.81 (AUC)
Yang Zhang et al. (40)	50/n.a.	To predict progression or reccurence in NFPAs	CE-T1WI and T2WI	^^^fuzzy c-mean (FCM) clustering algorithm	107/3	SVM	SVM	***0.78/n.a.(AUC)
Andrei Mouraviev et al. (117)	87/n.a.	To predict local recurrence following SRS	CE-T1WI and FLAIR	^^Elastix software	440/12	Random forest feature importance (RFI)	RF	#0.793/n.a. (AUC)
Kniep, H et al.(113)	189/n.a.	To predict tumor type in different BM with unknown primary lesion	CE-T1WI, T1WI and FLAIR	^^Analyze 11.0 Software	1423/20	Decision tree	RF	*0.64 (NSCLC)-0.82 (MM)/ n.a.(AUC)
Peng, LK et al. (116)	66/n.a.	To predict diagnosing treatment effect after stereotactic radiosurgery	CE-T1WI and FLAIR	^^a multiparametric deep learning (MPDL) network	51/5	IsoSVM	IsoSVM	**0.81/n.a.(AUC)
Ji Eun Park et al. (73)	85/35	To predict core signaling pathways in IDH wild-type GBM	T1WI, T2WI, DWI, FLAIR, CE-T1WI, and DSC	^^MITK software	71/5(RTK), 17/ 5(P53), 35/5 (retinoblastoma)	t-test, LASSO, RF	Logistic regression	(3-fold CV)0.87/0.88(RTK), 0.80/0.76(p53), 0.84/0.81 (retinoblastoma pathway) (AUC)
Pascal O. Zinn et al. (65)	46/47 (GBM), 40/ n.a.(mice)	To establish causality between POSTN status and MRI-extracted radiomic-features in GBM	FLAIR and CE-T1WI	^^3D Slicer software	2480/48 (GBM),17(mice)	LASSO	Binary logistic regression	**76.56%/n.a.(GBM), 92.26%/n.a.(mice)(AUC)
Chia-Feng Lu et al. (46)	214/70	To stratify the molecular subtypes of gliomas	CE-T1WI, FLAIR, T2WI, and DWI	^^n.a.	39212/(20-1960)	Two-sample t-test with pooled variance estimate	SVM	*87.7%-96.1%/80.0%-91.7%(accuracy)
Robin Gutsche et al. (47)	50/n.a.	To evaluate the repeatability of feature-based FET PET radiomics and investigate IDH genotype on feature repeatability	FET PET	^^^TBR≥1.6	1302/297	intraclass correlation coefficient	n.a.	n.a.
Yoon Seong Choi et al. (48)	727/439 (129 internal and 310 external)	To predict the IDH status of gliomas	CE-T1WI, T2WI and FLAIR	^^^CNN	24/20	CNN	CNN	***0.96/0.94(internal), 0.86 (external)(AUC)

(Continued)

TABLE 1 | Continued

Study	No. of patients involved (Training/ Test)	General purpose	CT or MRI contrast(s)	Segmentation method	Nomber of features (extraction/ selection)	Feature selection method	Classification method	Performance (Training/ Test)
Anna Luisa Di Stefano et al. (67)	66/78	To characterize the clinical, radiological, and molecular profile of F3T3 positive diffuse gliomas	T1WI, CE-T1WI and FLAIR	^ITK-SNAP software	2616/25	Cox proportional hazards models (OS)	RF(F3T3)	*F3T3: 0.87/0.745(AUC)
Xiaorui Su et al. (71)	75/25	To predict H3 K27M mutation status in midline gliomas.	FLAIR	^ITK-SNAP software	99/10	TPOT	TPOT	*** 0.788~0.867/0.60~ 0.84 (accuracy)
Zev A. Binder et al. (59)	260/n.a.	To investigate the negative survival impact of EGFRA289D/T/V	T1WI, CE-T1WI, T2WI, FLAIR, DTI, and DSC	^^GLISTRboost	2104/299/17	SVM/ radiographic interpretability	n.a.	n.a.
Jingwei Wei et al. (50)	74/31	To predict MGMT methylation status in astrocytoma	CE-T1WI, FLAIR and DWI	^ITK-SNAP software	3051/13	The minimum redundancy maximum relevance (mRMR)	logistic regression	#0.925/0.902(AUC)
Yiming Li et al. (51)	63/123(32 internal and 91 external)	To predict ATRX mutation in LGGs	T2WI	^MRIcro software	431/9	LASSO	SVM	#0.94/0.925(internal) and 0.725(external)(AUC)
Johannes Haubold et al. (52)	28/14	To predict tumor grading and mutational status of patients with cerebral gliomas	CE-T1WI, ADC, and 3D-FLAIR (SPACE), FET PET, and SWI, The water-content-based M0 map (MRF M0) , T1FLAIR, DWI	^^3D Slicer software	19284/32 (1p19q codeletion), 64 (IDH1), 8 (ATRX),16 (MGMT)	f score (ANOVA), chi-square, LCSI and randomized logistics regression (RandLR)	RF and SVM	*WHOI-IV:0.818/n.a.(AUC); Differentiation of LGG and HGG:0.85/n.a.(AUC); 1p19q codeletion: 0.9784/ n.a.(AUC); IDH1:0.88/ n.a.(AUC); ATRX:0.851/ n.a.(AUC); MGMT:0.757/ n.a.(AUC)
Luyuan Zhang et al. (57)	96/24	To identify the value of CIC mutations in gliomas	T1WI, T2WI, FLAIR and CE-T1WI	^FSL image viewer	6676/11	LASSO	logistic regression	*0.985/n.a.(AUC)
Changliang Su et al. (53)	220/n.a.	To differentiate among glioma subtypes and predict tumour proliferation	T2WI fast-echo images (T2FSE), T1WI, FLAIR, CE-T1WI, DWI, PWI and CBF	^^ImageJ	431/25	univariate analysis	logistic regression	#0.936/n.a.(AUC)
Yiming Li et al. (76)	69/40	To predict PTEN mutation status in GBM	T1WI, T2WI and CE-T1WI	^MRIcro software	862/6	The minimum redundancy maximum relevance (mRMR)	SVM	#0.925/0.787(AUC)

*^ manual segmentation; ^^ semi-automatic segmentation; ^^^ full-automatic segmentation; * 5-fold cross-validation; ** leave-one out cross-validation; *** 10-fold cross-validation; # cross validation not available;*

AUC, area under the receiver operating characteristic curve; GBM, glioblastoma; BM, brain metastases; PsP, pseudoprogression; CCC, concordance correlation coefficient; IDH, isocitrate dehydrogenase; PCNSL, primary central nervous system lymphoma; TBR, tumor-to-brain ratio; CNN, convolutional neural network; LASSO, least absolute shrinkage and selection operator; MGMT, O6-methylguanineDNA-methyltransferase; EGFR, epidermal growth factor receptor; TRC, treatment-related changes; RFE, recursive feature elimination; RF, random forest; SAM, significance analysis of microarrays; SVM, support vector machine; n.a., not available; PCA, principal component analysis;

MRI, magnetic resonance imaging; mRMR, minimum redundancy maximum relevance algorithm; MRS, magnetic resonance spectroscopy; DTI, diffusion tensor imaging; DWI, diffusion-weighted imaging; FLAIR, fluid-attenuated inversion recovery; PWI, perfusion-weighted imaging; T1, T1-weighted MRI; CE-T1WI, contrast-enhanced T1-weighted MRI; T2, T2-weighted MRI; DSC, dynamic susceptibility contrast; CBV, cerebral blood volume.

outstanding contrast capacity for the detection of brain tissues. Second, MRI has many different sequences respectively sensitive to different physiology parameters, such as blood flow and edema in surroundings, which indicate the tumor's microenvironment. Third, MRI can be implemented throughout the treatment noninvasively and assesses the progression and effect. Besides this anatomical imaging, the multimodal MRI with emerging sequences and technical developments like PET using either amino acid, choline, or fluorodeoxyglucose, as well as fusion PET/CT and PET/MRI scanners, provides a mass of functional

TABLE 2 | Radiomic.

Study	No. of patients involved (Training/ Test)	General purpose	CT or MRI contrast (s)	Segmentation method	Nomber of features (extraction/ selection)	Feature selection method	Classification method	Performance (Training/ Test)
Daesung Kang et al. (82)	112/42	To identify atypical PCNSL mimicking GBM	T1WI, T2WI, FLAIR, DWI, CE-T1WI and DSC	^^MITK software	1618/15 (ADC), 40 (CE-T1WI)	RFE(ADC); relief (CE-T1W1)	RF(ADC); LDA (CE-T1WI)	***0.983(AUC)/0.984 (internal AUC)0.944 (external AUC)
Guoqing Wu et al. (83)	67/35	To differentiate PCNSL and GBM and for IDH1 mutation estimation	T2WI and CE-T1WI	^^^CNN	968/49	Sparse representation	Collaborative sparse representation	**98.51%/94.51% (accuracy)
Zenghui Qian et al. (81)	227/185	To differentiate GBM from BM preoperatively	T1WI, T2WI and CE-T1WI	^MRIcro software	1303/24	LASSO	SVM	*0.945/0.90(AUC)
Jung Youn Kim et al. (86)	61/57 (23 internal and 34 external)	To differentiate PsP from early tumor progression in patients with GBM	CE-T1WI, FLAIR, ADC and CBV maps	^^MITK software	6472/12	LASSO logistic regression model	Student's t-test and the chi-square test	***0.9(AUC)/0.96(internal AUC), 0.85(external AUC)
Katrin Aslan et al. (91)	148/ n.a.(mice)	To predict treatment response and PsP in ICB-treated mice	T2WI and CE-T1WI	^Osirix or ITKsnap software	423/423	/	Gradient boosting	*82.7%/n.a. (accuracy)
Nabil Elshafeey et al. (84)	98/7	To discriminate PsP from progression in GBM	FLAIR, T1WI and CE-T1WI	^^3D Slicer software	620/60	Maximum Relevance Minimum Redundancy (MRMR)	SVM	**89%/n.a.(AUC)
Jinhua Cai et al. (90)	77/72 (41 internal and 31 external)	To predict the response to bevacizumab in patients with brain necrosis after radiotherapy	FLAIR	^^3D Slicer software	1301/18	LASSO	multivariate logistic	#0.916/0.912 (internal) and 0.827 (external)(AUC)
Philipp Lohmann et al. (85)	72/30	To differentiate PsP in glioma patients post-chemoradiation	FET PET	^^^TBR=1.4, 1,6, 1.8	944/4	RFE	TPOT	*0.74/0.74 (AUC)
Kai Wang et al. (87)	112/48	To discriminate tumor recurrence from radiation necrosis in glioma	CE-T1WI, FLAIR, 18F-FDG and 11C-MET PET	^ITK-SNAP software	396/15	LASSO	Multivariable logistic regression analysis	***0.988/0.914(AUC)
Zi-Qi Pan et al. (94)	82/70 (40 internal and 30 external)	To predict the response of individual GBM patients to radiotherapy	T1WI, CE-T1WI, T2WI, and FLAIR	^^GLISTR software	28496/8	Boruta algorithm	RF	***0.764/0.758 (external) (C-index)
Xing Liu et al. (103)	216/84	To predict the PFS in LGGs and investigate the corresponding genetic background	T2WI	^MRIcro software	431/9	Univariate Cox regression	LASSO Cox regression	***0.684/0.823(C-index)
Sara Dastmalchian et al. (95)	31/n.a.	To differentiate between intra-axial adult brain tumors and to predict survival in the GBM cohort	T1WI, T2WI, FLAIR and CE-T1WI	^Magnetic resonance fingerprinting	39/20	Spearman's rank correlation coefficient test	Wilcoxon rank sum	#Peripheric white matter regions in GBM and LGGs: 0.869/n.a.(AUC), Solid tumor regions in LGGs and metastases: 0.952/n.a. (AUC)
Zenghui Qian et al. (102)	85/148	To identify OS in patients with LGGs	T2WI	^MRIcro	55/6	univariate Cox regression	multivariate Cox regression	#0.92/0.70(C-index)
László Papp et al. (97)	42/28	To estimate survival for glioma	11C-MET PET	^^Hybrid 3D	112/56	Genetic algorithm and Nelder–Mead method	geometric probability covering algorithms	Monte Carlo cross-validation, n.a./0.9 (M36IEP AUC)

(Continued)

TABLE 2 | Continued

Study	No. of patients involved (Training/ Test)	General purpose	CT or MRI contrast (s)	Segmentation method	Nomber of features (extraction/ selection)	Feature selection method	Classification method	Performance (Training/ Test)
Sohi Bae et al. (96)	163/54	To predict survival in patients with GBM	T1WI, T2WI, FLAIR, 3D CE-T1WI and DTI	^^signal intensity thresholding, region growing, and edge detection	796/18	Random survival forest	Random survival forest	***Overall survival: n.a./ 0.652, Progression-free survival: n.a./ 0.590(iAUC)
Niha Beig et al. (98)	83/113 (male), 47/70 (female)	To develop sexually dimorphic radiomic risk score (RRS) models that are prognostic of OS	CE-T1WI, T2WI and FLAIR	^^CapTK	105/8 (male), 287/ 6 (female)	LASSO	Cox regression	#0.73/0.88(C-index, male), 0.73/0.69(C-index, female)
Maikel Verduin et al. (99)	142/46	To established prognostic clinical features, predict IDH-mutation, MGMT-methylation and EGFR amplification develop a prognostic model for OS in GBM	CE-T1WI and T2WI	^Osirix Lite and MiM software	1197/5	XGBoost	Cox-regression	***0.72/0.73(C-index)
Philipp Kickingereder et al. (100)	120/61	To stratify patients with GBM	CE-T1WI, T2WI and FLAIR	^^ITK-SNAP	1043/386	CCC	Lasso-penalized Cox model (Coxnet algorithm)	***0.103(all layers, OS IBS); 0.089(all layers, PFS IBS)

*^ manual segmentation; ^^ semi-automatic segmentation; ^^^ full-automatic segmentation; * 5-fold cross-validation; ** leave-one out cross-validation; *** 10-fold cross-validation; # cross validation not available;*

AUC, area under the receiver operating characteristic curve; GBM, glioblastoma; BM, brain metastases; PsP, pseudoprogression; CCC, concordance correlation coefficient; IDH, isocitrate dehydrogenase; PCNSL, primary central nervous system lymphoma; TBR, tumor-to-brain ratio; CNN, convolutional neural network; LASSO, least absolute shrinkage and selection operator; MGMT, O6-methylguanineDNA-methyltransferase; EGFR, epidermal growth factor receptor; TRC, treatment-related changes; RFE, recursive feature elimination; RF, random forest; SAM, significance analysis of microarrays; SVM, support vector machine; n.a., not available; PCA, principal component analysis;

MRI, magnetic resonance imaging; mRMR, minimum redundancy maximum relevance algorithm; MRS, magnetic resonance spectroscopy; DTI, diffusion tensor imaging; DWI, diffusion-weighted imaging; FLAIR, fluid-attenuated inversion recovery; PWI, perfusion-weighted imaging; T1, T1-weighted MRI; CE-T1WI, contrast-enhanced T1-weighted MRI; T2, T2-weighted MRI; DSC, dynamic susceptibility contrast; CBV, cerebral blood volume.

neuroradiological information towards tumor penetration boundaries and heterogeneity in brain tumor patients (9, 10).

Radiogenomics

In 2016, WHO published a new guideline on classifying CNS tumors; the molecular markers are especially spotlighted to describe brain tumor entities histology features for the first time, which leads to more precise tumor cataloging (11). Furthermore, target therapies and treatment strategies for malignant brain tumor patients are also predominantly dependent on specific molecular markers, emphasizing the importance of precision oncology. With the advance of big data and bioinformatics, it is possible to detect the correlations between gene expression and radiomics features, which is known as radiogenomics. Radiogenomics is based on a common hypothesis that the dissimilarities in phenotypes of ROIs can be attributed to gene-expression patterns (12). Panth et al. further proved that genetic changes that lead to phenotypic consequences can be reflected in variations of radiomics features (13). The main tasks of radiogenomics are to investigate the correlation between germline genotypic variance and the large clinical post-radiotherapy variability (14), as well as the correlation between specific imaging features and the inherent cellular pathophysiology (15).

Radiogenomics is analogous to the combination of radiology and genomics, but people should be aware that a sole radiomics analysis without biopsy genomics confirmation is not robust enough for definitive assessment of gene expression or other contents in ROIs. On the one hand, radiomics or radiogenomics only reveals the correlation between features and genetic alterations, not the causes. On the other hand, not all of the phenotypic differences are induced only by genetic alteration, but also epigenetic changes and other factors. Thus, they actually incorporate with histopathologic examinations or sequencing, which provides confirmatory information to improve clinical decision making. Radiogenomics holds great potential for an expanding translational technology, mainly due to three characteristics:

First of all, while genomic sequencing usually uses biopsy samples from one representative part of the tumor, the radiomics data is derived from the entire tumor lesion, so the outcome can capture radiomics features on the whole. As a result, radiogenomics supplements genomic sequencing with intratumoral heterogeneity and even intertumoral heterogeneity. Several recent studies have exhibited the role of radiogenomics in identifying regional genetic heterogeneity in malignant tumor with broad genetic diversity that led to treatment resistance such as glioblastoma (GBM) (16–18).

The following advantage lies in that radiogenomics are easy, rapid, noninvasive, and dynamic, as well as cost-effective. As imaging becomes routine for patients who are suspicious of brain tumor and the estimated error rate of cancer histopathology can be as high as 23% (19–22), quantitative imaging provides additional information to avoid observer variability and indicates actual biopsy sites. For those who have contraindications of biopsy, radiomics or radiogenomics is expected to serve as a secondary substitution to guide individualized medicine. As far as costs are concerned, radiomics are usually low cost compared to biopsy; it costs around 2000 dollars for a brain biopsy in China, but less than half of that for radiomics. In addition, brain biopsy is an invasive procedure that includes risks like bleeding, seizure, infection, and even paralysis or death for key lesions like basal area and brain stem. Consequently, radiomics outweighs biopsy in costs and risks. Since radiomics or radiogenomics is easy and rapid, they enable the monitoring of the change of gene expression in the tumor's different regions, which may potentially indicate the causes of gene mutation.

Finally, there are single features strongly related to genes, and a cluster of features not significantly correlative to genes, but that have the potential to provide the information with some sort of combination, which has made progress in predicting cancer immunotherapy. By combining CT radiomics features and genomic data based on CD8B, Sun et al. developed a novel radiomics-based biomarker to predict CD8 cell count and clinical outcomes of patients' response to anti-PD-1 or anti-PD-L1, when validated by further prospective randomized trials (23). In an AI-based radiomics study by Trebeschi et al., the biomarker mainly regarding tumor proliferation could predict anti-PD-1 therapy response with an AUC of up to 0.76 for both advanced melanoma and non-small-cell lung cancer (NSCLC) patients (24). It is also promising to see more robust radiomics-based biomarkers on targeted therapy in the future, such as antiangiogenic treatment with bevacizumab (25). To improve the resolution and confidence of features' subsets related to gene modification or expression, big data from multiple centers should be collected and integrated.

PITUITARY TUMOR

Pituitary adenomas (PAs) are among the leading types of brain tumors, and the foremost frequent lesion in sellar area. Usually, the hormone hypersecretion is assessed by immunocytochemical or hormone assays to distinguish secretory tumor from non-secretory ones. And the evaluation of tumor mass, such as accurate location and volume, is based on diagnostic imaging and visual field examination. Accurate clinical diagnosis derived from tumor characteristics helps individualized treatment. Radiomics will likely never replace histopathology or hormonal diagnosis for adenomas. However, separately, AI algorithms may replace the work performed by pathologists in interpreting microscopic analyses. Recent progress in pituitary tumors, gliomas, and brain metastases are arranged and summarized in **Tables 1**, **2**.

Evaluation of Aggression

Ki-67 is only expressed in the nucleus of proliferating cells and the Ki-67 labeling index is considered to be an ideal indicator for detecting cell proliferation activity. PAs may exhibit clinically invasive or aggressive behavior, accompanied by postoperative recurrence and resistance to multidimensional therapy, which is usually evaluated by Ki-67 labeling index through immunohistochemistry. Ugga et al. collected 89 patients with available Ki-67 data and performed k-nearest neighbors (k-NN) to build a radiomics classifier using 12/1128 quantitative radiomics features to evaluate the Ki-67 labeling index based on preoperative T2WI MRI, which is effective and accurate (accuracy 91.67%) (63).

Knosp grade describes the degree of PA invasion to bilateral cavernous sinuses (CS). Niu et al. predicted CS invasion in 194 PA patients (training cohort: 97; test cohort: 97) graded 2-3 with Knosp pre-operatively by radiomics approaches deriving from contrast-enhanced T1 (CE-T1) and T2WI MRI (64). They applied LASSO to select three important features and establish a classifier using support vector machine (SVM), which yielded decent AUC values (training cohort: 0.852, test cohorts: 0.826).

Grading

The regulation of adenohypophyseal cell differentiation and hormone secretion are operated by a series of transcription factors, including Tpit, Pit-1, and SF-1. Peng et al. involved 235 patients with pituitary adenoma (PA), and 18 quantitative imaging features were verified as significant to train SVM, k-NN, and Naïve Bayes (NBs) models to classify the transcription factor types of PAs. Among the three models, the SVM model showed the best performance (AUC 0.9549) whereas the K-NN (AUC 0.9266) and NBs (AUC 0.9324) models displayed lower performance and they found better performance in T2-weighted than Tl-weighted and CE-T1 (65).

Zhang et al. worked on differential diagnosis of non-functioning pituitary adenomas (NFPAs) subtypes from other subtypes preoperatively (66). They enrolled 75 patients as the training cohort and 37 patients as the test cohort, and derived complete T1-WI and CE-T1 MRI. The top three T1-WI imaging features, rather than CE-T1 imaging features, were ultimately selected using mRMR to fit a Radial Basis Function (RBF)-SVM predictive signature. A nomogram incorporated clinical characteristics and the radiomics signature corresponding to the best predictive model for individual prediction. Finally, the calibration of the nomogram was presented with a concordance index (CI) (training cohort: 0.854, test cohort: 0.857).

Prediction of Treatment Response

Prolactinoma is the most common secretory PA, with prime treatment being dopamine agonists (DA) such as bromocriptine. To assess prolactinoma patients' response to DA before initiating the treatment plan, Park et al. investigated a total of 177 prolactinoma patients' coronal T2-weighed MRIs and set up a radiomics predictor with an AUC of 0.81 in both training and test cohorts (26).

Acromegaly is a severe complication that leads to poor prognosis most frequently caused by somatotroph PAs that secrete excessive growth hormone (GH). For those who are not suitable for surgery or have severe symptoms, somatostatin receptor ligand (SRL) treatment is usually applied to reduce the volume of mass preoperatively (67–69). To predict the treatment response of SRL ahead of surgical resection, radiomics features from MRI are selected to predict the histological granulation pattern. Park et al. set up a radiomics prediction model based on 69 GH-secreting PA patients and the model showed an AUC of 73.7% (70).Fan et al. proposed a radiomics prediction model of invasive functional pituitary adenoma (IFPA) working on the treatment responses before surgery (71). The prediction model was incorporated with a radiomics signature based on seven selected features derived from MRI and Knosp grade of each IFPA patient. The discrimination abilities and calibration of this yielded good performance, with AUCs (training cohort: 0.832, test cohort: 0.811).

Prediction of Recurrence

Early progression/recurrence (P/R) is a poor prognostic factor of macro NFPAs that is reported in 25–55% patients after surgical resection (72–75). Zhang et al. established a SVM radiomics model based on three selected features from 50 patients' 214 preoperative and postoperative follow-ups MRI features extracted from CE-T1 and T2WI yielded an overall accuracy of 82% and AUC of 0.78 discriminating P/R NFPAs from non-P/R ones using the original mask tumor ROI (27). They also calculated SVM scores for each patient and found that higher scores may correlate with shorter PFS. Regarding SVM score for differentiation of P/R, the optimal cut-off value was calculated which means patients with SVM scores higher than 0.537 tended to exhibit shorter PSF and the corresponding AUC (0.87) were obtained.

Machado et al. applied 3D CE-T1 MRI of patients preoperatively and segmented two and three-dimensional regions. They evaluated the 255 extracted radiomics features of 27 patients with NFPA and selected six features for two-dimensional machine learning radiomics models and 13 for three-dimensional models (76). The results showed 3D-feature based models have superior discrimination ability to classify NFPAs recurrent and stable lesions than 2D-feature based models, with their accuracies of up to 96.3% compared to accuracies of 92.6% for models solely based on 2D features.

GLIOMAS

Gliomas account for the first leading primary brain and other CNS tumor in adults (25.1%) (77), making up approximately 80.8% among primary malignant brain and other CNS tumors with certain diagnoses (11).As the majority of gliomas (57.7%), GBM accounts for the foremost and lethal primary malignant brain tumor in neurosurgery, whose overall incidence rate is estimated 0.0032% in American adults (78). Though multimodal treatment protocols, including maximal safe surgical resection and adjuvant radiation therapy with concurrent chemotherapy mainly temozolomide (TMZ), are standardly given to GBM patients, they still suffer a crushingly adverse prognosis with 14.6 months of median OS (79).

Evaluation of Mutation Status

Isocitrate dehydrogenase (IDH) mutations are recognized as a good prognostic factor in early clinical intervention and can be integrated into routine clinical practice such as pathological examination *via* radiomics analysis, immunohistochemistry, flow cytometry, standard PCR, and/or sequencing techniques (80). All IDH-mutant diffuse astrocytic tumors are classified as a single type of tumor (astrocytoma, IDH mutation, grade 2/3/4) and all GBM only included IDH wild type tumors in the 2021 WHO CNS 5. The 1p19q co-deleted tumors are exclusively oligodendrogliomas. And most oligodendrogliomas with 1p/19q co-deleted, which indicates poor prognosis, are accompanied by IDH1 and IDH2 mutation. In 2016 WHO CNS 4, the guideline of gliomas classification incorporated five molecular subtypes of diffuse gliomas based on IDH mutation and 1p/19q codeletion status.

Lu et al. built a multilevel quantitative imaging model based on CE-T1 image, T2 FLAIR, T2WI, DWI, and ADC to recognize IDH and 1p/19q genotypes of glioma and further classification of five molecular types (33). The training cohort involved 214 patients and an additional independent cohort involved 70 patients for external test. The IDH and 1p/19q classifier using SVM models was established in the training cohort, yielding areas under receiver operating characteristic (ROC) curves ranging from 0.922 to 0.975, and accuracies ranging from 87.7% to 96.1%. Correspondingly in the test cohort it showed accuracies between 80.0% and 91.7%. On classifying five molecular subtypes, the trained classifier with the MR radiomics phenotypes as sole source yielded an accuracy of 81.8%, which further reached a higher accuracy of 89.2% in the existence of histology diagnosis. Gutsche et al. implemented FET PET to enhance the diagnostic performance of their radiomics features on IDH genotype identification (34). The repeatability of the features was evaluated by calculating the intraclass correlation coefficient (ICC) and 297 features with robust discrimination ability were finally selected.

Deep convolutional neural networks (CNNs) and radiomics share the same procedure but have separate characteristics regarding radiological evaluation on IDH genotypes. Choi et al. reviewed 1166 preoperative CE-T1, T2, and FLAIR neuroimaging of gliomas grading II-IV derived from three centers and developed a CNN-based fully automated model hybridizing conventional MRI that integrated 2D tumor signal intensity and quantitative radiological features from 3D tumor shape and location, which was reproducible and generalizable for noninvasive characterization of IDH status in gliomas (35). Based on CE-T1, T2WI, and FLAIR from 1166 patients with gliomas (training cohort:727; internal test cohort: 129; external test cohort: 310), 20 out of 24 extracted features were selected and the classifier reached an AUC of 0.96, 0.94, and 0.86

respectively for training cohort, internal validation cohort, and external validation cohort.

The DNA-repairing enzyme O-6-methylguanine-DNA methyltransferase (MGMT) restores cytotoxic lesions in DNA caused by temozolomide chemotherapy, thus leading to drug resistance. Methylation epigenetically silenced *MGMT* has a connection with a better treatment response and a better prognosis than the unmethylated promoter (81). Based on a group of 105 patients with grade II-IV astrocytoma, Wei et al. established a radiomics model for *MGMT* promoter methylation prediction with supreme power (AUC: training cohort: 0.925, test cohort: 0.902), which successfully bisected the group into high-risk and low-risk subgroups for OS followed by temozolomide chemotherapy (39). *ATRX* mutation is another good prognostic factor that usually appears in LGGs accompanied by *IDH* mutation. Li et al. included 95 patients with LGGs and built a radiomics predictor of ATRX alterations, which was subsequently validated in an external cohort of 91 patients with good performance (AUC: training cohort: 0.94, internal test cohort: 0.925 and external test cohort: 0.725) (40).

Haubold et al. assessed the potential of multimodal imaging and radiomics algorithms to predict the grading and common mutations including MGMT of 42 patients with suspicious primary brain tumor (41). They applied 18F-FET PET along with MR Fingerprinting and extracted 19284 features from each patient, which were further divided into 32 for 1p19q codeletion, 64 for *IDH1*, 8 for *ATRX*, and 16 for *MGMT*. And through 5-fold cross-validation the AUCs in predicting the mutation were respectively assessed, with 1p19q for 97.8%, *IDH1* for 88.7%, *ATRX* for 85.1%, and *MGMT* for 75.7%. The 2016 WHO grading model yielded an AUC of 81.8% while AUC of discriminating LGG from HGG was 85.2%. Su et al. further investigated grading along with proliferation levels in 220 patients with various grades of gliomas (43). When combining radiomics features of multi-contrast MRI (T2WI fast-echo images (T2FSE), T1WI, FLAIR, CE-T1WI, DWI, ADC, PWI and CBF), the models displayed the highest AUC (0.911 for LGGs and HGGs, 0.896 for grades II–III, 0.997 for grades II–IV, 0.881 for grades III–IV, and 0.936 for levels of Ki-67 labeling index).

It has been reported that *CIC* mutation promotes glioma cell proliferation, differentiation, and aggression and results in a poor outcome (82–84). However, Zhang et al. found that patients with LGGs (*IDH* mutation) or oligodendroglioma (*IDH* mutation and 1p/19q codeletion) combined with *CIC* mutations may have better prognosis (42). As shown in MRI, LGGs (*IDH* mutation) with *CIC* mutation illustrate visually less malignant manifestations, such as fairer necrosis and more homogeneity among the tumor volume. They further developed a radiomics model to predict the *CIC* alterations based on 11 features derived from 120 patients with LGGs (AUC: 0.985).

Epidermal growth factor receptor (EGFR) variants are reported in 57% of GBM specimens. Among these, a deletion from exons 2–7, EGFRvIII is the most frequent EGFR variant, and extracellular domain (ECD) missense mutations like A289D/T/V, R108G/K, and G598V are the most frequent EGFR deletion comprising 10%–15% of transcription products while the deletion was found to co-occur with amplification (85). $EGFR^{A289V}$ mutation has the most negative survival impact, which was reinforced by Binder's study involving 260 patients with *de novo* GBM (38). To investigate the negative prognostic effect of $EGFR^{A289D/T/V}$, Binder et al. firstly did quantitative imaging analysis comprising six different MRI modalities and extracted 2104 quantitative imaging phenomic features which were further reduced to a more manageable set of 299 using a multivariate classification framework (86). To promote radiographic interpretability, an experienced neuroradiologist further filtered these features into 17. The MRI signatures based on selected features presented a picture of decreased T1 signal but higher CE-T1 signal, higher T2 values, higher peak height (PH) values, and relative cerebral blood volume (rCBV) in $EGFR^{A289D/T/V}$ mutant tumors region, indicating higher water content, hyperproliferation, and increased invasion in tumor region with $EGFR^{A289V}$. The peritumoral edema region presented reduced fractional anisotropy (FA) generated by Diffusion Tensor Imaging (DTI) for cases with $EGFR^{A289D/T/V}$ mutations, suggesting decreased tissue organization and homogeneity. Taking MRI signature and following modified cell lines *in vivo*, they demonstrated higher proliferation, increased aggressiveness, and shorter OS in $EGFR^{A289D/T/V}$. To explore the mechanism *in vivo*, they inhibited two main signaling pathways of EGFR, RAS/RAK/ERK and PIK3CA/AKT, which revealed that A289V-induced EGFR activation mediates phosphorylated ERK and augments MMP1 expression which cause hyperproliferation and invasion. Finally, mAb806 targeting therapy was examined in $EGFR^{A289V}$ mice models and was proven to be a potential therapeutic option as the mAb806 antibody reduced the tumor burden, inhabited tumor growth, and improved animal survival.

Glioma cells connect their microenvironment in a two-way street, mainly through cytokines and matrix proteins. POSTN is a secretory extracellular matrix protein made up of gliomas cells. Previous studies showed POSTN plays a role in neovascularization, endothelial junction formation decrease, stem cell maintenance, and macrophage recruitment (87, 88). Subsequent studies revealed POSTN in glioma grade, recurrence, and resistance to bevacizumab monoclonal antibody against VEGF-A (89, 90). Zinn et al. found GBM patients with different POSTN expression levels in association with distinct imaging features, which can be utilized in radiomics for prediction (32). They extracted 2480 radiomics features respectively in GBM patients and GSC-derived orthotopic tumors mice and selected 48 and 17 features respectively to build two classifiers (GBM AUC: 76.56%; mice, AUC: 92.26%).

F3T3 is a novel fusion proto-oncogene incorporating FGFR3-TACC3 found in approximately 3% of gliomas that functions as an important part in the activation of oxidative phosphorylation and mitochondrial metabolism. Though the foremost energy metabolism pathway of tumor is anaerobic glycolysis, GBM with F3T3 mutation depends on noncanonical mitochondrial pathway. Thus, F3T3 may serve as a potential target for targeting therapies such as mitochondrial inhibitors (91). Stefano et al. showed that, in the midst of IDH-WT tumors,

F3T3-positive gliomas exhibit distinct molecular, radiological, and clinical features and possess a more optimistic clinical outcome independent of their grading. Their radiomics data composed 66 patients as training cohort and 78 patients from another institution as test cohort and identified F3T3-positive patients with good accuracy. They successfully built a classifier towards F3T3 mutation status (AUC: 0.87(training)/0.745(test)) and a model composing clinical, genetic, and radiomics profile to estimate the F3T3-positive patients' OS as presenting the best concordance (C-index: 0.81). They further implemented multiple optimization techniques (SCCAN) to inspect the tropism of F3T3 gliomas for specific intracranial ROIs and finally located cortical and subcortical regions, especially insula and temporal lobe (36).

H3 K27M mutant occurs within the histone-3 gene (H3F3A) wherein an amino acid recurrently converts from lysine to methionine in the site 27, and H3 K27M-positive diffuse midline glioma is listed separately graded as IV in 2016 WHO classification (92). According to a report, H3 K27M-positive gliomas in thalamus area tend to result in a shorter median OS in pediatric patients than in H3 K27M-WT ones (93). Furthermore, there have been several epigenetic-targeted treatments towards H3K27M and an immunological study provided evidence for immunotherapeutic approaches like mutation-specific vaccines targeting H3K27M (94). Su et al. carried out a retrospective study on automated classification of H3 K27M genotypes. The Tree-based Pipeline Optimization Tool (TPOT), a method that automatically conducts feature and model selection procedure, along with pipeline optimization, was highlighted. The study included 40 H3 K27M-positive patients and 60 WT, of whom 75% were randomly grouped into the training cohort while 25% into test cohort. After extracting 99 features from FLAIR, TPOT finally refined 10 more manageable radiomics features and generated ten prediction models. The optimal model is generated through comparison of accuracy metrics. The model exhibiting the best performance in the test cohort yielded the highest average precision of 0.911 and AUC of 0.903, while validation in an independent validation dataset observed an average precision of 0.855 and an AUC of 0.85 (37).

Monitoring the core signaling pathway of GBM may reveal the tumor evolution, allow early clinical intervention, and enhance patients' management (95). Park et al. built a radiogenomic classifier based on patients with IDH-WT GBM certified by next-generation sequencing (NGS), which noninvasively predicts retinoblastoma 1, p53, and Receptor tyrosine kinase (RTK) core signaling pathways (31). In this study, 85 patients were classified into the training cohorts in total and 35 into test cohorts, and their T1WI, T2WI, DWI, FLAIR, CE-T1WI, and perfusion MRI-like dynamic susceptibility contrast (DSC) were acquired for radiomics analysis. For each core signaling pathway, 71, 17, and 35 features passed extraction, and finally the top 5 features were selected respectively. Three models were evaluated, presented as AUC (RTK, training cohort: 0.87, test cohort: 0.88; p53, training cohort: 0.80, test cohort: 0.76; retinoblastoma, training cohort: 0.84, test cohort: 0.81). *PTEN* is a tumor suppressor gene participating in both ATK and RTK signaling pathways and the deficiency of *PTEN* is considered to be the main feature of GBM (96, 97). In addition to relying on sequencing and immunohistochemistry to detect *PTEN* alterations, Li et al. established a noninvasive radiomics method with good performance in guiding targeted therapy (AUC: training cohort: 0.925, test cohort: 0.787) (44).

Differential Diagnosis

The radiological features of gliomas usually lack specificity, involving spherical well-encapsulated shape with ring enhancement indicating tumor angiogenesis and prominent peritumoral edema. The lesions are mostly multiple and are usually located at watershed or grey-white junction with white matter fiber bundles erosion. Further, the radiological features between malignant gliomas and lymphomas are analogous and there is also similarity in abscesses, infections, demyelinating diseases, and vascular lesions. The radiological features are also likely to be interfered with by hemorrhage, melanin, and paramagnetic ions.

Brain metastases (BMs) take up the second most common type of malignant brain neoplasms in adults preceded by GBM (98, 99). Early diagnosis is the key to appropriate therapies since the strategies for these two tumors are distinct with different local control rates and intervention prognosis: the prior treatment for GBM is maximum-safe surgery resection, following adjuvant radiotherapy and chemotherapy (100), while regarding BM the more effective and less invasive treatment is stereotactic radiosurgery (101). Qian et al. assessed high-dimensional radiomics features from T1-WI, T2-WI, and CE-T1 to distinguish GBM from solitary BM (47). In the retrospective study, patient's population, including 242 GBM and 170 solitary BM, was randomly grouped (training cohort: 227, test cohorts: 185). An amount of 1303 radiomics features passed extraction, which were then refined by twelve feature selection methods. Thirteen classifiers were generated and all yielded excellent predictive efficacy with AUC≥0.95 in the training cohort. Through ROC curve analysis they found out that the combination of SVM and LASSO classifiers had the best prediction value in the test cohort (AUC: 0.90).

Primary central nervous system lymphoma (PCNSL) shares radiological similarities with GBM when solely using ADC parameter, due to some overlaps in ADC values. Kang et al. evaluated the feasibility of a radiomics model for the differentiation of atypical PCNSL and GBM based on ADC (45). The patient population in the training cohort contained 112 patients, while the population in the test cohort involved 42 patients for internal and 42 for external validation sets. They combined 12 feature selection methods with 8 classification methods using 5~50 selected features and optimized 8 ADC radiomics models. The prediction performance and stability were subsequently measured by each AUC and relative standard deviation (RSD) of each model. As a result, the ADC model combining RFE feature selection with RF classification yielded the highest diagnostic value with an AUC of 0.983 in the training cohort. The ADC model showed robustness exceling

expert readers and was further assessed respectively in the internal validation cohort (AUC 0.984) and external validation cohort (AUC 0.944) to promote generalizability.

To standardize the procedure and improve efficacy, Wu et al. generated a novel radiomics system utilizing feature extraction and selection methods and classification framework based on dictionary learning and sparse representation (46). Simply using T2WI and CE-T1, they tested the technical feasibility of the system using 49 selected radiomics features out of 968 features extracted from 102 patients with PCNSL or GBM (training: test=67:35). The sparse representation radiomics system had superior PCNSL and GBM differentiation performance (training cohort: 98.51% accuracy, test cohort: 94.51% accuracy). Furthermore, the IDH1 prediction performance of the novel system exceeded traditional methods based on calculation by 11%.

Prediction of Treatment Effect and Recurrence

Tumor recurrence in early posttreatment stage is commonly reported in HGGs. Assessing posttreatment MRI changes according to RANO standard within 0-72 hours is a common and effective method to evaluate the degree of surgery. Pseudoprogression (PsP) is a diagnostic dilemma presented as expanded and/or new regions of edema and enhancement, especially within the 2-5 months from the initiation of adjunctive therapies, which mimics tumor recurrence and radiation necrosis. The mechanism of PsP may be attributed to three factors: (1) non-tumor tissue chemoradiation damage, e.g., hemorrhage, ischemia, aseptic inflammation, edema and necrosis; (2) blood-brain barrier breakdown; or (3) other factors, e.g., signal artefacts from metal implants.

A recent retrospective radiomics study comprising 76 patients of histopathology-proved progressive disease (PD) and 22 of PsP from three centers by Elshafeey et al. provided evidence on perfusion MRI on accurately discriminating PD and PsP. Its reported model based solely on Ktrans maps had matching diagnostic value with the rCBV model in discriminating between PsP and PD. The final prediction model combining Ktrans with rCBV maps generated by SVM used the top 60 radiomics features ranking with MRMR, which achieved an accuracy of 90.82% and an AUC of 89.10% in discriminating between PsP and PD. Subsequent validation also showed statistical significance by LOOCV (AUC 89%) (50). Based on dynamic FET PET radiomics, Lohmann et al. aimed at establishing a reliable diagnostic test for differentiating PsP from tumor progression in gliomas patients (52). In the tumor segmentation process, data augmentation was implemented to increase the number of datasets from 34 patients to 102. The radiomics model was automatically generated using TPOT based on random forest classification, with an AUC of 0.74 in both training and test cohorts.

Kim et al. investigated the feasibility of multiparametric MRI radiomics incorporating diffusion and perfusion to identify tumor recurrence within 3 months following standard therapy (48). They developed and validated a radiomics model comprising CE-T1WI, FLAIR, ADC, and CBV maps using 61 patients as training cohort and 23 patients as validation cohort. Initially 6472 features were extracted and then 12 significant radiomics features passed selection using LASSO to construct the integrated model. And the model presented best diagnostic performance (AUC, 0.90) over any single imaging technique or parameter model. The internal validation (AUC, 0.96) and external validation (AUC, 0.85) cohorts strengthened the outcome.

Towards differentiating tumor recurrence from radiation necrosis, Wang et al. carried out a radiomics study involving 112 patients as training cohort and 48 patients as test cohort. The multidimensional quantitative model integrated clinical information (patients' individual features and gliomas grade) and radiomics information (MRI techniques (T1WI, T2WI, CE-T1WI and FLAIR) and PET images using both 18F-FDG and 11C-MET), while the radiomics model only included radiomics information. Fifteen textural features were selected from the images for the construction of radiomics model and integrated model. And the integrated model showed significant superiority over radiomics model (both training and test cohorts: $p < 0.001$) and was proved to be accurate and effective in the prediction of differentiating postoperative tumor recurrence from radiation necrosis (training cohort: AUC 0.988, test cohort: AUC 0.914) (53).

Brain necrosis after radiotherapy is a common complication in approximately 3%–24% of patients (102), mostly with primary or metastatic cancer of the head, neck, and CNS. Bevacizumab has shown its potential in symptomatic relief and radiographic response compared with general corticosteroid therapy in randomized study (103), however, some patients are unable to gain benefit or even worsen. To predict the treatment effect of bevacizumab in brain necrosis patients, Cai et al. developed a radiomics model based on a total of 149 patients including 42 as an external validation cohort, which yielded an AUC of 0.916 in the training cohort and 0.912 and 0.827 in the validation cohorts (51).

Immune checkpoint inhibitors (ICI) hold great promise for GBM treatment, however the suppressive microenvironment of GBM characterized by poor antigen presentation and low T-cell activation and infiltration limits the ICI application. Aslan et al. investigated mechanisms of resistance to ICIs blocking PD-1 and CTLA-4 and acquired immune heterogeneity in the allogeneic intracranial inoculated mice with Gl261 tumor cells (49). To determine the response of ICI in mice post inoculation and identity PsP, they built a radiomics signature based on CE-T1WI and T2-WI. From 101 mice inoculated with Gl261 tumor cell before and during ICI treatment, they extracted 423 features and built a gradient boosting classifier containing all 423 z-score-normalized radiomics features with an accuracy of 82.7%. Subsequent *in-vivo* and *ex-vivo* experiments proofed that PD-L1/PD-1/CD80 axis plays an important role in ICI resistance induced by CD4 T cell suppression, tumor-associated macrophages, and Treg extension in the microenvironment of GBM.

Predicting Patient's OS and Complications

According to a large-scale randomized trial, the median survival time of GBM is 14-15 months, which can be prolonged by adjuvant temozolomide with radiotherapy (104). However, the current radiotherapy plans ignore the biological heterogeneity of individuals and use the same dose, resulting in significant difference in patients' OS (105). Radiomics and radiogenomics can provide an imaging biomarker on predicting the individual radiotherapeutic response, which helps to adjust the dose and make a personalized treatment plan. By combining the clinical risk factors and radiomics signature which was built with 152 patients with GBM to predict the radiotherapeutic response, Pan et al. developed a nomogram to predict the OS, with C-indexes up to 0.764 and 0.758 respectively in the training cohort and external validation cohort (54).

The study by Dastmalchian et al., including 31 patients with GBM, LGGs, and metastases and 20 top selected features, proved that the radiomics approach has robust potentiality to differentiate between these tumors and to predict OS of GBM (56). They found a significant difference between patients with different selected features such as T1/T2 entropy and secondary features like high-gray run emphasis ($p < 0.05$). And the cut-off values dichotomizing the GBM patients' median survival were calculated by grading these features. For example, lower entropy values in solid tumor regions (p: 0.034) in T2 maps correlate with longer survival of 11 months and 6.7 months for those below the cut-off value, and higher entropy values in peritumoral white matter regions (p: 0.009) in T1 maps correlate with longer survival of 18 months and 6.8 months for those below the cut-off value.

Since the mutation of IDH1-R132H and MGMT in GBM patients is strongly associated with the OS and PFS, Bae integrated radiomics with clinical and genetic profiles and built several models to predict the prognosis of 217 GBM patients (training: test=3:1) (59). From 796 features derived in T1WI, T2WI, FLAIR, postcontrast 3D T1WI, and DTI, they selected 18 significant features and trained multilayers RSF models. Except for the integrated model, the model containing only radiomics features was the most significant with successful validation in the test cohort (OS: iAUC 0.652, PFS: iAUC 0.590).

Papp et al. evaluated the prediction value on dichotomized OS using an integrated model comprising 56 features including 11C-MET PET radiomics characteristics *in vivo*, histopathological characteristics *ex vivo*, along with patients' individual information to predict survival in glioma patients without treatment (58). The cut-off value determines 36 months as the survival prediction threshold and the prediction weight for each model was assessed in training cohorts and the validation cohorts. When it came to validation, they introduced the Monte Carlo cross-validation (MCCV) different from k-fold validation in that the sample may appear multiple times in the same set (training set/test set). The MCCV proved the highest AUC for the integrated model as 0.9, following the patient-based and histopathology-based models.

It has been reported that in GBM patients, females exhibit longer OS compared to males, which may be associated with hormonal, metabolic, and immune variances. Based on the discovery, Beig et al. developed sexually dimorphic radiomics risk score (RRS) models to predict patients' OS. The OS prediction model combines age molecular features, extent of resection, and RRS, showing good performance in both male and female cohorts [0.73/0.88(C-index, male), 0.73/0.69(C-index, female)]. By further analyzing radiogenomics associations between MRI-based phenotypes and transcriptomic data correspondingly, they found that RRS is associated with a series of biological activities including angiogenesis, apoptosis, cell differentiation, cell proliferation, and cell adhesion (60).

Verduin et al. involved the training (n = 142) and validation cohort (n = 46) to establish a combined model for prognosis of OS in IDH-WT GBM patients based on quantitative radiological features, qualitative Visually Accessible Rembrandt Images (VASARI) features, and clinical information. The accuracy and reproducibility of the combined model was analyzed using Harrell's C-index (training cohort: 0.72, validation cohort: 0.73). They additionally developed a prediction model towards molecular mutation status comprising IDH, MGMT methylation, and EGFR amplification in 95 patients for the training cohort and 38 patients for the validation cohort. In this model, 5 VASARI and 5 radiomics features mainly selected from T2WI were considered to be most prognostically relevant, with performance towards MGMT methylation (AUC: 0.667) and EGFR amplification (AUC: 0.707) yielding significance in external validation cohorts (61).

A radiomics study on PFS and OS stratification by Kickingereder et al. included 181 GBM patients available of imaging information (CE-T1, FLAIR, and T2WI), DNA methylation profiling (MGMT methylation status and global DNA methylation pattern), treatment (surgery, TMZ chemotherapy and/or radiation), and patients' individual information (62). In the prognostic analysis, a total of 386 features were selected independently in a test-retest MRI cohort. Subsequently, 8 of these were further used to construct a radiomics signature using sole radiomics information. When using only epigenetic and clinical information, the prediction error for PFS (29%) and 37% for OS (27%) is not appreciable, which were reduced by 36% for a model after integrating radiomics signature. The radiomics signature showed significance beyond models using other information ($P \leq 0.01$).

The epilepsy complicated by LGGs is mainly attributed to compression and stimulation of the brain tumors that cause the degeneration and gliosis of the brain cells around the tumors which constitute the epileptic foci complex. Wang Y aimed at predicting epilepsy types to guide more targeted antiepileptic therapy in a retrospective study. A novel radiomics nomogram was developed with 4 selected discriminative MRI features regarding location and molecular background in 205 LGG patients, which displayed excellent quantitative clinical prediction performance (AUC: 0.863) (106). Qian et al. suggested a radiomics risk score to alternatively identify the OS in LGGs. When combined with independent clinical prognostic parameters such as WHO grade, age at diagnosis, and seizure, the nomogram based on the risk score exhibited high prognostic accuracy (C-index: training cohort: 0.92,

test cohort: 0.70) (57). They subsequently implemented radiogenomic analysis of high-risk positively associated genes, further revealing the underlying correlated biological processes including hypoxia, angiogenesis, and apoptosis. For the prediction of PFS, Liu et al. worked out a practical nomogram incorporating clinicopathologic factors and a radiomics signature based on 300 patients with LGGs (C-index, training cohort: 0.684; validation cohort: 0.823) (55), and demonstrated similar biological processes through radiogenomic analysis.

BRAIN METASTASES

Approximately 20% of the cancer patients with other primary sites develop brain metastases, outnumbering primary brain tumors 10:1, but the actual statistic is estimated to be even more since plenty of them do not go through regular MRI examination. The top three extracranial primary cancer types with high intracranial metastatic tendency are lung cancer, breast cancer, and melanoma, which respectively have incidences up to 20-56%, 5-20%, and 7-16% (107–110). Meanwhile, the incidence of brain metastases' occurrence after primary cancer varies according to race, age, and primary cancer. Brain metastases' main symptom is parallel to space occupying lesions, which varies with the lesion location.

Evaluation of Mutation Status

Since there is inter-heterogeneity between the primary tumor and metastases, assessing the mutation status in the metastases region and comparing with the primary tumor are meaningful in guiding individualized treatment. EGFR inhibitors, such as erlotinib and gefitinib, received distinct responses in GBM and NSCLCs patients originated from different EGFR mutation sites. In NSCLCs, the mutation sites typically locate in the kinase domain that facilitate sensitivity to first-generation EGFR inhibitors. Unlike NSCLCs, the mutation sites of GBMs locate in extracellular domain that promote resistance (111).

Ahn used CE-T1 MRI to predict the EGFR mutation in histologically certified primary lung cancer patients' brain metastases (33 with EGFR WT, 29 with EGFR mutation) (112). Among all the combination of 7 feature selection methods and 4 classification methods, the RF classification model applying RF selection yielded highest AUC of 86.81% on predicting EGFR mutation status. Subsequent analyses subgrouping BMs by measurable size revealed smaller BMs correlate with better discrimination capacity (AUC 89.09% in the small BMs subgroup, combining SVM classification with RF selection).

Park used DTI and T1-contrast to classify the EGFR mutation in 99 BMs from 51 NSCLC patients, verified by biopsy. Among all the combinations of 5 feature selection methods and 4 classification methods, the linear discriminant algorithm classifier using 5 features selected by tree-based methods showed the best diagnostic performance, resulting in an AUC of 0.73 (113).

Chen did a retrospective study using CE-T1, T2WI, and FLAIR to predict the mutation on EGFR, ALK, and KRAS mutation in BMs from patients diagnosed primary lung cancer, verified by genotype testing. The model on EGFR, ALK, and KRAS incorporating both radiomics and clinical information resulted in AUC values of 0.912, 0.915, and 0.985 (114).

Identifying Primary Tumor

The clinical manifestations of BMs are analogous to primary brain tumors. Generally, systemic metastases, cachexia, and multiple foci in CNS may indicate BMs, however, there are up to 15% with unknown primary tumor (115). Kniep retrospectively studied 189 patients with primary breast cancer, lung cancer (NSCLC and SCLC), gastric cancer, or melanoma, who developed BMs, and analyzed CE-T1 and nonenhanced T1WI and FLAIR through machine learning algorithm (29). The results showed that all the RF classifiers surpassed senior neuroradiologists' reading. After combing radiomics and clinical data, the 5-class model showed best prediction performance with lowest AUC (0.64) for NSCLC and highest AUC (0.82) for melanoma.

Prediction of Recurrence

As noninvasive treatments such as radiation and chemotherapy have more extensive application in BMs, the most common application of radiomics in BM in recent years may be prediction of treatment and progression.

Prasanna proposed a novel entropy feature called co-occurrence of local anisotropic gradient orientations (COLLAGE) which is of great prognostic value in evaluating radiation necrosis and tumor recurrence on gadolinium-contrast T1WI (116). They proved in 75 patients with metastatic brain tumors that, with additional independent multisite validation, COLLAGE features exhibited statistically significant different skewness ($P < 0.05$) in recurrent tumor patients compared to patients with pure tumor and cerebral radiation necrosis > 80%.

Huang retrospectively analyzed 161 patients with NSCLC (576 brain metastases) postoperatively by Gamma Knife radiosurgery and found zone percentage related to progression (117). After feature selection by consensus clustering, analysis of univariate Cox proportional hazards model comprising clinical variables, and radiomics features revealed potential prognostic factors that were subsequently selected to build a multivariate Cox proportional hazards model, which indicated that a textural feature called higher zone percentage was independently pertained with higher local tumor control rates (HR 0.712; $P = 0.022$). Similar to the result, multivariate proportional hazards model in cause-specific condition also filtered higher zone percentage (HR 0.699; $P = 0.014$).

To predict diagnosing treatment effect after stereotactic radiosurgery, Peng investigated 82 lesions of BM with obvious progression from 66 patients who underwent SRS based on CE-T1 and T2WI MRI (30). Five top-performing radiomics features out of 51 extracted features filtered by univariate logistic regression were selected to build a subsequent hybrid IsoSVM model, which was assessed by the LOOCV (AUC 0.79). Mouraviev retrospectively analyzed 408 BM lesion in 87 patients who underwent SRS based on their pretreatment CE-

T1, T2WI, and FLAIR (28). For 440 extracted radiomics features, they applied RF feature importance and ranked these features for selection. The top 12 features comprising radiomics and clinical data are optimized for best prediction model, with the highest AUC (mean = 0.793).

LIMITATIONS

Radiomics is a rapidly expanding field and is still in extensive clinical exploration stage, with many obstacles to overcome. We may discuss the limitations from the aspects of standardization, robustness, repeatability, reproducibility, and generalizability.

Standardization is the basis of robustness, producibility, and generalizability. Current standards lack results validation, incomplete results reports, and unidentified confounding variables in the source database, especially for retrospective data. To solve the above problems and standardize radiomics-specific reporting, Lambin et al. put forward an evaluation system comprising 16 weighted metrics to determine the workflow completeness, model quality, and clinical adaptation potentiality of radiomics studies, in the form of the radiomics quality score (RQS) (12). The establishment of RQS extended a number of initiatives, such as the Transparent Reporting of a multivariable prediction model for Individual Prognosis OR Diagnosis (TRIPOD) consensus (118).

As is discussed in the Radiogenomics section, radiomics and radiogenomics can only identify the correlation, thus lacking robustness and credibility without tissue biopsy. As for radiomics itself, the accurate segmentation of ROIs is the most challenging step that largely affects the robustness of outcome. Due to a tumor's heterogeneity and polymorphism, manual segmentation is used in most imaging studies. Its advantages are high accuracy and fine delineation of irregular tumor boundaries, but it is greatly affected by subjective factors and is time-consuming and inefficient, with low repeatability. Recently, novel volume data segmentation methods based on deep learning models, such as CNNs named after the shape of the feature map structure (U-net (119) V-net (120), W-net, UNet++ (121) and Y-net), and DeepMedic have made a breakthrough in the clinical radiology segmentation. Current studies demonstrated the utility of CAD, which combines automated brain tumor segmentation with radiomics, in helping physicians to detect following initial observation (122, 123). Most normal tissues like bones and organs can be segmented semi-automatedly or fully automatedly.

However, current protocols of autosegmentation approaches are diverse and lack unified standards. From the studies we reviewed, more intelligent algorithms such as deep learning are rarely used in radiomics of brain tumor compared to lung cancer, prostate cancer, and colorectal cancer. Current segmentation, feature selection, and classification methods in brain tumors are mainly manual operations using shallow machine learning methods, such as random forest, SVM, and LASSO. What's more, the ratio of articles with test-retest analysis is low for currently available original research, which also adds doubts upon repeatability and reproducibility of radiomics analysis. Though the sensitivity, specificity, and/or AUC of the reviewed studies are considerable, few are prospective studies that were later followed up or confirmed by biopsy.

Another problem is that the clinical translation of radiomics studies in multicenter studies faces difficulty in repeatability and reproducibility. For example, MR images may capture noise caused by physiological motion, magnetic field, eddy current, and unsteadiness of the scanning hardware. Then during image reconstruction, noise is post-processed to be wiped out prior to ROIs determination (124). As MR images omit physical parameters, such as magnetic field strength and voxel size, the various settings of image acquisition, reconstruction algorithm, and image processing makes MR radiomics more challenging than CT to ensure repeatability and reproducibility (125–127).

To test the generalizability of prediction model, there are the internal validation and external validation using other centers' data. Current studies in brain tumor still lack big populations, especially from multiple centers. A systematic review by Park et al. evaluated 51 original radiomics research articles in neuro-oncology with RQS and showed that only 29.4% performed external validation, with few studies discussing clinical utility and none of them conducting phantom study or cost-effectiveness analysis (128).

There are also ethic problems in that, while the development of AI algorithms requires not only fundamental techniques, but also legislation and perhaps ethics, there are issues on whether researchers or governments should be motivated to share private validated data for machine learning (129). What's more, it is possible for AI algorithms to be tampered with improper intention to make profits.

FURTHER DEVELOPMENTS IN BRAIN TUMOR

Radiomics accelerates the development of precision medicine. In addition to providing accurate and well-organized personal radiological diagnostic information to identify different states for each patient, large quantities of features extracted from numerous pathologically confirmed patients can contribute a lot to a large-scaled database for tumor classification. The Picture Archiving and Communication System (PACS) has enabled the acquisition, display, processing, storage, transmission, and management of medical images to be digitized and networked in a uniform standard (130, 131), with parallel progress in Europe (132) and developing countries (133, 134). Additionally, there are strong public repositories which record these systematic electronic radiological data with open access, for instance, The Cancer Genome Atlas (TCGA) (135), The Cancer Imaging Archive (TCIA) (136), and The Quantitative Imaging Network (QIN) (137).

As the database improves, it can be used for deep learning to evolve over time to build more sophisticated classifiers and may help discover more internal connections between image features and gene expression. Open availability of source code and data is encouraged for current radiomics studies to promote technical development. The development of fully automated approaches based on deep learning will start from solving the most common

clinical problems with plenty of data (138). These clinical problems may concern occasions where professional neuroradiologists are in heavy demand or analyzation is too cumbersome for neuroradiologists, like predicting *IDH* mutation status in gliomas (35).

The development of radiomics also compels the development of histopathology. Sampling is a crucial step in identifying the tumor, but it relies on the location of the lesion and is often affected by operator's subjective factors and intra-heterogeneity and inter-heterogeneity of tumor. If radiomics is performed before sampling, it can segment the lesion and suggest the most interesting area so that we can puncture the target tissue accurately. Multicenter radiomic research requires establishing norms for the radiomics study protocols and for their reporting in the literature, which also supplements traditional imaging reports with quantitative indicators and more standard structures.

Based on multiple noninvasive biomarkers, more explicit characteristics of tumor can be assessed, and the progression of tumor can be recorded and visualized (139). For example, liquid biopsy enables the analysis of molecules or macrostructures in low concentration from body liquid that shows minimal invasiveness towards patients who are susceptible to tumor or cannot withstand biopsy. Cucchiara et al. integrated liquid biopsy and radiomics to monitor clonal heterogeneity of *EGFR*-Positive NSCLC (140). As a result, more individualized treatment plans can be tailored and patients with imperceptible premalignant lesions or who undergo surgery can also benefit, though the expense is another problem to be discussed. Future studies should focus on improving the sensitivity and specificity.

CONCLUSION

Radiomics was born from traditional radiology, bioinformatics, and machine learning and provides clinicians with economical, automatic, and accurate diagnosis on brain tumors by mining high-dimensional data correlated with lesions extracted from images. The overall imaging and evaluating by radiomics not only presents the inner heterogeneity of the lesion but also indicates the microenvironment surrounding the tumor region, making it possible to guide targeted agents before experiment or to be aligned with biopsy to maximize the clinical implications. Though many guidelines are published or being developed, there are still gaps in standard radiomics reporting. As more sophisticated segmentation and analyzation techniques are exploited, along with big data to reach multicenter interoperability, we believe radiomics will soon expand rapidly beyond a small research area and transform into a clinical surrogate diagnosis tool.

AUTHOR CONTRIBUTIONS

ZY and YZ reviewed the literature and wrote the manuscript. LL reviewed the literature and generated **Table 2**. YZ and ZL designed the research. All authors contributed to the article and approved the submitted version.

FUNDING

The research was funded by China Postdoctoral Science Foundation 2018M643006.

ACKNOWLEDGMENTS

We are grateful to all of those with whom we have had the pleasure to work with during this and other related projects.

REFERENCES

1. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer Statistics, 2021. *CA: Cancer J Clin* (2021) 71(1):7–33. doi: 10.3322/caac.21654
2. 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: AJNR Am Soc Neuroradiol (ASNR)* (2018) 39(2):208–16. doi: 10.3174/ajnr.a5391
3. 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 (Oxford Engl 1990)* (2012) 48(4):441–6. doi: 10.1016/j.ejca.2011.11.036
4. Chaddad A, Kucharczyk MJ, Daniel P, Sabri S, Jean-Claude BJ, Niazi T, et al. Radiomics in Glioblastoma: Current Status and Challenges Facing Clinical Implementation. *Front Oncol* (2019) 9:374. doi: 10.3389/fonc.2019.00374
5. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More Than Pictures, They Are Data. *Radiology* (2016) 278(2):563–77. doi: 10.1148/radiol.2015151169
6. Castellino RA. Computer Aided Detection (CAD): An Overview. *Cancer Imaging: Off Publ Int Cancer Imaging Soc e-MED* (2005) 5(1):17–9. doi: 10.1102/1470-7330.2005.0018
7. Nishikawa RM. Computer-Aided Detection and Diagnosis. In: *Digital Mammography*. Berlin Heidelberg: Springer (2010). p. 85–106. doi: 10.1007/978-3-540-78450-0_6
8. Jaffe C. Measures of Response: RECIST, WHO, and New Alternatives. *J Of Clin Oncol Amer Soc Clin Oncol* (2006) 24(20):3245–51. doi: 10.1200/JCO.2006.06.5599
9. Schucht P, Knittel S, Slotboom J, Seidel K, Murek M, Jilch A, et al. 5-ALA Complete Resections Go Beyond MR Contrast Enhancement: Shift Corrected Volumetric Analysis of the Extent of Resection in Surgery for Glioblastoma. *Acta Neurochirurgica* (2014) 156(2):305–12; discussion 312. doi: 10.1007/s00701-013-1906-7
10. Kunz M, Thon N, Eigenbrod S, Hartmann C, Egensperger R, Herms J, et al. Hot Spots in Dynamic (18)FET-PET Delineate Malignant Tumor Parts Within Suspected WHO Grade II Gliomas. *Neuro-Oncology* (2011) 13 (3):307–16. doi: 10.1093/neuonc/noq196
11. Louis DN, Perry A, Reifenberger G, von Deimling A, Figarella-Branger D, Cavenee WK, et al. The 2016 World Health Organization Classification of Tumors of the Central Nervous System: A Summary. *Acta Neuropathol* (2016) 131(6):803–20. doi: 10.1007/s00401-016-1545-1
12. 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
13. Panth KM, Leijenaar RTH, Carvalho S, Lieuwes NG, Yaromina A, Dubois L, et al. Is There a Causal Relationship Between Genetic Changes and Radiomics-Based Image Features? An In Vivo Preclinical Experiment

With Doxycycline Inducible GADD34 Tumor Cells. *Radiother Oncol: J Eur Soc Ther Radiol Oncol* (2015) 116(3):462–6. doi: 10.1016/j.radonc.2015.06.013
14. Rosenstein BS, West CM, Bentzen SM, Alsner J, Andreassen CN, Azria D, et al. Radiogenomics: Radiobiology Enters the Era of Big Data and Team Science. *Int J Radiat Oncol Biol Phys* (2014) 89(4):709–13. doi: 10.1016/j.ijrobp.2014.03.009
15. Rutman AM, Kuo MD. Radiogenomics: Creating a Link Between Molecular Diagnostics and Diagnostic Imaging. *Eur J Radiol* (2009) 70(2):232–41. doi: 10.1016/j.ejrad.2009.01.050
16. Hu LS, Ning S, Eschbacher JM, Baxter LC, Gaw N, Ranjbar S, et al. Radiogenomics to Characterize Regional Genetic Heterogeneity in Glioblastoma. *Neuro-Oncology* (2017) 19(1):128–37. doi: 10.1093/neuonc/now135
17. Zinn PO, Singh SK, Kotrotsou A, Hassan I, Thomas G, Luedi MM, et al. A Coclinical Radiogenomic Validation Study: Conserved Magnetic Resonance Radiomic Appearance of Periostin-Expressing Glioblastoma in Patients and Xenograft Models. *Clin Cancer Res an Off J Am Assoc Cancer Res* (2018) 24(24):6288–99. doi: 10.1158/1078-0432.CCR-17-3420
18. Kickingereder P, Bonekamp D, Nowosielski M, Kratz A, Sill M, Burth S, et al. Radiogenomics of Glioblastoma: Machine Learning-Based Classification of Molecular Characteristics by Using Multiparametric and Multiregional MR Imaging Features. *Radiology* (2016) 281(3):907–18. doi: 10.1148/radiol.2016161382
19. Clauson J, Hsieh YC, Acharya S, Rademaker AW, Morrow M. Results of the Lynn Sage Second-Opinion Program for Local Therapy in Patients With Breast Carcinoma. Changes Manage Determinants Where Care Is Delivered. *Cancer* (2002) 94(4):889–94. doi: 10.1002/cncr.10318
20. Nguyen PL, Schultz D, Renshaw AA, Vollmer RT, Welch WR, Cote K, et al. The Impact of Pathology Review on Treatment Recommendations for Patients With Adenocarcinoma of the Prostate. *Urologic Oncol* (2004) 22(4):295–9. doi: 10.1016/S1078-1439(03)00236-9
21. Sirota RL. Mandatory Second Opinion Surgical Pathology at a Large Referral Hospital. *Cancer* (2000) 89(1):225–6.
22. Staradub VL, Messenger KA, Hao N, Wiley EL, Morrow M. Changes in Breast Cancer Therapy Because of Pathology Second Opinions. *Ann Surg Oncol* (2002) 9(10):982–7. doi: 10.1007/BF02574516
23. Sun R, Limkin EJ, Vakalopoulou M, Dercle L, Champiat S, Han SR, et al. A Radiomics Approach to Assess Tumour-Infiltrating CD8 Cells and Response to Anti-PD-1 or Anti-PD-L1 Immunotherapy: An Imaging Biomarker, Retrospective Multicohort Study. *Lancet Oncol* (2018) 19(9):1180–91. doi: 10.1016/S1470-2045(18)30413-3
24. Trebeschi S, Drago SG, Birkbak NJ, Kurilova I, Călin AM, Delli Pizzi A, et al. Predicting Response to Cancer Immunotherapy Using Noninvasive Radiomic Biomarkers. *Ann Oncol: Off J Eur Soc Med Oncol* (2019) 30(6):998–1004. doi: 10.1093/annonc/mdz108
25. 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: an Off J Am Assoc Cancer Res* (2016) 22(23):5765–71. doi: 10.1158/1078-0432.CCR-16-0702
26. Ugga L, Cuocolo R, Solari D, Guadagno E, D'Amico A, Somma T, et al. Prediction of High Proliferative Index in Pituitary Macroadenomas Using MRI-Based Radiomics and Machine Learning. *Neuroradiology* (2019) 61(12):1365–73. doi: 10.1007/s00234-019-02266-1
27. Niu J, Zhang S, Ma S, Diao J, Zhou W, Tian J, et al. Preoperative Prediction of Cavernous Sinus Invasion by Pituitary Adenomas Using a Radiomics Method Based on Magnetic Resonance Images. *Eur Radiol* (2019) 29(3):1625–34. doi: 10.1007/s00330-018-5725-3
28. Peng A, Dai H, Duan H, Chen Y, Huang J, Zhou L, et al. A Machine Learning Model to Precisely Immunohistochemically Classify Pituitary Adenoma Subtypes With Radiomics Based on Preoperative Magnetic Resonance Imaging. *Eur J Radiol* (2020) 125:108892. doi: 10.1016/j.ejrad.2020.108892
29. Zhang S, Song G, Zang Y, Jia J, Wang C, Li C, et al. Non-Invasive Radiomics Approach Potentially Predicts non-Functioning Pituitary Adenomas Subtypes Before Surgery. *Eur Radiol* (2018) 28(9):3692–701. doi: 10.1007/s00330-017-5180-6
30. Park YW, Eom J, Kim S, Kim H, Ahn SS, Ku CR, et al. Radiomics With Ensemble Machine Learning Predicts Dopamine Agonist Response in Patients With Prolactinoma. *J Clin Endocrinol Metab* (2021) 106(8):e3069–77. doi: 10.1210/clinem/dgab159
31. Bhayana S, Booth GL, Asa SL, Kovacs K, Ezzat S. The Implication of Somatotroph Adenoma Phenotype to Somatostatin Analog Responsiveness in Acromegaly. *J Clin Endocrinol Metab* (2005) 90(11):6290–5. doi: 10.1210/jc.2005-0998
32. Paragliola RM, Corsello SM, Salvatori R. Somatostatin Receptor Ligands in Acromegaly: Clinical Response and Factors Predicting Resistance. *Pituitary* (2017) 20(1):109–15. doi: 10.1007/s11102-016-0768-4
33. Bakhtiar Y, Hirano H, Arita K, Yunoue S, Fujio S, Tominaga A, et al. Relationship Between Cytokeratin Staining Patterns and Clinico-Pathological Features in Somatotropinomae. *Eur J Endocrinol* (2010) 163(4):531–9. doi: 10.1530/EJE-10-0586
34. Park YW, Kang Y, Ahn SS, Ku CR, Kim EH, Kim SH, et al. Radiomics Model Predicts Granulation Pattern in Growth Hormone-Secreting Pituitary Adenomas. *Pituitary* (2020) 23(6):691–700. doi: 10.1007/s11102-020-01077-5
35. Fan Y, Liu Z, Hou B, Li L, Liu X, Liu Z, et al. Development and Validation of an MRI-Based Radiomic Signature for the Preoperative Prediction of Treatment Response in Patients With Invasive Functional Pituitary Adenoma. *Eur J Radiol* (2019) 121:108647. doi: 10.1016/j.ejrad.2019.108647
36. O'Sullivan EP, Woods C, Glynn N, Behan LA, Crowley R, O'Kelly P, et al. The Natural History of Surgically Treated But Radiotherapy-Naïve Nonfunctioning Pituitary Adenomas. *Clin Endocrinol* (2009) 71(5):709–14. doi: 10.1111/j.1365-2265.2009.03583.x
37. Roelfsema F, Biermasz NR, Pereira AM. Clinical Factors Involved in the Recurrence of Pituitary Adenomas After Surgical Remission: A Structured Review and Meta-Analysis. *Pituitary* (2012) 15(1):71–83. doi: 10.1007/s11102-011-0347-7
38. Dekkers OM, Hammer S, de Keizer RJW, Roelfsema F, Schutte PJ, Smit JWA, et al. The Natural Course of non-Functioning Pituitary Macroadenomas. *Eur J Endocrinol* (2007) 156(2):217–24. doi: 10.1530/eje.1.02334
39. Ferrante E, Ferraroni M, Castrignanò T, Menicatti L, Anagni M, Reimondo G, et al. Non-Functioning Pituitary Adenoma Database: A Useful Resource to Improve the Clinical Management of Pituitary Tumors. *Eur J Endocrinol* (2006) 155(6):823–9. doi: 10.1530/eje.1.02298
40. Zhang Y, Ko C, Chen J, Chang K, Chen T, Lim S, et al. Radiomics Approach for Prediction of Recurrence in Non-Functioning Pituitary Macroadenomas. *Front Oncol* (2020) 10:590083. doi: 10.3389/fonc.2020.590083
41. Machado LF, Elias PCL, Moreira AC, Dos Santos AC, Murta Junior LO. MRI Radiomics for the Prediction of Recurrence in Patients With Clinically non-Functioning Pituitary Macroadenomas. *Comput Biol Med* (2020) 124:103966. doi: 10.1016/j.compbiomed.2020.103966
42. Ostrom QT, Patil N, Cioffi G, Waite K, 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-Oncology* (2020) 22(12 Suppl 2):iv1–96. doi: 10.1093/neuonc/noaa200
43. Ostrom QT, Gittleman H, Fulop J, Liu M, Blanda R, Kromer C, et al. CBTRUS Statistical Report: Primary Brain and Central Nervous System Tumors Diagnosed in the United States in 2008-2012. *Neuro-oncology Oxford Univ Press* (2015) 17 Suppl 4(Suppl 4):iv1–62. doi: 10.1093/neuonc/nov189
44. Stupp R, Mason WP, van den Bent MJ, Weller M, Fisher B, Taphoorn MJB, et al. Radiotherapy Plus Concomitant and Adjuvant Temozolomide for Glioblastoma. *N Engl J Med* (2005) 352(10):987–96. doi: 10.1056/NEJMoa043330
45. Gupta R, Webb-Myers R, Flanagan S, Buckland ME. Isocitrate Dehydrogenase Mutations in Diffuse Gliomas: Clinical and Aetiological Implications. *J Clin Pathol* (2011) 64(10):835–44. doi: 10.1136/jclinpath-2011-200227
46. Lu C, Hsu F, Hsieh KL, Kao YJ, Cheng S, Hsu JB, et al. Machine Learning-Based Radiomics for Molecular Subtyping of Gliomas. *Clin Cancer Res an Off J Am Assoc* (2018) 24(18):4429–36. doi: 10.1158/1078-0432.CCR-17-3445
47. Gutsche R, Scheins J, Kocher M, Bousabarah K, Fink GR, Shah NJ, et al. Evaluation of FET PET Radiomics Feature Repeatability in Glioma Patients. *Cancers* (2021) 13(4):647. doi: 10.3390/cancers13040647

48. Choi YS, Bae S, Chang JH, Kang S, Kim SH, Kim J, et al. Fully Automated Hybrid Approach to Predict the IDH Mutation Status of Gliomas via Deep Learning and Radiomics. *Neuro-Oncology* (2020) 23(2):304–13. doi: 10.1093/neuonc/noaa177
49. Hegi ME, Diserens A, Godard S, Dietrich P, Regli and S. Ostermann L, et al. Clinical Trial Substantiates the Predictive Value of O-6-Methylguanine-DNA Methyltransferase Promoter Methylation in Glioblastoma Patients Treated With Temozolomide. *Clin Cancer Res an Off J Am Assoc Cancer Res* (2004) 10(6):1871–4. doi: 10.1158/1078-0432.ccr-03-0384
50. Wei J, Yang G, Hao X, Gu D, Tan Y, Wang X, et al. A Multi-Sequence and Habitat-Based MRI Radiomics Signature for Preoperative Prediction of MGMT Promoter Methylation in Astrocytomas With Prognostic Implication. *Eur Radiol* (2019) 29(2):877–88. doi: 10.1007/s00330-018-5575-z
51. Li Y, Liu X, Qian Z, Sun Z, Xu K, Wang K, et al. Genotype Prediction of ATRX Mutation in Lower-Grade Gliomas Using an MRI Radiomics Signature. *Eur Radiol* (2018) 28(7):2960–8. doi: 10.1007/s00330-017-5267-0
52. Haubold J, Demircioglu A, Gratz M, Glas M, Wrede K, Sure U, et al. Non-Invasive Tumor Decoding and Phenotyping of Cerebral Gliomas Utilizing. *Eur J Nucl Med Mol Imaging* (2020) 47(6):1435–45. doi: 10.1007/s00259-019-04602-2
53. Su C, Jiang J, Zhang S, Shi J, Xu K, Shen N, et al. Radiomics Based on Multicontrast MRI can Precisely Differentiate Among Glioma Subtypes and Predict Tumour-Proliferative Behaviour. *Eur Radiol* (2019) 29(4):1986–96. doi: 10.1007/s00330-018-5704-8
54. Chan AK, Pang JC, Chung NY, Li KK, Poon WS, Chan DT, et al. Loss of CIC and FUBP1 Expressions are Potential Markers of Shorter Time to Recurrence in Oligodendroglial Tumors. *Modern Pathol an Off J United States Can Acad Pathol Inc* (2014) 27(3):332–42. doi: 10.1038/modpathol.2013.165
55. Gleize V, Alentorn A, Connen de Kérillis L, Labussière M, Nadaradjane AA, Mundwiller E, et al. CIC Inactivating Mutations Identify Aggressive Subset of 1p19q Codeleted Gliomas. *Ann Neurol* (2015) 78(3):355–74. doi: 10.1002/ana.24443
56. Padul V, Epari S, Moiyadi A, Shetty P, Shirsat NV. ETV/Pea3 Family Transcription Factor-Encoding Genes are Overexpressed in CIC-Mutant Oligodendrogliomas. *Genes Chromosomes Cancer* (2015) 54(12):725–33. doi: 10.1002/gcc.22283
57. Zhang L, Giuste F, Vizcarra JC, Li X, Gutman D. Radiomics Features Predict CIC Mutation Status in Lower Grade Glioma. *Front Oncol* (2020) 10:937. doi: 10.3389/fonc.2020.00937
58. Brennan CW, Verhaak RGW, McKenna A, Campos B, Noushmehr H, Salama SR, et al. The Somatic Genomic Landscape of Glioblastoma. *Cell* (2013) 155(2):462–77. doi: 10.1016/j.cell.2013.09.034
59. Binder ZA, Thorne AH, Bakas S, Wileyto EP, Bilello M, Akbari H, et al. Epidermal Growth Factor Receptor Extracellular Domain Mutations in Glioblastoma. *Cancer Cell* (2018) 34(1):163–77.e7. doi: 10.1016/j.ccell.2018.06.006
60. Gaonkar B, Davatzikos C. Analytic Estimation of Statistical Significance Maps for Support Vector Machine Based Multi-Variate Image Analysis and Classification. *NeuroImage* (2013) 78:270–83. doi: 10.1016/j.neuroimage.2013.03.066
61. Zhou W, Ke SQ, Huang Z, Flavahan W, Fang X, Paul J, et al. Periostin Secreted by Glioblastoma Stem Cells Recruits M2 Tumour-Associated Macrophages and Promotes Malignant Growth. *Nat Cell Biol* (2015) 17(2):170–82. doi: 10.1038/ncb3090
62. Malanchi I, Santamaria-Martínez A, Susanto E, Peng H, Lehr H, Delaloye J, et al. Interactions Between Cancer Stem Cells and Their Niche Govern Metastatic Colonization. *Nature* (2011) 481(7379):85–9. doi: 10.1038/nature10694
63. Park SY, Piao Y, Jeong KJ, Dong J, de Groot JF. Periostin (POSTN) Regulates Tumor Resistance to Antiangiogenic Therapy in Glioma Models. *Mol Cancer Ther* (2016) 15(9):2187–97. doi: 10.1158/1535-7163.MCT-15-0427
64. Mikheev AM, Mikheeva SA, Trister AD, Tokita MJ, Emerson SN, Parada CA, et al. Periostin is a Novel Therapeutic Target That Predicts and Regulates Glioma Malignancy. *Neuro-Oncology Oxford Univ Press* (2015) 17(3):372–82. doi: 10.1093/neuonc/nou161
65. Zinn PO, Singh SK, Kotrotsou A, Hassan I, Thomas G, Luedi MM, et al. A Coclinical Radiogenomic Validation Study: Conserved Magnetic Resonance Radiomic. *Clin Cancer Res: An Off J Am Assoc* (2018) 24(24):6288–99. doi: 10.1158/1078-0432.CCR-17-3420
66. Frattini V, Pagnotta SM, Tala, Fan JJ, Russo MV, Lee SB, et al. A Metabolic Function of FGFR3-TACC3 Gene Fusions in Cancer. *Nature* (2018) 553(7687):222–7. doi: 10.1038/nature25171
67. Di Stefano AL, Picca A, Saragoussi E, Bielle F, Ducray F, Villa C, et al. Clinical, Molecular, and Radiomic Profile of Gliomas With FGFR3-TACC3 Fusions. *Neuro-Oncology* (2020) 22(11):1614–24. doi: 10.1093/neuonc/noaa121
68. Louis DN, Giannini C, Capper D, Paulus W, Figarella-Branger D, Lopes MB, et al. cIMPACT-NOW Update 2: Diagnostic Clarifications for Diffuse Midline Glioma, H3 K27M-Mutant and Diffuse Astrocytoma/Anaplastic Astrocytoma, IDH-Mutant. *Acta Neuropathol* (2018) 135(4):639–42. doi: 10.1007/s00401-018-1826-y
69. Ryall S, Krishnatry R, Arnoldo A, Buczkowicz P, Mistry M, Siddaway R, et al. Targeted Detection of Genetic Alterations Reveal the Prognostic Impact of H3K27M and MAPK Pathway Aberrations in Paediatric Thalamic Glioma. *Acta Neuropathol Commun* (2016) 4(1):93. doi: 10.1186/s40478-016-0353-0
70. Ochs K, Ott M, Bunse T, Sahm F, Bunse L, Deumelandt K, et al. K27M-Mutant Histone-3 as a Novel Target for Glioma Immunotherapy. *Oncoimmunology* (2017) 6(7):e1328340. doi: 10.1080/2162402X.2017.1328340
71. Su X, Chen N, Sun H, Liu Y, Yang X, Wang W, et al. Automated Machine Learning Based on Radiomics Features Predicts H3 K27M Mutation in. *Neuro-Oncology* (2020) 22(3):393–401. doi: 10.1093/neuonc/noz184
72. Cai X, Sughrue ME. Glioblastoma: New Therapeutic Strategies to Address Cellular and Genomic Complexity. *Oncotarget* (2018) 9(10):9540–54. doi: 10.18632/oncotarget.23476
73. Park JE, Kim HS, Park SY, Nam SJ, Chun S, Jo Y, et al. Prediction of Core Signaling Pathway by Using Diffusion- and Perfusion-Based MRI. *Radiology* (2020) 294(2):388–97. doi: 10.1148/radiol.2019190913
74. Abounader R. Interactions between PTEN and Receptor Tyrosine Kinase Pathways and Their Implications for Glioma Therapy. *Expert Rev Anticancer Ther* (2009) 9(2):235–45. doi: 10.1586/14737140.9.2.235
75. Nakamura JL, Karlsson A, Arvold ND, Gottschalk AR, Pieper RO, Stokoe D, et al. PKB/Akt Mediates Radiosensitization by the Signaling Inhibitor LY294002 in Human Malignant Gliomas. *J Neuro-Oncology* (2005) 71(3):215–22. doi: 10.1007/s11060-004-1718-y
76. Li Y, Liang Y, Sun Z, Xu K, Fan X, Li S, et al. Radiogenomic Analysis of PTEN Mutation in Glioblastoma Using Preoperative Multi-Parametric Magnetic Resonance Imaging. *Neuroradiology* (2019) 61(11):1229–37. doi: 10.1007/s00234-019-02244-7
77. Platta CS, Khuntia D, Mehta MP, Suh JH. Current Treatment Strategies for Brain Metastasis and Complications From Therapeutic Techniques: A Review of Current Literature. *Am J Clin Oncol* (2010) 33(4):398–407. doi: 10.1097/COC.0b013e318194f744
78. Ohgaki H, Kleihues P. Epidemiology and Etiology of Gliomas. *Acta Neuropathol* (2005) 109(1):93–108. doi: 10.1007/s00401-005-0991-y
79. Weller M, van den Bent M, Hopkins K, Tonn JC, Stupp R, Falini A, et al. EANO Guideline for the Diagnosis and Treatment of Anaplastic Gliomas and Glioblastoma. *Lancet Oncol* (2014) 15(9):e395–403. doi: 10.1016/S1470-2045(14)70011-7
80. Gaudy-Marqueste C, Carron R, Delsanti C, Loundou A, Monestier S, Archier E, et al. On Demand Gamma-Knife Strategy can be Safely Combined With BRAF Inhibitors for the Treatment of Melanoma Brain Metastases. *Ann Oncol* (2014) 25(10):2086–91. doi: 10.1093/annonc/mdu266
81. Qian Z, Li Y, Wang Y, Li L, Li R, Wang K, et al. Differentiation of Glioblastoma From Solitary Brain Metastases Using Radiomic. *Cancer Lett* (2019) 451:128–35. doi: 10.1016/j.canlet.2019.02.054
82. Kang D, Park JE, Kim Y, Kim JH, Oh JY, Kim J, et al. Diffusion Radiomics as a Diagnostic Model for Atypical Manifestation of Primary. *Neuro-Oncology* (2018) 20(9):1251–61. doi: 10.1093/neuonc/noy021
83. Wu G, Chen Y, Wang Y, Yu J, Lv X, Ju X, et al. Sparse Representation-Based Radiomics for the Diagnosis of Brain Tumors. *IEEE Trans Med Imaging* (2018) 37(4):893–905. doi: 10.1109/TMI.2017.2776967

84. Elshafeey N, Kotrotsou A, Hassan A, Elshafei N, Hassan I, Ahmed S, et al. Multicenter Study Demonstrates Radiomic Features Derived From Magnetic Resonance. *Nat Commun* (2019) 10(1):3170. doi: 10.1038/s41467-019-11007-0
85. Lohmann P, Elahmadawy MA, Gutsche R, Werner J, Bauer EK, Ceccon G, et al. FET PET Radiomics for Differentiating Pseudoprogression From Early Tumor Progression in Glioma Patients Post-Chemoradiation. *Cancers* (2020) 12(12):3835. doi: 10.3390/cancers12123835
86. 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. *Neuro-Oncology* (2019) 21(3):404–14. doi: 10.1093/neuonc/noy133
87. Wang K, Qiao Z, Zhao X, Li X, Wang X, Wu T, et al. Individualized Discrimination of Tumor Recurrence From Radiation Necrosis in Glioma. *Eur J Nucl Med Mol Imaging* (2020) 47(6):1400–11. doi: 10.1007/s00259-019-04604-0
88. Chao ST, Ahluwalia MS, Barnett GH, Stevens GHJ, Murphy ES, Stockham AL, et al. Challenges With the Diagnosis and Treatment of Cerebral Radiation Necrosis. *Int J Radiat Oncol Biol Phys* (2013) 87(3):449–57. doi: 10.1016/j.ijrobp.2013.05.015
89. Xu Y, Rong X, Hu W, Huang X, Li Y, Zheng D, et al. Bevacizumab Monotherapy Reduces Radiation-Induced Brain Necrosis in Nasopharyngeal Carcinoma Patients: A Randomized Controlled Trial. *Int J Radiat Oncol Biol Phys* (2018) 101(5):1087–95. doi: 10.1016/j.ijrobp.2018.04.068
90. 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: an Off J Am Assoc Cancer Res* (2020) 26 (20):5438–47. doi: 10.1158/1078-0432.CCR-20-1264
91. Aslan K, Turco V, Blobner J, Sonner JK, Liuzzi AR, Núñez NG, et al. Heterogeneity of Response to Immune Checkpoint Blockade in Hypermutated Experimental Gliomas. *Nat Commun* (2020) 11(1):931. doi: 10.1038/s41467-020-14642-0
92. Stupp R, Hegi ME, Mason WP, van den Bent MJ, Taphoorn MJB, Janzer RC, et al. Effects of Radiotherapy With Concomitant and Adjuvant Temozolomide Versus Radiotherapy Alone on Survival in Glioblastoma in a Randomised Phase III Study: 5-Year Analysis of the EORTC-NCIC Trial. *Lancet Oncol* (2009) 10(5):459–66. doi: 10.1016/S1470-2045(09)70025-7
93. Scott JG, Berglund A, Schell MJ, Mihaylov I, Fulp WJ, Yue B, et al. A Genome-Based Model for Adjusting Radiotherapy Dose (GARD): A Retrospective, Cohort-Based Study. *Lancet Oncol* (2017) 18(2):202–11. doi: 10.1016/S1470-2045(16)30648-9
94. Pan Z, Zhang S, Wang X, Jiao Y, Qiu J. Machine Learning Based on a Multiparametric and Multiregional Radiomics Signature Predicts Radiotherapeutic Response in Patients With Glioblastoma. *Behav Neurol* (2020) 2020:1712604. doi: 10.1155/2020/1712604
95. Dastmalchian S, Kilinc O, Onyewadume L, Tippareddy C, McGivney D, Ma D, et al. Radiomic Analysis of Magnetic Resonance Fingerprinting in Adult Brain Tumors. *Eur J Nucl Med Mol Imaging* (2020) 48(3):683–93. doi: 10.1007/s00259-020-05037-w
96. Bae S, Choi YS, Ahn SS, Chang JH, Kang S, Kim EH, et al. Radiomic MRI Phenotyping of Glioblastoma: Improving Survival Prediction. *Radiology* (2018) 289(3):797–806. doi: 10.1148/radiol.2018180200
97. Papp L, Pötsch N, Grahovac M, Schmidbauer V, Woehrer A, Preusser M, et al. Glioma Survival Prediction With Combined Analysis of In Vivo (11)C-MET PET Features. *J Nucl Med Off Publication Soc Nucl Med* (2018) 59 (6):892–9. doi: 10.2967/jnumed.117.202267
98. Beig N, Singh S, Bera K, Prasanna P, Singh G, Chen J, et al. Sexually Dimorphic Radiogenomic Models Identify Distinct Imaging and Biological Pathways That Are Prognostic of Overall Survival in Glioblastoma. *Neuro-Oncology* (2021) 23(2):251–63. doi: 10.1093/neuonc/noaa231
99. Verduin M, Primakov S, Compter I, Woodruff HC, van Kuijk SMJ, Ramaekers BLT, et al. Prognostic and Predictive Value of Integrated Qualitative and Quantitative Magnetic Resonance Imaging Analysis in Glioblastoma. *Cancers* (2021). doi: 10.3390/cancers13040722
100. Kickingereder P, Neuberger U, Bonekamp D, Piechotta PL, Götz M, Wick A, et al. Radiomic Subtyping Improves Disease Stratification Beyond Key Molecular, Clinical. *Neuro-Oncology* (2018) 20(6):848–57. doi: 10.1093/neuonc/nox188
101. Wang Y, Wei W, Liu Z, Liang Y, Liu X, Li Y, et al. Predicting the Type of Tumor-Related Epilepsy in Patients With Low-Grade Gliomas: A Radiomics Study. *Front In Oncol* (2020) 10:235. doi: 10.3389/fonc.2020.00235
102. Qian Z, Li Y, Sun Z, Fan X, Xu K, Wang K, et al. Radiogenomics of Lower-Grade Gliomas: A Radiomic Signature as a Biological Surrogate for Survival Prediction. *Aging* (2018) 10(10):2884–99. doi: 10.18632/aging.101594
103. Liu X, Li Y, Qian Z, Sun Z, Xu K, Wang K, et al. A Radiomic Signature as a non-Invasive Predictor of Progression-Free Survival in Patients With Lower-Grade Gliomas. *NeuroImage. Clin* (2018) 20:1070–7. doi: 10.1016/j.nicl.2018.10.014
104. Nayak L, Lee EQ, Wen PY. Epidemiology of Brain Metastases. *Curr Oncol Rep* (2012) 14(1):48–54. doi: 10.1007/s11912-011-0203-y
105. Berghoff AS, Schur S, Füreder LM, Gatterbauer B, Dieckmann K, Widhalm G, et al. Descriptive Statistical Analysis of a Real-Life Cohort of 2419 Patients With Brain Metastases of Solid Cancers. *ESMO Open* (2016) 1(2):e000024.
106. Sperduto PW, Chao ST, Sneed PK, Luo X, Suh J, Roberge D, et al. Diagnosis-Specific Prognostic Factors, Indexes, and Treatment Outcomes for Patients With Newly Diagnosed Brain Metastases: A Multi-Institutional Analysis of 4,259 Patients. *Int J Radiat Oncol Biol Phys* (2010) 77(3):655–61. doi: 10.1016/j.ijrobp.2009.08.025
107. Barnholtz-Sloan JS, Sloan AE, Davis FG, Vigneau FD, Lai P, Sawaya RE. Incidence Proportions of Brain Metastases in Patients Diagnosed (1973 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
108. Vivanco I, Robins HI, Rohle D, Campos C, Grommes C, Nghiemphu PL, et al. Differential Sensitivity of Glioma- Versus Lung Cancer-Specific EGFR Mutations to EGFR Kinase Inhibitors. *Cancer Discov* (2012) 2(5):458–71. doi: 10.1158/2159-8290.CD-11-0284
109. Ahn S, Kwon H, Yang J, Park M, Cha Y, Suh S, et al. Contrast-Enhanced T1-Weighted Image Radiomics of Brain Metastases may Predict EGFR Mutation Status in Primary Lung Cancer. *Sci Rep* (2020) 10(1):8905. doi: 10.1038/s41598-020-65470-7
110. Park Y, An C, Lee J, Han K, Choi D, Ahn S, et al. Diffusion Tensor and Postcontrast T1-Weighted Imaging Radiomics to Differentiate the Epidermal Growth Factor Receptor Mutation Status of Brain Metastases From non-Small Cell Lung Cancer. *Neuroradiology* (2020) 63(3):343–52. doi: 10.1007/s00234-020-02529-2
111. Chen B, Jin T, Ye N, Mambetsariev I, Daniel E, Wang T, et al. Radiomic Prediction of Mutation Status Based on MR Imaging of Lung Cancer Brain Metastases. *Magn Reson Imaging* (2020) 69:49–56. doi: 10.1016/j.mri.2020.03.002
112. Balestrino R, Rudà R, Soffietti R. Brain Metastasis From Unknown Primary Tumour: Moving From Old Retrospective Studies to Clinical Trials on Targeted Agents. *Cancers* (2020) 12(11):3350. doi: 10.3390/cancers12113350
113. Kniep H, Madesta F, Schneider T, Hanning U, Schonfeld M, Schon G, et al. Radiomics of Brain MRI: Utility in Prediction of Metastatic Tumor Type. *Radiol Radiol Soc North Am* (2019) 290(2):479–87. doi: 10.1148/radiol.2018180946
114. Prasanna P, Rogers L, Lam T, Cohen M, Siddalingappa A, Wolansky L, et al. Disorder in Pixel-Level Edge Directions on T1WI Is Associated With the Degree of Radiation Necrosis in Primary and Metastatic Brain Tumors: Preliminary Findings. *Am J Of Neuroradiol Amer Soc Neuroradiol* (2019) 40 (3):412–7. doi: 10.3174/ajnr.A5958
115. Huang C, Lee C, Yang H, Lin C, Wu H, Chung W, et al. Radiomics as Prognostic Factor in Brain Metastases Treated With Gamma Knife Radiosurgery. *J Of Neuro-Oncology* (2020) 146(3):439–49. doi: 10.1007/s11060-019-03343-4
116. Peng L, Parekh V, Huang P, Lin D, 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 Of Radiat Oncol Biol Phys* (2018) 102(4):1236–43. doi: 10.1016/j.ijrobp.2018.05.041
117. Mouraviev A, Detsky J, Sahgal A, Ruschin M, Lee Y, Karam I, et al. Use of Radiomics for the Prediction of Local Control of Brain Metastases After Stereotactic Radiosurgery. *Neuro-Oncology* (2020) 22(6):797–805. doi: 10.1093/neuonc/noaa007
118. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis Or

Diagnosis (TRIPOD). *Ann Internal Med* (2015) 162(10):735–6. doi: 10.7326/L15-5093-2
119. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation, Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. In: *Lecture Notes in Computer Science*, vol. 9351. Cham: Springer (2015). doi: 10.1007/978-3-319-24574-4_28
120. Milletari F, Navab N, Ahmadi S. (2016). V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation, in: *2016 Fourth International Conference on 3D Vision (3DV)*, . pp. 565–71. doi: 10.1109/3DV.2016.79
121. Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J. *UNet++: A Nested U-Net Architecture for Medical Image Segmentation. Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support : 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction With MICCAI 2018* Vol. 11045. Granada, Spain, S (2018) p. 3–11. doi: 10.1007/978-3-030-00889-5_1
122. Chen W, Liu B, Peng S, Sun J, Qiao X. Computer-Aided Grading of Gliomas Combining Automatic Segmentation and Radiomics. *Int J Biomed Imaging* (2018) 2018:2512037. doi: 10.1155/2018/2512037
123. Fathallah-Shaykh HM, DeAtkine A, Coffee E, Khayat E, Bag AK, Han X, et al. Diagnosing Growth in Low-Grade Gliomas With and Without Longitudinal Volume Measurements: A Retrospective Observational Study. *PloS Med* (2019) 16(5):e1002810. doi: 10.1371/journal.pmed.1002810
124. You X, Cao N, Lu H, Mao M, Wang W. Denoising of MR Images With Rician Noise Using a Wider Neural Network and Noise Range Division. *Magn Reson Imaging* (2019) 64:154–9. doi: 10.1016/j.mri.2019.05.042
125. Zwanenburg A. Radiomics in Nuclear Medicine: Robustness, Reproducibility, Standardization, and How to Avoid Data Analysis Traps and Replication Crisis. *Eur J Nucl Med Mol Imaging* (2019) 46(13):2638–55. doi: 10.1007/s00259-019-04391-8
126. Mayerhoefer ME, Materka A, Langs G, Häggström I, Szczypiński P, Gibbs P, et al. Introduction to Radiomics. *J Nucl* (2020) 61(4):488–95. doi: 10.2967/jnumed.118.222893
127. Yang F, Dogan N, Stoyanova R, Ford JC. Evaluation of Radiomic Texture Feature Error Due to MRI Acquisition and Reconstruction: A Simulation Study Utilizing Ground Truth. *Physica Med* (2018) 50:26–36. doi: 10.1016/j.ejmp.2018.05.017
128. Park JE, Kim HS, Kim D, Park SY, Kim JY, Cho SJ, et al. A Systematic Review Reporting Quality of Radiomics Research in Neuro-Oncology: Toward Clinical Utility and Quality Improvement Using High-Dimensional Imaging Features. *BMC Cancer BioMed Cent* (2020) 20(1):29. doi: 10.1186/s12885-019-6504-5
129. Chang K, Balachandar N, Lam C, Yi D, Brown J, Beers A, et al. Distributed Deep Learning Networks Among Institutions for Medical Imaging. *J Am Med Inf Association: JAMIA* (2018) 25(8):945–54. doi: 10.1093/jamia/ocy017
130. Bryan S, Weatherburn G, Watkins J, Roddie M, Keen J, Muris N, et al. Radiology Report Times: Impact of Picture Archiving and Communication Systems. *AJR Am J Roentgenol* (1998) 170(5):1153–9. doi: 10.2214/ajr.170.5.9574575
131. Mansoori B, Erhard KK, Sunshine JL. Picture Archiving and Communication System (PACS) Implementation, Integration & Benefits in an Integrated Health System. *Acad Radiol* (2012) 19(2):229–35. doi: 10.1016/j.acra.2011.11.009
132. Lemke HU. PACS Developments in Europe. *Computerized Med Imaging Graphics: Off J Computerized Med Imaging Soc* (2003) 27(2-3):111–20. doi: 10.1016/s0895-6111(02)00084-8
133. Sarbaz M, Hosseini N, Kimiafar K. Views of Users Towards the Quality of Picture Archiving and Communication System (PACS) in a Developing Country. *Stud Health Technol Inf* (2019) 258:100–4.
134. Elahi A, Dako F, Zember J, Ojetayo B, Gerus DA, Schweitzer A, et al. Overcoming Challenges for Successful PACS Installation in Low-Resource Regions: Our Experience in Nigeria. *J Digital Imaging* (2020) 33(4):996–1001. doi: 10.1007/s10278-020-00352-y
135. Cancer Genome Atlas Research Network. Comprehensive Genomic Characterization Defines Human Glioblastoma Genes and Core Pathways. *Nature* (2008) 455(7216):1061–8. doi: 10.1038/nature07385
136. Prior FW, Clark K, Commean P, Freymann J, Jaffe C, Kirby J, et al. TCIA: An Information Resource to Enable Open Science. in: *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference.* (2013) 2013:1282–5. doi: 10.1109/EMBC.2013.6609742
137. Kalpathy-Cramer J, Freymann JB, Kirby JS, Kinahan PE, Prior FW. Quantitative Imaging Network: Data Sharing and Competitive AlgorithmValidation Leveraging The Cancer Imaging Archive. *Trans Oncol* (2014) 7(1):147–52. doi: 10.1593/tlo.13862
138. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL. Artificial Intelligence in Radiology. *Nat Rev Cancer* (2018) 18(8):500–10. doi: 10.1038/s41568-018-0016-5
139. Yousefi B, LaRiviere MJ, Cohen EA, Buckingham TH, Yee SS, Black TA, et al. Combining Radiomic Phenotypes of Non-Small Cell Lung Cancer With Liquid Biopsy Data may Improve Prediction of Response to EGFR Inhibitors. *Sci Rep* (2021) 11(1):9984. doi: 10.1038/s41598-021-88239-y
140. Cucchiara F, Petrini I, Romei C, Crucitta S, Lucchesi M, Valleggi S, et al. Combining Liquid Biopsy and Radiomics for Personalized Treatment of Lung Cancer Patients. State Art New Perspectives. *Pharmacol Res* (2021) 169:105643. doi: 10.1016/j.phrs.2021.105643

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: 19 October 2021
doi: 10.3389/fonc.2021.737901

Knowledge-Based Planning for Robustly Optimized Intensity-Modulated Proton Therapy of Head and Neck Cancer Patients

Yihang Xu, Jonathan Cyriac, Mariluz De Ornelas, Elizabeth Bossart, Kyle Padgett, Michael Butkus, Tejan Diwanji, Stuart Samuels, Michael A. Samuels and Nesrin Dogan *

Department of Radiation Oncology, University of Miami Miller School of Medicine, Miami, FL, United States

OPEN ACCESS

Edited by:
Jose Eduardo Villarreal Barajas, Royal Devon and Exeter Hospital, United Kingdom

Reviewed by:
Hamid Mammar, Institut Curie, France
Yuanpeng Zhang, Nantong University, China

***Correspondence:**
Nesrin Dogan
ndogan@med.miami.edu

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

Received: *07 July 2021*
Accepted: *27 September 2021*
Published: *19 October 2021*

Citation:
Xu Y, Cyriac J, De Ornelas M, Bossart E, Padgett K, Butkus M, Diwanji T, Samuels S, Samuels MA and Dogan N (2021) Knowledge-Based Planning for Robustly Optimized Intensity-Modulated Proton Therapy of Head and Neck Cancer Patients. Front. Oncol. 11:737901. doi: 10.3389/fonc.2021.737901

Purpose: To assess the performance of a proton-specific knowledge-based planning (KBP) model in the creation of robustly optimized intensity-modulated proton therapy (IMPT) plans for treatment of advanced head and neck (HN) cancer patients.

Methods: Seventy-three patients diagnosed with advanced HN cancer previously treated with volumetric modulated arc therapy (VMAT) were selected and replanned with robustly optimized IMPT. A proton-specific KBP model, RapidPlanPT (RPP), was generated using 53 patients (20 unilateral cases and 33 bilateral cases). The remaining 20 patients (10 unilateral and 10 bilateral cases) were used for model validation. The model was validated by comparing the target coverage and organ at risk (OAR) sparing in the RPP-generated IMPT plans with those in the expert plans. To account for the robustness of the plan, all uncertainty scenarios were included in the analysis.

Results: All the RPP plans generated were clinically acceptable. For unilateral cases, RPP plans had higher CTV_primary V100 (1.59% ± 1.24%) but higher homogeneity index (HI) (0.7 ± 0.73) than had the expert plans. In addition, the RPP plans had better ipsilateral cochlea Dmean (–5.76 ± 6.11 Gy), with marginal to no significant difference between RPP plans and expert plans for all other OAR dosimetric indices. For the bilateral cases, the V100 for all clinical target volumes (CTVs) was higher for the RPP plans than for the expert plans, especially the CTV_primary V100 (5.08% ± 3.02%), with no significant difference in the HI. With respect to OAR sparing, RPP plans had a lower spinal cord Dmax (–5.74 ± 5.72 Gy), lower cochlea Dmean (left, –6.05 ± 4.33 Gy; right, –4.84 ± 4.66 Gy), lower left and right parotid V20Gy (left, –6.45% ± 5.32%; right, –6.92% ± 3.45%), and a lower integral dose (–0.19 ± 0.19 Gy). However, RPP plans increased the Dmax in the body outside of CTV (body-CTV) (1.2 ± 1.43 Gy), indicating a slightly higher hotspot produced by the RPP plans.

Conclusion: IMPT plans generated by a broad-scope RPP model have a quality that is, at minimum, comparable with, and at times superior to, that of the expert plans. The RPP plans demonstrated a greater robustness for CTV coverage and better sparing for several OARs.

Keywords: knowledge-based planning, intensity-modulated proton therapy (IMPT), robust optimization, advanced head and neck cancer, plan quality validation

INTRODUCTION

Head and neck (HN) cancer therapy is both challenging and complicated due to the proximity of clinical target volumes (CTVs) to various critical organs such as the oral cavity, pharynx, larynx, parotids, spinal cord, and brainstem. Radiation therapy for HN cancer is an often used treatment paradigm as an adjuvant to surgery or chemotherapy. Intensity-modulated radiation therapy (IMRT), volumetric modulated arc therapy (VMAT), and intensity-modulated proton therapy (IMPT), all of which can deliver a highly conformal dose to the tumor while sparing organs at risk (OARs), are advanced radiation therapy techniques commonly used for treatment of HN cancer. Both VMAT and IMRT utilize photons to irradiate the patients, while the IMPT utilizes protons. The physical property of proton beams that can eliminate "exit dose" beyond the Bragg peak allows for steeper dose gradients and better OAR sparing than the photon-based therapy. It is well documented that IMPT offers a superior dose distribution as well as reduced toxicity as compared with IMRT and VMAT in the treatment of HN cancers (1, 2). Like IMRT, IMPT utilizes inverse planning optimization to achieve dosimetric objectives. However, the complexity of IMPT planning makes the quality of the IMPT plans very dependent on planner experience and skill, especially for plans in complex anatomy such as the HN region. This may lead to larger variations in plan quality and suboptimal dose distributions (3–5).

Knowledge-based planning (KBP) tools, which incorporate prior treatment planning experience, have the potential to improve the quality and consistency of treatment plans (6–10). One of the commercially available KBP systems [RapidPlan™ (RP) Varian Medical Systems, Palo Alto, CA] employs a dose–volume histogram (DVH) estimation model trained from a library of high-quality treatment plans. It was demonstrated by numerous studies that RP is able to generate IMRT and VMAT plans comparable with or better than the expert plans for a range of treatment sites (11–16). Recently, a proton-specific KBP system [RapidPlanPT™ (RPP), Varian Medical Systems, Palo Alto, CA] was developed to accommodate the physical traits of protons (e.g., no dose beyond the Bragg peak) into the DVH estimation model (17). A small number of publications have explored the usefulness of the RPP for HN cancer. Delaney et al. originally described the principle of RPP and demonstrated the feasibility of generating clinically acceptable planning target volume (PTV)-based IMPT plans by RPP for HN patients (17, 18). In their studies, a relatively narrow scope model was trained and evaluated, where IMPT plans with the same dose prescription and standardize field setup were applied. We believe that more studies are necessary to validate the RPP model reliability before it can be put into clinic use at this early stage. In the present work, we built an RPP model with a wide variety of HN proton plans (e.g., customized field setup, different prescriptions, and both unilateral and bilateral cases). This is a more "broad-scope" model than previously done, and we assessed its performance in the creation of robustly optimized IMPT plans for the HN cancer patients with different dose prescriptions and tumor localization.

MATERIALS AND METHODS

Patient Cohort and Intensity-Modulated Proton Therapy Planning

Seventy-three patients with advanced HN cancer located in the mid/lower HN region, including base of the tongue, tonsil, oropharynx, hypopharynx, parotid, and larynx, were included in this study. These patients were previously treated with VMAT using simultaneous integrated boost (SIB) technique and were enrolled in a retrospective institutional review board (IRB) approved protocol. Thirty of the patients underwent unilateral HN treatment, and the remaining were treated with bilateral HN irradiation. For all patients, contrast and non-contrast planning CTs were acquired in a supine position with 1.5-mm slice thickness using the Siemens Somatom 16 slice CT simulator. All gross tumor volumes (GTVs), CTVs, and OARs, including the spinal cord, brainstem, parotids, constrictors, mandible, cochlea, larynx, carotids, and oral cavity, were delineated on the contrast CT, and these volumes were subsequently transferred to the non-contrast CT. For bilateral treatment, patients were treated with three dose levels: the primary CTV prescribed to 70 Gy; the secondary CTV prescribed to 66, 63, or 60 Gy; and the tertiary CTV prescribed to 56 Gy. For unilateral cases, either one or two dose levels were prescribed with some combination of doses at the levels of 66, 60, 55, 54, and 50 Gy.

For each patient, IMPT plans were generated using multifield optimization (MFO) technique. The IMPT plans employed two to four fields depending on the target extent and anatomy. The field number and arrangement were selected by the expert planners based mainly on the tumor anatomy and location. For each field, a field-specific target was created encompassing all CTVs. These field-specific targets were then modified to avoid having beams entering through the chin area or going through teeth. Streaking artifacts caused by dental implants were delineated and overridden to an appropriate density value. The non-linear universal proton optimizer (NUPO 15.6, Eclipse,

Varian Medical Systems) was utilized for optimization along with the proton convolution superposition algorithm (PCS 15.6, Eclipse, Varian Medical Systems) for dose calculation. A relative biological effectiveness (RBE) of 1.1 was used to weight the dose. The spot spacing was set to 0.425 times the energy-dependent in-air full width at half maximum (FWHM) spot size at the isocenter. All IMPT plans were robustly optimized using ±3 mm setup uncertainty (in cardinal directions) along with ±3% proton range uncertainty, resulting in 12 uncertainty scenarios. The targets were the only structures selected to be robustly optimized. The worst-case scenario was required to achieve V95 > 95% (95% of the volume receiving more than 95% of the prescription dose) for the CTVs while keeping the normal tissue constraints as low as possible. The dose constraints used for the OARs are shown in **Table 1**. All plans were normalized such that 95% of the primary CTV volume was covered by the 100% of the prescription dose (V100 = 95%). All proton plans were created by an experienced proton dosimetrist and reviewed by a medical physicist.

Knowledge-Based Planning Model Configuration

The proton-specific KBP optimization tool RPP (Eclipse TPS, ver. 16.1, Varian Medical Systems) was used to create the KBP library. RPP consists of two phases for model configuration: the data extraction phase and the model training phase. In the data extraction phase, the geometric and dosimetric features of selected structures are parameterized for use in model training. During the model training phase, the DVH estimation algorithm is applied to create a DVH estimation model. Individual structure objectives and priorities may be set or generated based on the training set and their principal components. As described in Delaney et al., RPP incorporates a simplified spread-out Bragg peak into the model and utilizes the geometry-based expected dose (GED) metric to estimate the distance of the different voxels in each structure from the target surfaces. Delaney et al. have described RPP modelling in greater detail as well as the differences between the photon-based model and the proton-based model in their work (17), so these details will not be included here.

In our study, 53 IMPT plans consisting of 20 unilateral cases and 33 bilateral HN cases were included in the proton RPP model library. A defined objective list was implemented in the model after initial model training as shown in **Table 2**. The model quality was assessed using model generated plots such as DVH plots, regression and residual plots based on principal component analysis (PCA), and some additional metrics (19). Coefficient of determination (R^2) and average chi-square (χ^2) were applied to measure the goodness of fit of the model for each trained OAR, where the R^2 indicates the correlation between dosimetric and geometric features, while χ^2 represents the difference between the original and estimated data (19).

Model Validation

The 20 (10 unilateral cases and 10 bilateral cases) patients who were not included in the model training served as the model-validation group. For each patient used in model validation, RPP plans were created using the same beam arrangement as the corresponding expert plans. Optimization was first performed using an autogenerated objective list by the RPP. One to two additional optimization iterations were performed to improve the CTV coverage or OAR sparing with small changes to the original objective list for some patients if aforementioned dose constraints were not met. The RPP plans were normalized applying the same normalization as the expert plan (V100 = 95%).

The RPP plans were assessed and compared with the expert plans using the same clinical dose–volume constraints for CTVs and OARs. Additionally, we assessed the integral dose deposited in the structure, which removed the CTV volume from the external volume contour (body-CTV). The homogeneity index (HI) was also evaluated for RPP-based IMPT plans and compared with that of the expert plans. In this work, the HI was defined as (20, 21)

$$HI = \frac{D_{2\%} - D_{98\%}}{D_p} \times 100$$

where $D_{2\%}$ is the dose to 2% of the CTV, $D_{98\%}$ is the dose to 98% of the CTV, and D_p is the prescription dose for the CTV. The closer the HI value is to zero, the more homogenous the plan is. In order to take the plan robustness into consideration, averaged dosimetric indices over all scenarios (12 uncertainty scenarios plus the nominal scenario) were calculated for each patient, and comparisons was carried out between expert and RPP plans. All comparisons were performed by two-sided Wilcoxon signed-rank test. A p-value <0.05 was considered statistically significant.

RESULTS

Model Training Results

Table 3 reports the training results for the model. The R^2 was low for some structures such as the brainstem, larynx, and spinal cord, but the proximity of χ^2 values (mean ± SD, 1.08 ± 0.02) to 1 indicates that the model is of good quality. **Figure 1** shows the residual plots for some structures. The residual plots show how the original DVH of a structure differs from the estimated DVH, and they were used as a more realistic evaluation of potential influential points that can significantly affect the outcome of the DVH estimation model. Though previous studies have shown that removal of outliers from a good-quality KBP model library with sufficient population often does not have a significant impact on plan quality, outlier cases such as the one marked by the arrow in the constrictor plot were evaluated to determine if the patient needed to be re-planned (12, 22). After review, we believe that they did not need to be excluded, as most outliers were due to an anatomical difference or a difference in the relative location of the object to the CTV; e.g., a large part of constrictor overlapped with the CTV for the arrowed case. Thus, we decided not to remove any of the outliers from the model.

Model Validation Results

Most IMPT plans generated by the expert planners and RPP met the clinical constraints in **Table 1**. Some constraints were not met

TABLE 1 | CTV and OAR dose constraints for the nominal IMPT plan.

	Bilateral		Unilateral
CTV		V100 > 95%	
	Dmax < 115%		Dmax < 113%
Brainstem		Dmax < 54 Gy	
Left cochlea		Dmean < 40 Gy	
Right cochlea		Dmean < 40 Gy/50 Gy	
Constrictors	Dmean < 50 Gy		Dmean < 40 Gy
Larynx	Dmean < 50 Gy		Dmean < 30 Gy
Mandible	Dmax < 75 Gy		Dmax < 60 Gy
Oral cavity	Dmean < 50 Gy		Dmean < 30 Gy
Spinal cord		Dmax < 48 Gy	
Left parotid		Dmean < 26 Gy	
		V20 Gy < 50%	
Right parotid		Dmean < 26 Gy	
		V20 Gy < 50%	

V95 represents the relative volume receiving equal or more than the 95% of prescription dose; Dmax represents the maximum dose or relative dose delivered to the structure; V20Gy and Dmean represent the relative volume of the structure receiving more than 20 Gy and mean dose to the volume, respectively.
CTV, clinical treatment volume; OAR, organ at risk; IMPT, intensity-modulated proton therapy.

TABLE 2 | Objectives implemented in the model.

Structure		Relative volume	Absolute/relative dose of specific target prescription	Priority
CTV_primary	Upper	0.0%	102.0%	Generated
	Upper	25.0%	101.0%	Generated
	Lower	100.0%	100.0%	Generated
	Lower	97.0%	100.5%	Generated
	Lower (RO)	95.0%	95.0%	Generated
CTV_secondary	Upper	5%	103%	Generated
	Lower	100%	100%	Generated
	Lower	97%	101%	Generated
	Lower (RO)	95.0%	95.0%	Generated
CTV_tertiary	Upper	5%	103%	Generated
	Lower	100%	100%	Generated
	Lower	97%	101%	Generated
	Lower (RO)	95.0%	95.0%	Generated
Brainstem	Upper	0%	35 Gy	Generated
	Upper	Generated	30% of CTV_primary prescription	Generated
	Upper	Generated	15% of CTV_primary prescription	Generated
Left cochlear	Upper	Generated	25 Gy	Generated
	Mean		5 Gy	Generated
Right cochlear	Upper	Generated	25 Gy	Generated
	Mean		5 Gy	Generated
Constrictor	Upper	Generated	25% of CTV_primary prescription	Generated
	Upper	50%	Generated	Generated
Larynx	Upper	0%	68.5 Gy	Generated
	Mean		53.5 Gy	Generated
Mandible	Upper	0%	101% of CTV_primary prescription	Generated
Oral cavity	Upper	5%	Generated	Generated
	Mean		50 Gy	Generated
Left parotid	Upper	Generated	25% of CTV_primary prescription	Generated
	Upper	50%	Generated	Generated
	Mean		25% of CTV_primary prescription	Generated
Right parotid	Upper	Generated	25% of CTV_primary prescription	Generated
	Upper	50%	Generated	Generated
	Mean		25% of CTV_primary prescription	Generated
Spinal cord	Upper	0%	30 Gy	Generated
	Upper	Generated	30% of CTV_primary prescription	Generated
	Upper	Generated	15% of CTV_primary prescription	Generated
	Upper gEUD		10 Gy	Generated
Spinal cord+3 mm	Upper	0%	35 Gy	Generated
	Upper gEUD		12 Gy	Generated
Submandibular	Mean		26 Gy	Generated

RO, for robust optimization; gEUD, generalized equivalent uniform dose.

TABLE 3 | Model training results.

	Trained number	R^2	χ^2
Brainstem	51	0.504	1.062
Left cochlea	50	0.743	1.056
Right cochlea	51	0.727	1.066
Constrictor	51	0.628	1.088
Larynx	44	0.440	1.076
Mandible	51	0.783	1.102
Oral cavity	51	0.742	1.070
Left parotid	53	0.747	1.120
Right parotid	51	0.892	1.103
Spinal cord	53	0.529	1.060
Spinal cord+3 mm	44	0.556	1.066
Submandibular	25	0.973	1.095
CTV_primary	53		
CTV_secondary	47		
CTV_tertiary	36		

for very few cases due to close proximity of some OARs to CTVs. After review on these cases, these plans were clinically acceptable. **Table 4** summarizes the comparison of dosimetric indices presented as mean ± SD between the RPP-generated plans and the expert plans for 10 unilateral (**Table 4A**) and 10 bilateral (**Table 4B**) cases in the validation group. The range of each dosimetric index was also presented in brackets as (min, max) in **Table 4**. The dosimetric indices from nominal plans as well as the averaged dosimetric indices over all scenarios are listed in **Table 4**. To take the plan robustness into consideration, we will only focus on the results of averaged dosimetric indices over all scenarios. **Figures 2A, B** show the difference of averaged dose–volume indices over all scenarios between the RPP and expert plans for unilateral cases, and **Figures 2C, D** show the differences for bilateral cases.

In unilateral cases, RPP plans achieved more robust CTV coverage with a moderately higher CTV_primary V100 (1.59% ± 1.24%), whereas the expert plans were more homogeneous with a slightly lower CTV_primary HI (0.7 ± 0.73) than the RPP plans. The Dmax of the mandible from RPP plans was marginally higher than that of the expert plans (0.62 ± 0.25 Gy), but the RPP plans had a better ipsilateral cochlea Dmean (−5.76 ± 6.11 Gy). For other OAR dose–volume indices, there was no statistically significant difference between RPP and expert plans. In the bilateral cases, the V100 for all CTVs prescribed with different dose levels was higher for the RPP plans, especially for CTV_primary V100 (5.08% ± 3.02%), indicating that RPP plans were more robust for CTV coverage. There was no statistically significant difference for the CTV_primary HI between the expert and RPP plans.

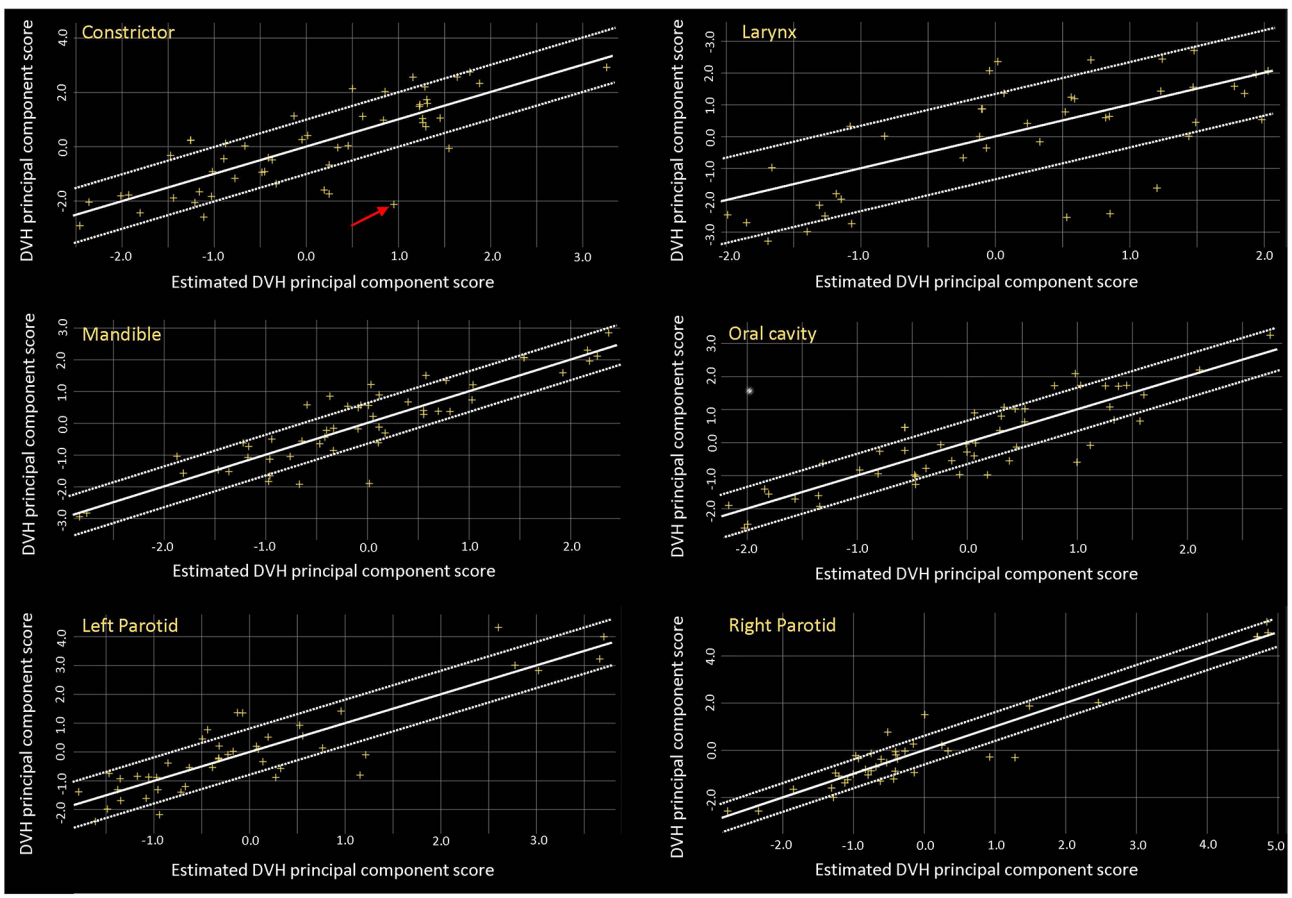


FIGURE 1 | Residual plots for some of the OARs. OARs, organs at risk.

TABLE 4A | Dosimetric comparison between RPP plans and expert plans for 10 unilateral cases in the validation group.

	Nominal plan				All scenarios averaged			
	Expert	RPP	RPP – Expert	p-Value	Expert	RPP	RPP – Expert	p-Value
CTV_primary V100 (%)	95 ± 0	95 ± 0	0 ± 0	1.000	89.69 ± 2.62 (86.18, 93.23)	91.28 ± 2.27 (86.4, 93.45)	1.59 ± 1.24	0.002
CTV_primary V95 (%)	99.74 ± 0.27 (99.13, 100)	99.43 ± 0.41 (98.65, 100)	−0.3 ± 0.29	0.012	98.99 ± 0.84 (97.03, 100)	98.83 ± 0.6 (97.87, 100)	−0.15 ± 0.53	0.426
CTV_primary HI	5.81 ± 1.31 (2.57, 7.39)	6.53 ± 1.56 (2.98, 9.04)	0.73 ± 0.48	0.002	8.01 ± 2.68 (3.16, 13.48)	8.71 ± 2.72 (3.51, 14.07)	0.7 ± 0.73	0.020
CTV_secondary V100 (%)	99.55 ± 0.73 (98.71, 99.97)	97.44 ± 2.63 (94.7, 99.95)	−2.11 ± 2.78	0.250	98.05 ± 1.85 (95.99, 99.57)	94.97 ± 3.68 (91.43, 98.78)	−3.09 ± 4.43	0.500
CTV_secondary V95 (%)	99.96 ± 0.08 (99.87, 100)	99.79 ± 0.18 (99.65, 100)	−0.16 ± 0.18	0.500	99.64 ± 0.44 (99.13, 99.93)	99.39 ± 0.48 (99.09, 99.94))	−0.25 ± 0.44	1.000
Brainstem Dmax (Gy)	9.02 ± 8.42 (1.15, 26.01)	8.01 ± 6.31 (0.97, 18.96)	−1.01 ± 3.22	0.734	9.29 ± 8.37 (1.08, 26.02)	8.41 ± 6.39 (0.92, 19.4)	−0.88 ± 3.14	0.652
Spinal cord Dmax (Gy)	14 ± 8.15 (1.35, 25.11)	14.73 ± 7.5 (0.97, 26.65)	0.73 ± 2.92	0.106	14.51 ± 8.05 (1.78, 25.34)	15.17 ± 7.34 (3.99, 26.8)	0.66 ± 2.77	0.160
Mandible Dmax (Gy)	60 ± 5.95 (50.69, 69.35)	60.8 ± 6.02 (50.92, 70.01)	0.81 ± 0.6	0.008	60.09 ± 5.78 (50.75, 68.72)	60.71 ± 5.8 (51.06, 69.26)	0.62 ± 0.25	0.004
Ipsilateral cochlea Dmean (Gy)	15.45 ± 12.25 (1.06, 39.28)	9.57 ± 6.82 (0.9, 23.01)	−5.89 ± 6.19	0.012	15.62 ± 12.23 (1.16, 39.61)	9.86 ± 6.86 (1, 23.56)	−5.76 ± 6.11	0.012
Constrictor Dmean (Gy)	12 ± 8.93 (3.74, 32.31)	12.41 ± 9.22 (3.49, 33.97)	0.41 ± 1.29	0.426	12.06 ± 8.86 (3.92, 32.25)	12.47 ± 9.13 (3.66, 33.83)	0.41 ± 1.28	0.426
Larynx Dmean (Gy)	5.21 ± 6.48 (0.01, 15.85)	5.5 ± 7.31 (0.01, 19.15)	0.29 ± 1.42	0.770	5.3 ± 6.57 (0.01, 16.11)	5.58 ± 7.37 (0.01, 19.26)	0.27 ± 1.38	0.770
Oral cavity Dmean (Gy)	4.17 ± 5.31 (0, 17.4)	4.46 ± 5.13 (0, 16.8)	0.29 ± 0.73	0.375	4.23 ± 5.31 (0, 17.41)	4.51 ± 5.13 (0, 16.82)	0.28 ± 0.73	0.375
Ipsilateral parotid Dmean (Gy)	20.1 ± 14.73 (3.32, 38.86)	20.61 ± 14.16 (4.74, 38.84)	0.51 ± 0.63	0.250	20.15 ± 14.72 (3.4, 38.87)	20.67 ± 14.14 (4.81, 38.85)	0.52 ± 0.63	0.250
Ipsilateral parotid V20Gy (%)	42.62 ± 32.94 (6.02, 85.26)	43.77 ± 31.93 (8.98, 85.59)	1.15 ± 1.22	0.125	42.67 ± 32.84 (6.12, 85.07)	43.82 ± 31.75 (9.09, 85.26)	1.14 ± 1.26	0.125
Body-CTV Dmax (Gy)	63.55 ± 5.25 (53.06, 70.57)	64.26 ± 4.62 (55.75, 70.82)	0.71 ± 1.66	0.322	64.19 ± 5.12 (54.27, 71.83)	64.53 ± 4.74 (56.44, 71.97)	0.33 ± 1.07	0.375
Body-CTV Dmean (Gy)	1.72 ± 0.93 (0.74, 3.37)	1.71 ± 0.93 (0.82, 3.36)	−0.01 ± 0.08	0.846	1.72 ± 0.93 (0.74, 3.38)	1.71 ± 0.93 (0.82, 3.37)	−0.01 ± 0.08	0.846

The data are presented as mean ± standard deviation, and the range is presented in brackets as (min, max). Only two of 10 unilateral patients contained CTV_secondary. RPP, RapidPlanPT.

The red-colored values indicate statistically significant difference (p-Value< 0.05).

With respect to OAR sparing, RPP plans had significantly superior dosimetric indices for several OARs such as the spinal cord Dmax (−5.74 ± 5.72 Gy), the left and right cochlear Dmean (left cochlea, −6.05 ± 4.33 Gy; right cochlea, −4.84 ± 4.66 Gy), the left and right parotid V20Gy (left parotid, −6.45% ± 5.32%; right parotid, −6.92% ± 3.45%), and the integral dose (−0.19 ± 0.19 Gy). However, RPP plans increased the Dmax for the body (body-CTV) (1.2 ± 1.43), revealing that the RPP plans produced a slightly higher hotspot in normal tissue than the expert plans. In this study, only patients with HN cancers in low/mid region (e.g., laryngeal to tonsillar lesions) were included. Therefore, eyes, optic nerves, and optic chiasm received relatively low doses when compared with our clinical dose constraints (Dmax < 45 Gy). For eyes, the Dmax was less than 3.5 Gy for all cases; and for optic nerves and chiasm, the Dmax was always less than 0.5 Gy except for one unilateral case where the left optic nerve had a Dmax = 7.15 Gy in the expert plan and 6.34 Gy in the RPP plan.

Figures 3A and **3B** present the mean DVH, including all scenarios over 10 unilateral patients from the validation group, and **Figures 3C, D** present the mean DVH over 10 bilateral patients. It can be observed that for unilateral cases, the RPP plans had comparable DVH with the expert plans, though the RPP plans achieved a superior DVH for the ipsilateral cochlea DVH. Concerning the DVH of bilateral cases, the RPP plans did a better job of sparing the brainstem, spinal cord, cochlea, and parotids.

Figure 4 shows the dose distributions of the RPP and expert plans for an example bilateral case from the validation group. **Figures 4A–C** indicate that although the RPP delivered a slightly higher dose to the oral cavity (oral cavity Dmean = 25.99 *vs.* 23.62 Gy), the RPP achieved better sparing for parotids, especially for right parotid (left parotid Dmean = 22.96 *vs.* 25.03 Gy, right parotid Dmean = 10.84 *vs.* 15.23 Gy). In **Figures 4D** and **E**, both the RPP and expert plans met the constraints for cochlea (Dmean < 40 Gy), but the RPP plan achieved much better sparing than the expert plan (left cochlea Dmean = 13.28 *vs.* 17.52 Gy, right cochlea Dmean = 4.7 *vs.* 18.21 Gy).

DISCUSSION

This work demonstrated that a proton-specific KBP model, RPP, can generate high-quality IMPT plans for the HN cancer patients. One of the benefits of employing RPP is its high efficiency. On average, it required about 20 min to generate the prediction, optimizations, and dose calculation when utilizing the RPP model. In comparison, it typically took more than 2 h to complete HN IMPT plans by experienced dosimetrists in our

TABLE 4B | Dosimetric comparison between RPP plans and expert plans for 10 bilateral cases in the validation group.

	Nominal plan				All scenarios averaged			
	Expert	RPP	RPP – Expert	p-Value	Expert	RPP	RPP – Expert	p-Value
CTV_primary V100 (%)	95 ± 0	95 ± 0	0 ± 0	1.000	82.03 ± 2.72(77.87, 87.7)	87.1 ± 3.06(81.65, 89.92)	5.08 ± 3.02	0.002
CTV_primary V95 (%)	99.86 ± 0.42(98.66, 100)	99.83 ± 0.42(98.63, 100)	−0.04 ± 0.04	0.008	99.52 ± 0.67(97.7, 99.91)	99.51 ± 0.51(98.23, 99.94)	−0.01 ± 0.24	0.625
CTV_primary HI	4.04 ± 1.44(3.22, 8.01)	4.39 ± 1.64(2.68, 8.71)	0.35 ± 0.47	0.027	6.63 ± 2.23(5.16, 12.77)	6.22 ± 2.11(4.55, 11.46)	−0.4 ± 0.77	0.131
CTV_secondary V100 (%)	95.15 ± 1.76(92.44, 97.44)	95.85 ± 2.28(91.03, 98.68)	0.7 ± 2.01	0.160	91.26 ± 2.91(86.67, 94.61)	93.18 ± 3.53(84.23, 96.15)	1.92 ± 2.19	0.037
CTV_secondary V95 (%)	99.82 ± 0.45(98.54, 99.99)	99.65 ± 0.44(98.49, 99.93)	−0.17 ± 0.17	0.004	99.37 ± 0.59(97.82, 99.83)	99.21 ± 0.52(97.95, 99.74)	−0.16 ± 0.27	0.160
CTV_tertiary V100 (%)	95.97 ± 1.97(92.02, 98.99)	96.8 ± 2.08(92.51, 99.45)	0.83 ± 2.69	0.846	88.83 ± 3.28(83, 94.33)	92.92 ± 3.62(85.09, 98.01)	4.09 ± 5.33	0.037
CTV_tertiary V95 (%)	99.87 ± 0.13(99.65, 99.99)	99.62 ± 0.25(99.21, 99.92)	−0.26 ± 0.21	0.002	99.34 ± 0.31(98.77, 99.75)	99.08 ± 0.43(98.55, 99.7)	−0.26 ± 0.4	0.131
Brainstem Dmax (Gy)	35.25 ± 3.89(28, 41.49)	33.08 ± 4.03(28.64, 40.75)	−2.17 ± 3.05	0.084	35.81 ± 3.79(28.65, 41.59)	33.32 ± 3.99(29.3, 41.24)	−2.48 ± 3.04	0.035
Spinal cord Dmax (Gy)	41.86 ± 3.01(35.59, 45.68)	36.13 ± 4.52(27.1, 41.49)	−5.73 ± 5.62	0.006	42.37 ± 3.13(36.07, 45.81)	36.9 ± 4.43(28.78, 43)	−5.47 ± 5.72	0.014
Mandible Dmax (Gy)	70.63 ± 3.33(61.87, 74.2)	70.8 ± 3.93(61.32, 75.02)	0.18 ± 1.16	0.865	70.43 ± 3.36(61.78, 74.24)	70.77 ± 4.01(61.16, 75.02)	0.34 ± 1.06	0.625
Left cochlea Dmean (Gy)	13.61 ± 5.17(1.87, 19.63)	7.38 ± 2.94(2.06, 13.28)	−6.23 ± 4.43	0.004	13.67 ± 5.12(2.04, 19.59)	7.62 ± 2.98(2.25, 13.67)	−6.05 ± 4.33	0.006
Right cochlea Dmean (Gy)	13.15 ± 5.84(2.95, 18.4)	8.24 ± 4.13(2.73, 15.62)	−4.91 ± 4.79	0.002	13.25 ± 5.77(3.03, 18.4)	8.41 ± 4.03(2.83, 15.54)	−4.84 ± 4.66	0.002
Constrictor Dmean (Gy)	48.72 ± 1.76(45.43, 51.34)	46.75 ± 3.57(41.33, 50.88)	−1.97 ± 2.32	0.049	48.84 ± 1.77(45.43, 51.52)	46.93 ± 3.52(41.63, 51.08)	−1.91 ± 2.27	0.049
Larynx Dmean (Gy)	44.44 ± 2.76(39.94, 48.67)	43.14 ± 3.46(37.66, 47.85)	−1.01 ± 2.34	0.297	44.69 ± 2.79(40.12, 48.96)	43.46 ± 3.47(37.89, 48.18)	−0.96 ± 2.29	0.375
Oral cavity Dmean (Gy)	20.61 ± 15.36(5.19, 49.14)	20.69 ± 15.54(4.68, 49.14)	0.07 ± 1.17	0.846	20.62 ± 15.29(5.24, 48.99)	20.69 ± 15.47(4.74, 48.98)	0.07 ± 1.16	0.922
Left parotid Dmean (Gy)	22.37 ± 1.07(20.05, 23.71)	19.35 ± 3.14(14.08, 25.03)	−3.01 ± 2.39	0.014	22.66 ± 1.16(19.99, 24.24)	19.72 ± 3.19(14.26, 25.31)	−2.94 ± 2.34	0.014
Left parotid V20Gy (%)	44.74 ± 2.04(41.86, 47.92)	38.11 ± 6.14(27.83, 46.76)	−6.63 ± 5.43	0.006	45.16 ± 2.21(42.32, 48.84)	38.71 ± 6.19(28.11, 47.13)	−6.45 ± 5.32	0.010
Right parotid Dmean (Gy)	20.88 ± 2.45(15.23, 23.09)	17.88 ± 3.16(10.84, 21.82)	−3 ± 1.46	0.002	21.17 ± 2.56(15.22, 23.65)	18.23 ± 3.24(10.95, 22.4)	−2.94 ± 1.42	0.002
Right parotid V20Gy (%)	42.62 ± 5.14(29.26, 49.01)	35.63 ± 6.71(19.24, 42.55)	−6.98 ± 3.45	0.002	43.18 ± 5.32(29.29, 49.71)	36.26 ± 6.83(19.48, 43.48)	−6.92 ± 3.45	0.002
Body-CTV Dmax (Gy)	73 ± 1.13(71.52, 75.2)	74.69 ± 1.21(73.02, 77.19)	1.69 ± 1.39	0.008	73.85 ± 1.28(72.39, 76.82)	75.04 ± 1.21(73.43, 77.23)	1.2 ± 1.43	0.049
Body-CTV Dmean (Gy)	5.59 ± 0.95(4.68, 7.25)	5.4 ± 1.04(4.45, 7.12)	−0.19 ± 0.19	0.014	5.59 ± 0.96(4.68, 7.25)	5.4 ± 1.04(4.45, 7.13)	−0.19 ± 0.19	0.014

The data are presented as mean ± standard deviation, and the range is presented in brackets as (min max).
RPP, RapidPlanPT.
The red-colored values indicate statistically significant difference (p-Value< 0.05).

study. Moreover, our results indicate that the plans generated by the RPP have greater robustness with respect to the CTV coverage when certain uncertainty parameters (3% range uncertainty and 3-mm setup uncertainty) were applied, which is consistent with the results from our previous study (23). For the unilateral cases, the RPP plans achieved comparable OAR sparing with the expert plans except for the ipsilateral cochlea where the RPP plans delivered lower dose. Regarding bilateral cases, the RPP plans improved the sparing in brainstem, spinal cord, cochlea, parotids, and constrictors without reducing the plan homogeneity when compared with the expert plans. In addition, the reduction of integral dose, which is one of the main advantages of proton therapy, was observed in the RPP plans compared with the expert plans for bilateral cases. As a tradeoff, the bilateral RPP plans produced a slightly higher hotspot than the expert plans in the body outside of CTVs (1.69 ± 1.39 Gy). In general, our results are consistent with the previous studies illustrating that the RPP plans were at least equivalent to if not better than the expert plans (18, 24, 25).

Different from the earlier studies by Delaney et al., which employed cases with the same prescription and standardized beam arrangement for model training and validation (17, 18), our model was more broad-scope in that it included cases prescribed with varying dose levels and using different customized beam arrangements. The results suggest that this broad-scope model can create IMPT plans of good quality in the HN regardless of the beam arrangement and prescription. It has been demonstrated that the

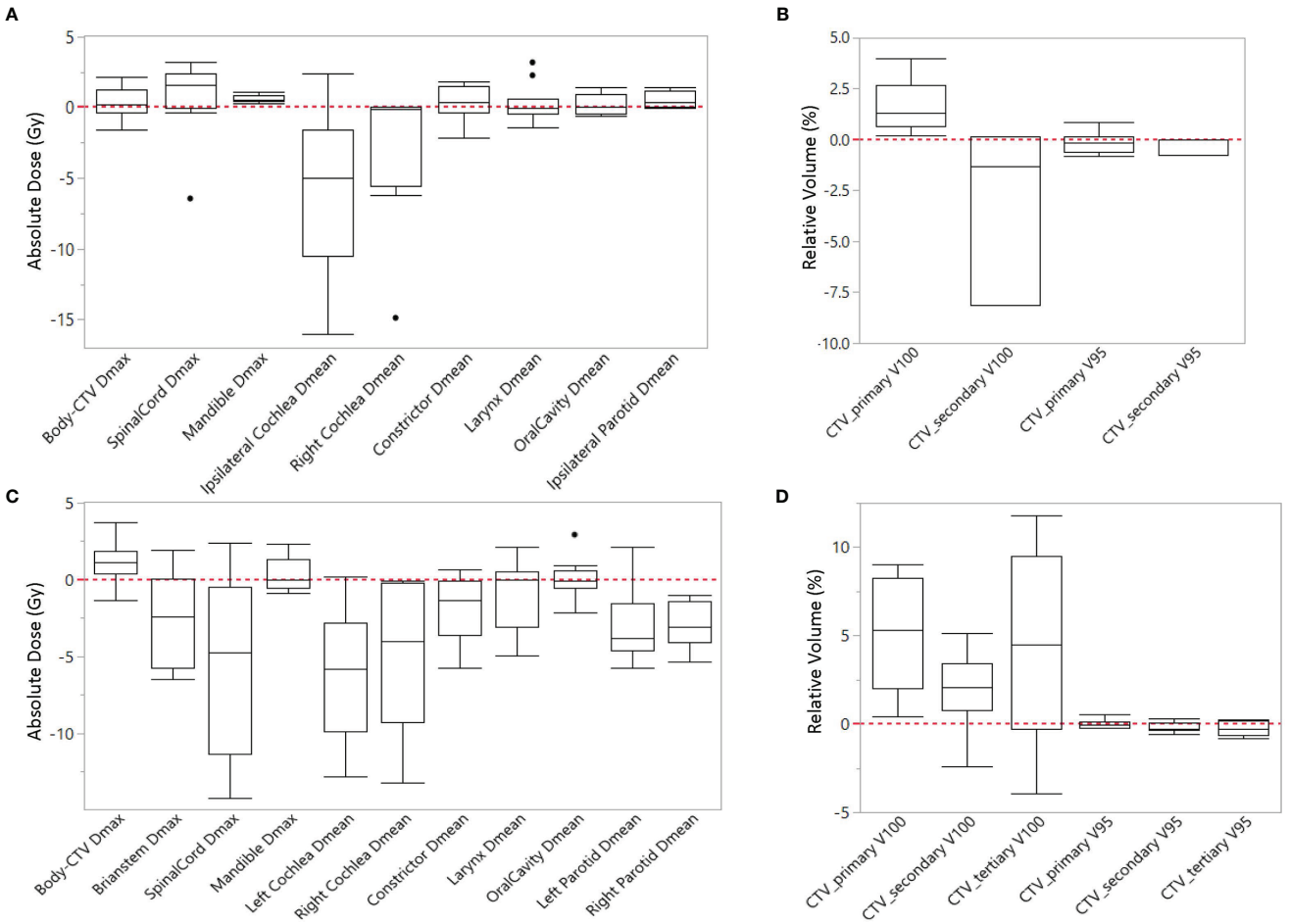


FIGURE 2 | Difference of averaged dose-volume indices over all scenarios between RPP plans and expert plans for **(A, B)** unilateral cases, and **(C, D)** bilateral cases.

quality of VMAT plans for HN cancer created by the RP model was independent of prescription and beam geometry (26). A study comparing a proton model trained with customized beam number and arrangement to another model trained with standardized beam number and arrangement for hepatocellular carcinoma treatment indicated that two models performed equivalently with no statistically significant difference for almost all dose–volume parameters (24). However, as the quality of the proton plans is more dependent on the beam arrangement than the photon plans, the impact of employing IMPT plans with different beam arrangements *versus* standardized beam arrangement in the model should be investigated for HN IMPT model. Future work on the integration of an automated beam angle selection algorithm, which is under investigation (27), should be done to see if there can be further improvement in plan quality and efficiency.

Concerning the treatment area, this study combined both unilateral and bilateral cases in the model training. In some photon KBP studies, models trained by combining unilateral and bilateral plans showed high quality in treatment of HN cancer (28). Another investigation revealed that a photon model trained by unilateral cases was able to generate high-quality VMAT plans for bilateral breast treatments (29). It is yet not clear whether a combined model or a specific model is better in generation of HN IMPT plans. In photon therapy, one study showed that specific model resulted in improved quality for liver cancer (30), while another study revealed that there was no difference of the quality between a specific model and a combined model for prostate cancer (31). Therefore, it is worthwhile to explore whether there is any benefit of utilizing specific models by separating unilateral and bilateral cases for IMPT plan generation in HN cancer treatment.

This study included 53 cases for model training, and no outlier was removed from the model. Potential outliers identified by the RPP system indicate that the plan has a statistically significant difference as compared with the whole population in the model. However, earlier studies by Delaney et al. and Hussein et al. compared the quality of the plans generated by an outlier-free model to a model without outlier removal and demonstrated that the impact of a small number of outliers does not significantly impact the plan quality (12, 22). Our previous investigation by Bossart et al. (31) also showed that the differences between refined KBP model generated by eliminating the dosimetric outliers and the original KBP generated plans were insignificant. According to the results, we believe that 53 patients should be enough to generate a

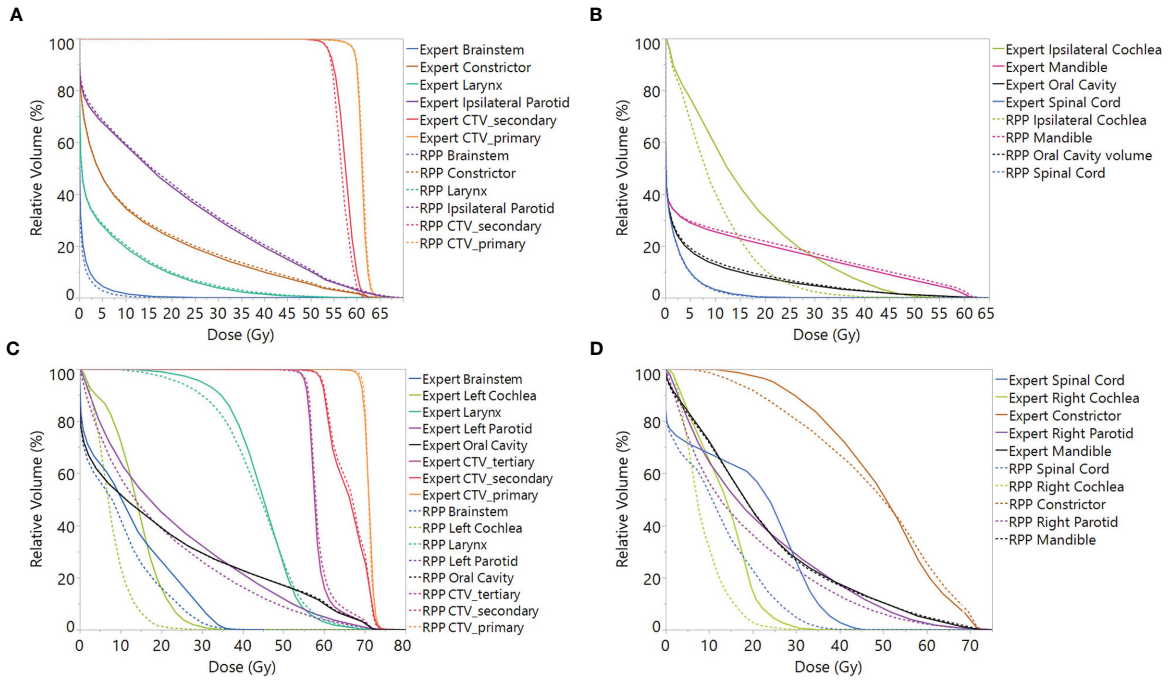


FIGURE 3 | Mean DVH including all scenarios over validation patients for **(A, B)** unilateral cases and **(C, D)** bilateral cases. Note: In order to average the CTV DVH over all patients, for each unilateral case, the DVH of CTV_primary was normalized to V60Gy = 95%, and the DVHs of CTV_secondary was normalized to V54Gy = 95%, while for bilateral cases, the CTV_secondary was normalized to V60Gy = 95%. DVH, dose–volume histogram; CTV, clinical treatment volume.

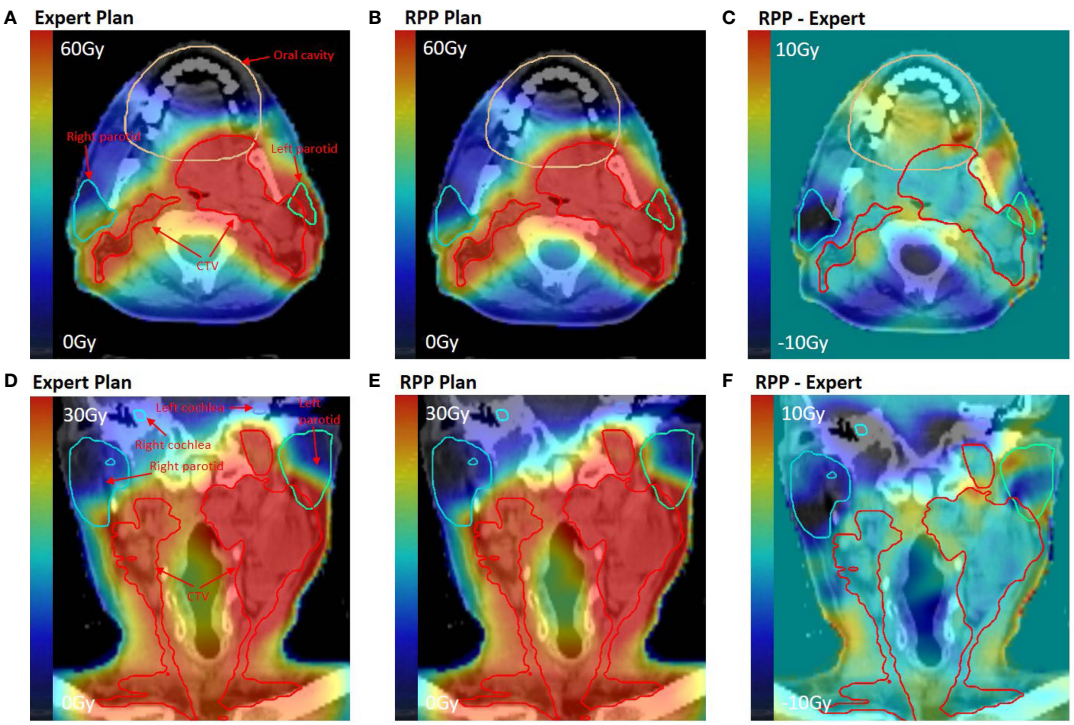


FIGURE 4 | Dose distribution for an example bilateral case in the validation group. **(A, D)** Dose distribution of the expert plan. **(B, E)** Dose distribution of the RPP plan. **(C, F)** Dose difference map between RPP and expert plan. **(A–C)** On axial plane, while **(D–F)** on coronal plane.

reliable model, but it would be necessary to investigate the influence of the model size on the IMPT plan quality.

One limitation of this study is that the 20 patients included in the validation set consisted of 10 unilateral cases and 10 bilateral cases, which may not be sufficient to confirm the reliability of the model, as it is at early stage for RPP exploration. That being said, many publications on KBP photon and proton models have included small numbers of plans for validation, and the vendor's recommendation is 10 validation cases to prove the model is working sufficiently (12, 16, 17, 24, 31, 32).

CONCLUSION

This work explored the performance of a broad-scope proton-specific KBP model to generate robustly optimized IMPT plans for HN cancer patients. The results demonstrated that the IMPT plans created by the model have high quality that is at least comparable and even, in some ways, superior to that of the expert plans. The IMPT plans generated by the model had greater robustness for CTV coverage and better sparing for several OARs. More studies should be done to evaluate the RPP model reliability.

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 University of Miami Institutional Review Board. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

ND designed the study, oversaw the whole study, and trained and validated the model library. YX carried out all data analysis and wrote the manuscript. JC, MO, and MB created IMPT expert plans for model training and validation. EB provided expertise on the proton model training and validation. KP provided his expertise in data analysis. TD provided his expertise in the study design and evaluated the IMPT plans. SS and MS provided contouring of the cases included in this study and reviewed the assessment of the cases. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported in part by a research grant from Varian Medical Systems, Palo Alto, CA (GR013242).

REFERENCES

1. Moreno AC, Frank SJ, Garden AS, Rosenthal DI, Fuller CD, Gunn GB, et al. Intensity Modulated Proton Therapy (IMPT) – The Future of IMRT for Head and Neck Cancer. *Oral Oncol* (2019) 88:66–74. doi: 10.1016/j.oraloncology.2018.11.015
2. Manzar GS, Lester SC, Routman DM, Harmsen WS, Petersen MM, Sloan JA, et al. Comparative Analysis of Acute Toxicities and Patient Reported Outcomes Between Intensity-Modulated Proton Therapy (IMPT) and Volumetric Modulated Arc Therapy (VMAT) for the Treatment of Oropharyngeal Cancer. *Radiother Oncol* (2020) 147:64–74. doi: 10.1016/j.radonc.2020.03.010
3. Das IJ, Cheng CW, Chopra KL, Mitra RK, Srivastava SP, Glatstein E. Intensity-Modulated Radiation Therapy Dose Prescription, Recording, and Delivery: Patterns of Variability Among Institutions and Treatment Planning Systems. *J Natl Cancer Institute* (2008) 100(5):300–7. doi: 10.1093/jnci/djn020
4. Nelms BE, Robinson G, Markham J, Velasco K, Boyd S, Narayan S, et al. Variation in External Beam Treatment Plan Quality: An Inter-Institutional Study of Planners and Planning Systems. *Pract Radiat Oncol* (2012) 2(4):296–305. doi: 10.1016/j.prro.2011.11.012
5. Chung HT, Lee B, Park E, Lu JJ, Xia P. Can All Centers Plan Intensity-Modulated Radiotherapy (IMRT) Effectively? An External Audit of Dosimetric Comparisons Between Three-Dimensional Conformal Radiotherapy and IMRT for Adjuvant Chemoradiation for Gastric Cancer. *Int J Radiat Oncol Biol Phys* (2008) 71(4):1167–74. doi: 10.1016/j.ijrobp.2007.11.040
6. Zhu X, Ge Y, Li T, Thongphiew D, Yin FF, Wu QJ. A Planning Quality Evaluation Tool for Prostate Adaptive IMRT Based on Machine Learning. *Med Phys* (2011) 38(2):719–26. doi: 10.1118/1.3539749
7. Moore KL, Brame RS, Low DA, Mutic S. Experience-Based Quality Control of Clinical Intensity-Modulated Radiotherapy Planning. *Int J Radiat Oncol Biol Phys* (2011) 81(2):545–51. doi: 10.1016/j.ijrobp.2010.11.030
8. Chanyavanich V, Das SK, Lee WR, Lo JY. Knowledge-Based IMRT Treatment Planning for Prostate Cancer. *Med Phys* (2011) 38(5):2515–22. doi: 10.1118/1.3574874
9. 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
10. Good D, Lo J, Lee WR, Wu QJ, Yin FF, Das SK. A Knowledge-Based Approach to Improving and Homogenizing Intensity Modulated Radiation Therapy Planning Quality Among Treatment Centers: An Example Application to Prostate Cancer Planning. *Int J Radiat Oncol Biol Phys* (2013) 87(1):176–81. doi: 10.1016/j.ijrobp.2013.03.015
11. Nwankwo O, Mekdash H, Sihono DSK, Wenz F, Glatting G. Knowledge-Based Radiation Therapy (KBRT) Treatment Planning *Versus* Planning by Experts: Validation of a KBRT Algorithm for Prostate Cancer Treatment Planning. *Radiat Oncol* (2015) 10(1):111. doi: 10.1186/s13014-015-0416-6
12. Hussein M, South CP, Barry MA, Adams EJ, Jordan TJ, Stewart AJ, et al. Clinical Validation and Benchmarking of Knowledge-Based IMRT and VMAT Treatment Planning in Pelvic Anatomy. *Radiother Oncol* (2016) 120(3):473–9. doi: 10.1016/j.radonc.2016.06.022
13. Cagni E, Botti A, Micera R, Galeandro M, Sghedoni R, Orlandi M, et al. Knowledge-Based Treatment Planning: An Inter-Technique and Inter-System Feasibility Study for Prostate Cancer. *Physica Med* (2017) 36:38–45. doi: 10.1016/j.ejmp.2017.03.002
14. Powis R, Bird A, Brennan M, Hinks S, Newman H, Reed K, et al. Clinical Implementation of a Knowledge Based Planning Tool for Prostate VMAT. *Radiat Oncol* (2017) 12(1):1–8. doi: 10.1186/s13014-017-0814-z
15. Li N, Carmona R, Sirak I, Kasaova L, Followill D, Michalski J, et al. Highly Efficient Training, Refinement, and Validation of a Knowledge-Based Planning Quality-Control System for Radiation Therapy Clinical Trials. *Int J Radiat Oncol Biol Phys* (2017) 97(1):164–72. doi: 10.1016/j.ijrobp.2016.10.005

16. Gits HC, Dow J, Matuszak M, Feng MU-S. Knowledge-Based Planning for Stereotactic Body Radiation Therapy (SBRT) of the Liver. *J Clin Oncol* (2016) 34(4_suppl):375. doi: 10.1200/jco.2016.34.4_suppl.375
17. Delaney AR, Verbakel WF, Lindberg J, Koponen TK, Slotman BJ, Dahele M. Evaluation of an Automated Proton Planning Solution. *Cureus* (2018) 10(12): e3696. doi: 10.7759/cureus.3696
18. Delaney AR, Dong L, Mascia A, Zou W, Zhang Y, Yin L, et al. Automated Knowledge-Based Intensity-Modulated Proton Planning: An International Multicenter Benchmarking Study. *Cancers* (2018) 10(11):420. doi: 10.3390/cancers10110420
19. Fogliata A, Belosi F, Clivio A, Navarria P, Nicolini G, Scorsetti M, et al. On the Pre-Clinical Validation of a Commercial Model-Based Optimisation Engine: Application to Volumetric Modulated Arc Therapy for Patients With Lung or Prostate Cancer. *Radiother Oncol* (2014) 113(3):385–91. doi: 10.1016/j.radonc.2014.11.009
20. ICRU. Prescribing, Recording, and Reporting Intensity-Modulated Photon-Beam Therapy (IMRT). ICRU Report 83. *J ICRU* (2010) 10(1):1–107. doi: 10.1093/jicru/10.1.Report83
21. Kataria T, Sharma K, Subramani V, Karrthick KP, Bisht SS. Homogeneity Index: An Objective Tool for Assessment of Conformal Radiation Treatments. *J Med Phys* (2012) 37(4):207–13. doi: 10.4103/0971-6203.103606
22. Delaney AR, Tol JP, Dahele M, Cuijpers J, Slotman BJ, Verbakel WF. Effect of Dosimetric Outliers on the Performance of a Commercial Knowledge-Based Planning Solution. *Int J Radiat Oncol Biol Phys* (2016) 94(3):469–77. doi: 10.1016/j.ijrobp.2015.11.011
23. Xu Y, Brovold N, Cyriac J, Bossart E, Padgett K, Butkus M, et al. Assessment of Knowledge-Based Planning for Prostate Intensity Modulated Proton Therapy. *Int J Particle Ther* (2021) 8(2):62–72. doi: 10.14338/IJPT-20-00088.1
24. Cozzi L, Vanderstraeten R, Fogliata A, Chang F-L, Wang P-M. The Role of a Knowledge Based Dose–Volume Histogram Predictive Model in the Optimisation of Intensity-Modulated Proton Plans for Hepatocellular Carcinoma Patients. *Strahlentherapie und Onkol* (2020) 197(4):332–42. doi: 10.1007/s00066-020-01664-2
25. Celik E, Baus W, Baues C, Schröder W, Clivio A, Fogliata A, et al. Volumetric Modulated Arc Therapy *Versus* Intensity-Modulated Proton Therapy in Neoadjuvant Irradiation of Locally Advanced Oesophageal Cancer. *Radiat Oncol* (2020) 15:1–10. doi: 10.1186/s13014-020-01570-y
26. Fogliata A, Reggiori G, Stravato A, Lobefalo F, Franzese C, Franceschini D, et al. RapidPlan Head and Neck Model: The Objectives and Possible Clinical Benefit. *Radiat Oncol* (2017) 12(1):73. doi: 10.1186/s13014-017-0808-x
27. Gu W, O'Connor D, Nguyen D, Yu VY, Ruan D, Dong L, et al. Integrated Beam Orientation and Scanning-Spot Optimization in Intensity-Modulated Proton Therapy for Brain and Unilateral Head and Neck Tumors. *Med Phys* (2018) 45(4):1338–50. doi: 10.1002/mp.12788
28. Lee SC, South CP, Adams EJ. Clinical Validation of Rapid Plan Head & Neck Model. *Oncology* (2018) 120(3):473–9. doi: 10.1016/S0167-8140(18)32201-1
29. Fogliata A, Nicolini G, Bourgier C, Clivio A, De Rose F, Fenoglietto P, et al. Performance of a Knowledge-Based Model for Optimization of Volumetric Modulated Arc Therapy Plans for Single and Bilateral Breast Irradiation. *PLoS One* (2015) 10(12):e0145137. doi: 10.1371/journal.pone.0145137
30. Yu G, Li Y, Feng Z, Tao C, Yu Z, Li B, et al. Knowledge-Based IMRT Planning for Individual Liver Cancer Patients Using a Novel Specific Model. *Radiat Oncol* (2018) 13(1):52. doi: 10.1186/s13014-018-0996-z
31. Bossart E, Duffy M, Simpson G, Abramowitz M, Pollack A, Dogan N. Assessment of Specific *Versus* Combined Purpose Knowledge Based Models in Prostate Radiotherapy. *J Appl Clin Med Phys* (2018) 19(6):209–16. doi: 10.1002/acm2.12483
32. Celik E, Baues C, Claus K, Fogliata A, Scorsetti M, Marnitz S, et al. Knowledge-Based Intensity-Modulated Proton Planning for Gastroesophageal Carcinoma. *Acta Oncol* (2020) 60(3):285–92. doi: 10.1080/0284186X.2020.1845396

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: 21 October 2021
doi: 10.3389/fonc.2021.765652

Multiparameter MRI Radiomics Model Predicts Preoperative Peritoneal Carcinomatosis in Ovarian Cancer

Xiao Yu Yu[1], Jialiang Ren[2], Yushan Jia[3], Hui Wu[3], Guangming Niu[3]*, Aishi Liu[3], Yang Gao[3], Fene Hao[3] and Lizhi Xie[3]*

[1] Affiliated Hospital, Inner Mongolia Medical University, Hohhot, China, [2] Department of Pharmaceuticals Diagnosis, GE Healthcare (China), Shanghai, China, [3] Department of Radiology, Inner Mongolia International Hospital, Hohhot, China

OPEN ACCESS

Edited by:
Jose Eduardo Villarreal Barajas, Royal Devon and Exeter Hospital, United Kingdom

Reviewed by:
Kundan Thind, Alberta Health Services, Canada
Weiwei Zong, Henry Ford Health System, United States

***Correspondence:**
Hui Wu
terrywuhui@sina.com
Guangming Niu
Cjr.niuguangming@vip.163.com

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

Received: *27 August 2021*
Accepted: *04 October 2021*
Published: *21 October 2021*

Citation:
Yu XY, Ren J, Jia Y, Wu H, Niu G, Liu A, Gao Y, Hao F and Xie L (2021) Multiparameter MRI Radiomics Model Predicts Preoperative Peritoneal Carcinomatosis in Ovarian Cancer. Front. Oncol. 11:765652. doi: 10.3389/fonc.2021.765652

Objectives: To evaluate the predictive value of radiomics features based on multiparameter magnetic resonance imaging (MP-MRI) for peritoneal carcinomatosis (PC) in patients with ovarian cancer (OC).

Methods: A total of 86 patients with epithelial OC were included in this retrospective study. All patients underwent FS-T2WI, DWI, and DCE-MRI scans, followed by total hysterectomy plus omentectomy. Quantitative imaging features were extracted from preoperative FS-T2WI, DWI, and DCE-MRI images, and feature screening was performed using a minimum redundancy maximum correlation (mRMR) and least absolute shrinkage selection operator (LASSO) methods. Four radiomics models were constructed based on three MRI sequences. Then, combined with radiomics characteristics and clinicopathological risk factors, a multi-factor Logistic regression method was used to construct a radiomics nomogram, and the performance of the radiomics nomogram was evaluated by receiver operating characteristic curve (ROC) curve, calibration curve, and decision curve analysis.

Results: The radiomics model from the MP-MRI combined sequence showed a higher area under the curve (AUC) than the model from FS-T2WI, DWI, and DCE-MRI alone (0.846 *vs*. 0.762, 0.830, 0.807, respectively). The radiomics nomogram (AUC=0.902) constructed by combining radiomics characteristics and clinicopathological risk factors showed a better diagnostic effect than the clinical model (AUC=0.858) and the radiomics model (AUC=0.846). The decision curve analysis shows that the radiomics nomogram has good clinical application value, and the calibration curve also proves that it has good stability.

Conclusion: Radiomics nomogram based on MP-MRI combined sequence showed good predictive accuracy for PC in patients with OC. This tool can be used to identify peritoneal carcinomatosis in OC patients before surgery.

Keywords: ovarian cancer, peritoneal carcinomatosis, radiomics, magnetic resonance imaging, predictions

INTRODUCTION

Ovarian cancer (OC) is the fifth most common cancer in women and the most common gynecological tumor. Epithelial ovarian cancer (EOC) is the most common OC subtype accounting for 90% of all OC. It is characterized by extensive and rapid intra-abdominal carcinomatosis and has a poor prognosis and high mortality. The 5-year survival rate of EOC is only 30% (1–4). If the patient can detect PC at an early stage, it will be able to buy sufficient treatment time for the patient and effectively control the patient's condition from further deterioration. Preoperative detection of peritoneal carcinomatosis (PC) is essential to avoid unnecessary resection and choose the best treatment method for patients with EOC.

National Comprehensive Cancer Network guidelines recommend that all peritoneal surfaces suspected of carcinomatosis should be selectively removed. Many ovarian cancers do not have ascites when they have peritoneal carcinomatosis. Because there are many peritoneal folds, smaller peritoneal metastatic nodules can be easily misdiagnosed, which affects the treatment and prognosis of patients. The diagnosis of peritoneal implants mainly relies on open exploratory surgery and laparoscopy; nonetheless, the existing results are hardly consistent. In addition, laparoscopic surgery, which is invasive and expensive, carries certain risks, such as intraoperative tumor capsule rupture, incision carcinomatosis (5, 6). Therefore, there is an urgent need for an accurate non-invasive technique to assess the PC's condition.

As an alternative method, computed tomography (CT) is usually used for preoperative examination. However, CT has limited sensitivity and may easily overlook carcinomatosis below 1 cm (7). Radiomics, an emerging and promising research field based on quantitative imaging technology, can provide decision support for oncology by extracting high-throughput quantitative radiological features from medical images (8). This low-cost and non-invasive technique has been successfully used for tumor diagnosis, staging, treatment monitoring, and treatment plan formulation (9–12). This method has been successfully applied to preoperatively predict peritoneal carcinomatosis of gastric cancer (13, 14) and may be potentially used for ovarian cancer. So far, no personalized prediction model has been developed for peritoneal carcinomatosis of ovarian cancer. This study evaluated the value of multi-parameter MRI radiomics in predicting preoperative peritoneal carcinomatosis in patients with ovarian cancer. We established a combined clinical-radiomics model to help improve decision-making and guide individualized treatment.

Abbreviations: OC, ovarian cancer; EOC, epithelial ovarian cancer; PC, peritoneal carcinomatosis; AJCC, American Joint Committee on Cancer; MP-MRI, multiparameter magnetic resonance imaging; ROI, region of interest; ICC, inter-group correlation coefficient; AUC, area under the curve; DCA, decision curve analysis.

MATERIALS AND METHODS

Patient Information

This retrospective study was approved by the ethical review committee of our hospital, and informed consent was obtained from patients.

From June 2015 to May 2020, 350 consecutive EOC patients were retrieved retrospectively in our hospital's image archiving and communication system (PACS, GE). The inclusion/exclusion criteria and patient recruitment process are shown in **Figure 1**. Inclusion criteria: 1) Histopathologically confirmed epithelial ovarian cancer; 2) Receive MRI examination one week before surgery. Exclusion criteria were: 1) past treatment history of ovarian cancer (n=69); 2) the histopathology was non-epithelial OC (n=39); 3) no dynamic enhanced MRI of the pelvis before treatment (n=34); 4) presence of clear PC signs on pelvic MRI (n=45); 5) the clinical data of CA125 were incomplete (n=31); 6) there are other distant carcinomatosis (n=46). Finally, 86 patients (age 33-82, median age 54) were enrolled in the study.

All patients underwent full hysterectomy with double appendages and increased omentectomy. The existence of PC was unanimously determined by pathologists and gynecologists according to AJCC (American Joint Committee on Cancer) guidelines. Finally, 39 out of 47 patients were detected with peritoneal carcinomatosis. The clinicopathological characteristics, including age, preoperative CA125 level, abdominal symptoms, menopausal history, genetic history, and type, were obtained from patients' medical records.

Two radiologists with 3 and 15 years of experience in female pelvic MRI imaging, who were blind to the pathological results but knew whether the patient was diagnosed with EOC, reviewed the MRI images and recorded the following: (1) unilateral or bilateral ovarian tumors; (2) tumor size (the volume of the largest layer); (3) T2 signal (low and high signals were divided by the signal strength near the myometrium); (4) enhancement degree. The radiological characteristics were selected according to the criteria of Guo HL et al. (15). Bilateral lesions were determined using the same pathological type according to the pathological surgical results; the largest tumor is finally selected for analysis. The ADC value was obtained according to the method of Thomassin et al. (16). The clinical and tumor characteristics of the patients are summarized in **Tables 1**, **2**.

Imaging Acquisition and Preprocessing

All MRI examinations were performed on a 3.0T system (Discovery MR750, GE Healthcare), using an 8-channel phased body coil. Before scanning, the patient with moderately filled bladder was placed in a supine position. Patients were also asked to fast 4-6 hours before the examination, and intestinal preparation (lactulose and magnesium sulfate) was used to reduce bowel movements. Scans were then performed along the pubic bone to the iliac spine. MR scanning parameters on the 3.0-T scanner DCE imaging of the pelvis was performed after administration of 0.1 mmol/kg of body weight of gadolinium chelate (Gadovist; Bayer). Images were acquired at postcontrast

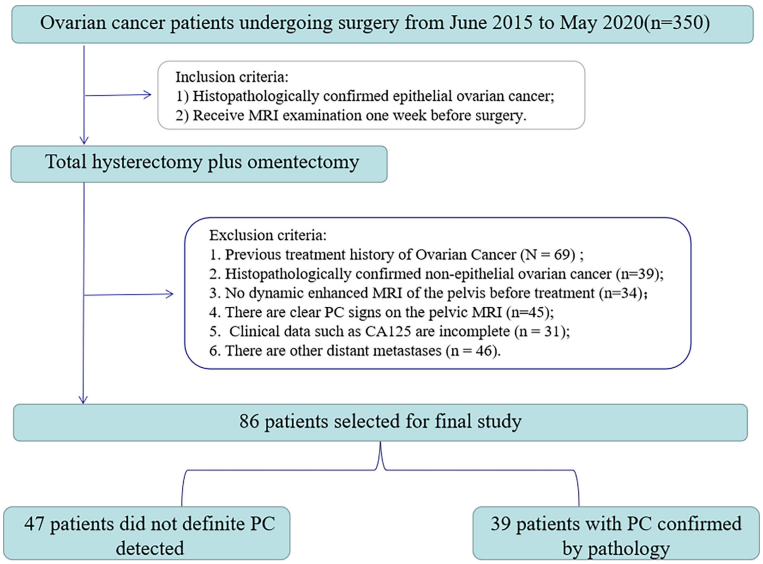


FIGURE 1 | Flowchart of patient selection.

TABLE 1 | Single-factor analysis of clinicopathological characteristics of 86 EOC patients.

Features	Without PC	With PC	p value
Number of patients	47	39	
Age (mean ± SD, years)	51.7 ± 8.8	56.8 ± 10.6	0.017
CA125 (median ± IQR, μ/ml)	213.1 (75.0-397.4)	1237. (608.7-2247.9)	<0.001
Genetic history (%)			0.950
Yes	7 (14.9%)	6 (15.4%)	
No	40 (85.1%)	33 (84.6%)	
Menopause(%)			0.381
Yes	27 (57.4%)	26 (66.7%)	
No	20 (42.6%)	13 (33.3%)	
Abdominal symptoms (%)			0.988
Yes	29 (61.7%)	24 (61.5%)	
No	18 (38.3%)	15 (38.5%)	
Type (%)			0.087
Type I	19 (40.4%)	9 (23.1%)	
Type II	28 (59.6%)	30 (76.9%)	

SD, standard deviation; IOR, interquartile range; ADC, apparent diffusion coefficient; PC, peritoneal metastasis.

TABLE 2 | Single-factor analysis of MR imaging characteristics of 86 EOC patients.

Features	Without PC	With PC	p value
Size (median ± IQR, mm3)	884.7 (239.9-1123.8)	596.4 (120.0- 844.2)	0.093
ADC (Average ± SD, mm2/s)	0.001 ± 0.001	0.001 ± 0.000	0.088
Location (%)			0.123
Unilateral	34 (72.3%)	22 (56.4%)	
Bilateral	13 (27.7%)	17 (43.6%)	
T2 homogeneity (%)			0.203
Low	14 (29.8%)	7 (17.9%)	
High	33 (70.2%)	32 (82.1%)	
T1 enhancement (%)			0.057
Mild	20 (42.6%)	9 (23.1%)	
Obvious	27 (57.4%)	30 (76.9%)	

SD, standard deviation; IOR, interquartile range; ADC, apparent diffusion coefficient; PC, peritoneal metastasis.

enhancement 120 seconds in the axial plane. This protocol obtained axial FS-T2WI, DWI, and DCE-MRI images. Detailed information about the acquisition parameters is shown in **Table S1**.

Before image segmentation, preprocessing was required. First, according to the research of Qian et al. (17), the fs-T2WI and DWI sequence images of each patient were selected for registration to the DCE-MRI (late arterial stage only) image. Then, the planar resolution of each mode was uniformly resampled to 1x1x1mm. Finally, the method of Cohen (18) was applied to normalize the image contrast of each mode to correct the factors that may affect the intensity unevenness. All the processing was performed on the 3D Slicer (version 4.10.2, funded by the National Institutes of Health) software.

MRI Radiomics Feature Extraction and Selection

Ovarian cancer lesions were performed by two radiologists with 3 years (A) and 15 years (B) experience in abdominal imaging respectively on each layer of DWI (b=1000s/mm2) to perform 3D manual manipulation of the primary tumor along the edge of the lesion segmentation. The region of interest (ROI) covered the entire tumor. The FS-T2WI and DCE-MRI were compared to avoid the cystic, necrotic, or hemorrhage area of the tumor (see **Appendix 1**).

PyRadiomics was used (19) to extract the radiologic signatures. Wavelet (8 filtering parameters) and Laplace of Gaussian (LoG, 2 filtering parameters) transformations were applied on the original image, respectively. Then, 1037 features were extracted from 11 different image types, including (1) gray histogram features; (2) morphological features; (3) gray level co-occurrence matrix (GLCM) features; (4) gray level run length matrix (GLRLM) functions; (5) grayscale area matrix (GLSZM) features. After that, all eigenvalues were normalized using Z-Score transformation. In order to ensure the reproducibility of the model results and reduce the over-fitting or selection bias in the radiomics model, the intra- and inter-group correlation coefficient (ICC) was used to evaluate the characteristics of retention stability and high repeatability, and the ICC threshold was set to 0.75. Then, the minimum redundancy maximum correlation (mRMR) was used to sort the remaining features, and each sequence retained the best top 20 features (20–22). Next, the least absolute shrinkage selection operator (LASSO) (23) method was used to screen the radiological features used to evaluate PC status. Finally, multi-factor stepwise logistic regression was used, and the Akaike Information Criterion (AIC) was used as the stopping condition to determine the best combination of radiological characteristics and clinical data (23). The workflow is shown in **Figure 2**.

Establishment of Radiomics and Clinical Models

The radiomic characteristics screened by the above method were incorporated into the multivariate Logistic regression analysis to establish a radiomics model. All the above steps were performed on the radiomics model extracted from FS-T2WI, DWI, and DCE-MRI separately and the combined model. In addition, for comparison, a Logistic regression analysis model containing clinical data was also established. Finally, the radiomics features were combined with clinical data to construct a hybrid model. In order to provide visualization and a personalized tool for predicting the probability of ovarian cancer peritoneal carcinomatosis, we have drawn a nomogram. The calibration curve was used to evaluate the calibration of the nomogram, and the Hosmer-Lemeshow test was performed. Decision curve analysis was used to calculate the net income of different models under different threshold probabilities.

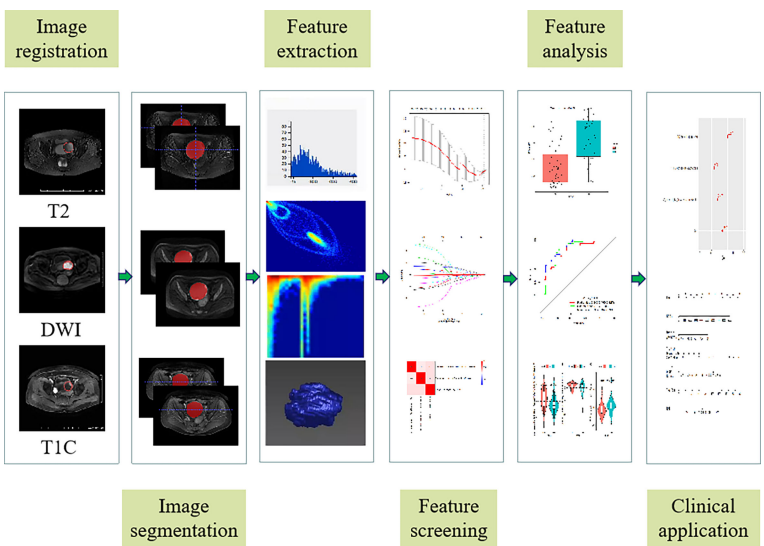


FIGURE 2 | Radiomics signature workflow.

Statistical Analysis

All statistical analyses were performed using R (Version 3.6.3, Statistical Computing Basis). Independent sample t-test or Mann-Whitney U test were used to examine the differences in measurement data uses, and the chi-square test or Fisher's exact test to evaluate the count data differences. The receiver operating characteristic curve (ROC) of the model was drawn, and the area under the curve (AUC) and 95% confidence interval (CI) were calculated to quantify the discriminant ability of the model. Delong test was used to compare the AUC among different models. The diagnostic sensitivity, specificity, accuracy, positive likelihood ratio, and negative likelihood ratio were also examined. Calibration curves were used to evaluate the predictive performance of each model. Decision curve analysis (DCA) was used to evaluate the net benefits of each model under different threshold probabilities and to evaluate the clinical applicability of each model. A two-sided p-value less than 0.05 was considered to be statistically significant.

RESULTS

Clinical and Tumor Characteristics of Patients

The clinicopathological and radiological characteristics of the patients are shown in **Tables 1** and **2**. There were significant differences in age and preoperative CA125 levels between ovarian cancer (OC) with PC and OC without PC (p=<0.001-0.017). However, no differences were found for abdominal symptoms, menopausal history, genetic history, type, lesion location, tumor size, T2 signal, and ADC between the two groups.

Evaluation of Radiomics

According to the standard of ICC>0.75, FS-T2WI, DWI, and DCE-MRI sequences retained 508, 557, and 508 radiomic features, respectively. Minimum redundancy maximum correlation and Lasso regression were then performed on the selected omics features to adjust penalty parameters through 10 cross-validation and select non-zero coefficient features related to PC status. Finally, 2 FS-T2WI features, 1 DWI feature, and three DCE-MRI features were used to build a model (**Appendix Figure 2**). Consequently, after removing features by multivariate Logistic regression, 3 features were retained, and a combined multi-sequence model was established (**Appendix Figure 3**).

The combined model from multiple sequences showed better distinguishing ability than the single model (when using a single sequence). The ROC curves of the four models are shown in **Figure 3A**. The AUC values of Fs-T2WI, DWI, CE-T1WI, and the combined model were 0.762 (0.662-0.861), 0.830 (0.745-0.914), 0.807 (0.717-0.898), and 0.846(0.765-0.927), respectively. The performance of the models is shown in **Table S2**. The DCA curves and calibration curves of the four models are shown in **Appendix Figure 4**.

Model Comparison and Nomogram Performance

Multivariate analysis of clinical data and radiomic characteristics showed that preoperative CA125 level, DWI_HLH_glszm_SizeZoneNonUniformityNormalized, T1C_glszm_LowGrayLevelZoneEmphasis, T1C_LHL_ngtdm_Contrast were significant predictors (**Figure 4**). As a result, they were fused into a nomogram (**Figure 5**). The AUC of the radiology nomogram was higher than that of the clinical model and the radiomic model (0.902, 95%CI: 0.846-0.858), indicating that the radiology nomogram can effectively distinguish the presence or absence of peritoneal carcinomatosis. The predictive performance of the clinical model was not significantly different from that of the omics model (AUC=0.858 *vs*. AUC=0.846). **Figure 3B** summarizes the diagnostic performance and ROC analysis results of these three models.

The calibration curve of the nomogram showed a good agreement between the predicted value and the observed value.

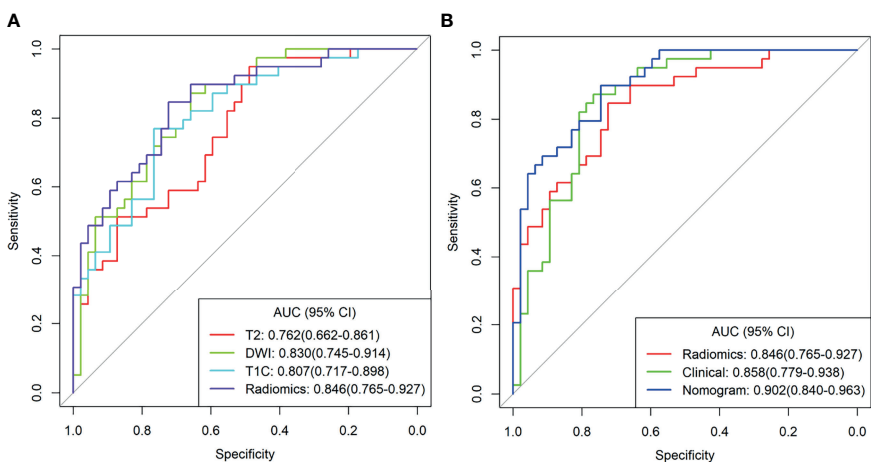


FIGURE 3 | **(A)** T2WI radiation model, DWI radiation model, T1C Radiation Model, and combined radiation model. **(B)** Clinical Model, combined radiological model, and Nomogram Receiver Operating Characteristic curve.

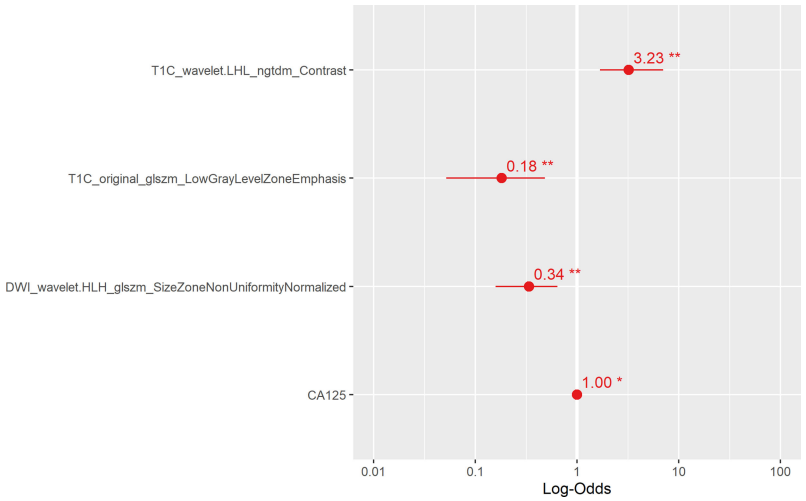


FIGURE 4 | The main radiological features extracted in this study and the results of multiple logistic regression of preoperative CA125. The horizontal line is the 95% confidence interval of the study, and the small dot in the center of the horizontal line is the point of the OR value.

The Hosmer-Lemeshow test was not significant ($p>0.05$), indicating a good degree of fit (**Figure 6A**). Decision curves were used to compare the benefits of nomograms, radiomics models, and clinical models, and we found that when the threshold probability of DCA curves was 37%-85%, nomograms had better predictive performance than clinical models and omics models (**Figure 6B**).

DISCUSSION

In this study, a nomogram radiomics model for preoperative prediction of peritoneal carcinomatosis of EOC was proposed based on clinical data and radiomics features reflecting primary tumors' characteristics. Our research shows that the multi-sequence combination model is better than the single-sequence

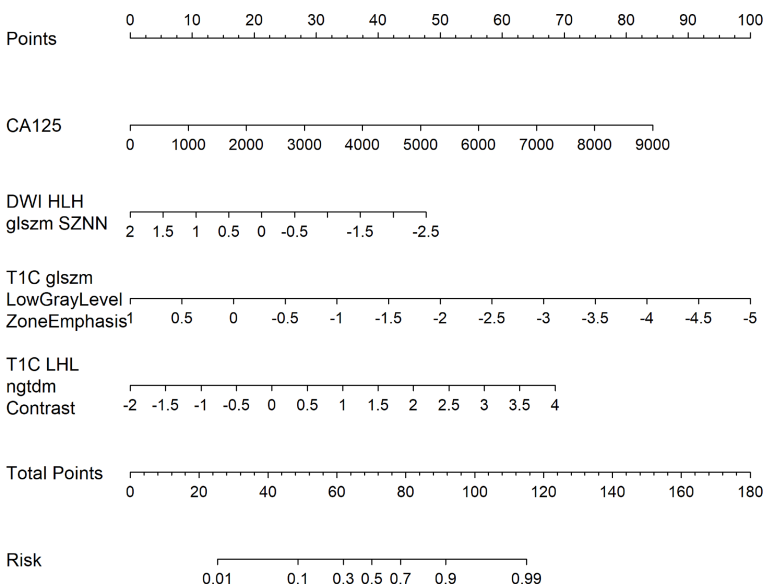


FIGURE 5 | Radiology nomogram. The radiology nomogram prediction model predicts the probability of PC in patients with epithelial ovarian cancer. The model is developed in a training group with radiomic characteristics and one clinical feature. How to use: (1) locate the patient's CA125 and then draw a straight line on the top dot axis to obtain a score related to CA125; (2) the patient's radiologic score is found on the characteristic axis of Radiology, and a line is drawn vertically up along the "point" axis. The process is repeated for each variable. (3) Sum up the sum of the four major risk factors. (4) Find the final sum on the Total Point axis and draw a straight line down to assess PC's risk in patients with epithelial ovarian cancer.

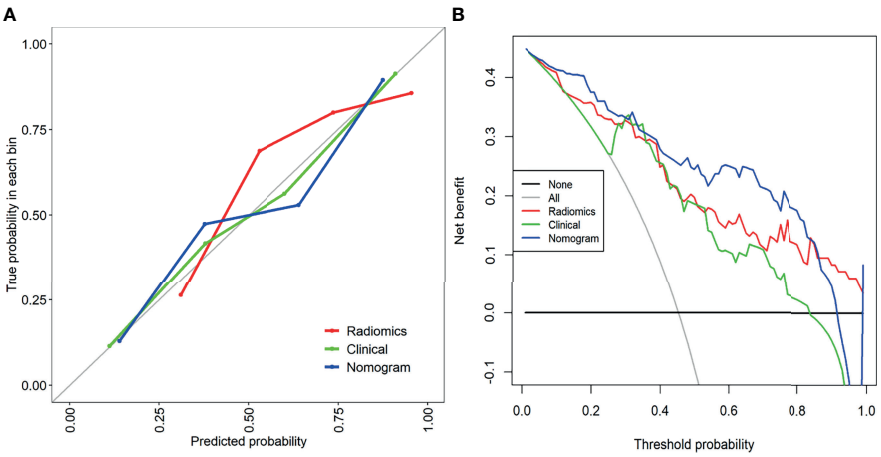


FIGURE 6 | **(A)** The calibration curve of the clinical model, combined radiology model, and nomogram. It is a curve with the model predicted PC probability as the X-axis and the actual PC probability as the Y-axis. The degree of coincidence between the calibration curve depicted and the 45-degree straight line reflects the predictive performance of each model. **(B)** Decision curve analysis of the clinical model, combined radiomics model, and mixed model. The Y-axis represents net income. The blue line represents the radiographic nomogram. Red lines represent radiomics models. The green line represents the model that contains only clinical features. The gray line represents the assumption that all patients have LN metastasis. The thin black line indicates the hypothesis that no patients have PC metastasis.

model. The constructed nomogram provides an easy-to-use, non-invasive and individualized tool for PC diagnosis and provides decision support for clinicians.

Some studies have explored the value of MRI in the evaluation of PC in epithelial ovarian cancer. At present, all published studies, including recent studies, have been conducted with comparable or fewer patient sample sizes, mainly focusing on routine imaging (15, 24, 25). Our study extracted more than 3000 features from MP-MRI images and evaluated the MR imaging features of patients with epithelial ovarian cancer in FS-T2WI, DWI, DCE-MRI, and the combination of the three. When FS-T2WI, DWI, and DCE-MRI were combined, the diagnostic efficiency of the presence or absence of peritoneal carcinomatosis was the highest, which is consistent with previous reports in the literature (26, 27). The DWI of our study is very close to the AUC of radiomics. This is because in the absence of ascites, some small lesions are usually better seen on the DWI image than on the standard T1 and T2 weighted images. To the best of our knowledge, this is the first study that combines anatomical, diffusion, and perfusion MRI and uses radiomics analysis based on the primary tumor to predict PC. Therefore, the MP-MRI combined model reveals more detailed tumor information and can more accurately predict ovarian cancer's PC status.

In our study, preoperative CA125 level was considered to be an independent predictor of PC carcinomatosis. Age has a certain potential for predicting PC, but it is not as effective as CA125. Therefore, a clinical model based on CA125 was established. Our data showed that the risk of PC with high levels of CA125 was significantly higher, which is consistent with the results reported in the previous literature (28, 29). At the same time, in order to facilitate clinical use, we have developed a nomogram containing preoperative CA125 levels and radiomic characteristics, with an AUC value of 0.902, for predicting PC. The DCA curve showed a satisfactory net income. The calibration curve showed that the stability was slightly inferior to the clinical model. This is because the clinical model was established based on a single variable, and CA125 was a continuous variable with a small fluctuation range. The nomogram integrates clinical data and radiomics characteristics and contains multi-dimensional quantitative and detailed information, so the results obtained are more objective and accurate.

Peritoneal carcinomatosis are common in the stomach, gallbladder, pancreas, lungs, intestines, uterus, and ovaries (30–32). Most of the patients with early-stage ovarian cancer PC have no specific symptoms; these patients are usually diagnosed when in an advanced stage. It is one of the main causes of ovarian cancer morbidity and mortality (32). So far, many studies have evaluated the PC status of patients with ovarian cancer (15, 24, 25). CT is a common tool used to detect PC, but its missed diagnosis rate is high. Although ^{18}F-FDG PET/CT has been shown to achieve good results in evaluating PC status, it is not widely used because of the high cost. MRI has a high resolution for soft tissues and can clearly show anatomical relationships.

Studies have recently shown that radiomics can predict peritoneal carcinomatosis in cancer patients (13, 14). Dong et al. (13) developed a personalized nomogram to identify the occult peritoneal carcinomatosis of advanced gastric cancer, achieving good results, which is similar to our findings, with few differences: first, we used MP-MRI to extract features, which can reflect tumor information more comprehensively and in more detail; second, we established a single sequence model and a combined model. Through comparison, we found that the combined model had better performance in predicting PC. In addition, the nomogram we established also showed good clinical practicability and ability to provide a diagnostic basis for predicting the PC status of ovarian cancer before clinical surgery.

This study has a few limitations. First, it is a retrospective study with small sample size. In the future, more patients are needed to provide more reliable evidence for clinical applications. Secondly, the research model's establishment and verification is a single-center, and further data from multiple centers are needed for external validation. In addition, we only performed radiomics analysis on late-arterial phase of contrast-enhanced MRI. Perhaps venous MRI images may provide more useful radiomics information, which needs to be discussed further. Finally, only primary tumors were selected in our study, and the radiohistological features of the peritoneum were not routinely used. MRI texture analysis of peritoneum needs to be further studied to explore its value.

In summary, we established nomograms based on preoperative CA125 and radiographic characteristics from primary tumors, which can be used to predict peritoneal carcinomatosis in EOC patients preoperatively. This effective and easy-to-use new approach provides a non-invasive and reliable tool for EOC patients to develop individualized treatment plans.

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 Biomedical Medical Research Ethics Committee of Inner Mongolia Medical University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

XY and JR substantial contributions to the conception or design of the work; or the acquisition, analysis or interpretation of data for the work. HW and GN drafting the work or revising it critically for important intellectual content. YJ, FH, and LX provide approval for publication of the content. YG and AL agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. All authors contributed to the article and approved the submitted version.

FUNDING

This article was funded by the Inner Mongolia Autonomous Region Fund of Natural Science (2021MS08026).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2021.765652/full#supplementary-material

REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer Statistics, 2018. *CA Cancer J Clin* (2018) 68(1):7–30. doi: 10.3322/caac.21387
2. Scalici JM, Atkins KA, Petroni G, Duska RL, Atkins AK, Saks JE, et al. Mesothelium Expression of Vascular Cell Adhesion Molecule-1 (VCAM-1) Is Associated With an Unfavorable Prognosis in Epithelial Ovarian Cancer (EOC). *Cancer* (2017) 23(6):977–84. doi: 10.1002/cncr.30415
3. Ahmed N, Stenvers Kaye L. Getting to Know Ovarian Cancer Ascites: Opportunities for Targeted Therapy-Based Translational Research. *Front Oncol* (2013) 3(3):256. doi: 10.3389/fonc.2013.00256
4. Cannistra, Stephen A. Cancer of the Ovary. *N Engl J Med* (2004) 351 (24):1550–9. doi: 10.1056/NEJMra041842
5. Smorgick N, Barel O, Halperin R, Schneider D, Pansky M. Laparoscopic Removal of Adnexal Cysts: Is it Possible to Decrease Inadvertent Intraoperative Rupture Rate? *Am J Obstet Gynecol* (2009) 200(3):237.e1–3. doi: 10.1016/j.ajog.2008.10.030
6. Zivanovic O, Sonoda Y, Diaz John P, Levine DA, Brown CL, Chi DS, et al. The Rate of Port-Site Metastases After 2251 Laparoscopic Procedures in Women With Underlying Malignant Disease. *Gynecol Oncol* (2008) 111(3):431–7. doi: 10.1016/j.ygyno.2008.08.024
7. Suidan Rudy S, Ramirez Pedro T, Sarasohn Debra M, Sarasohn, Jerrold B, Teitcher, et al. A Multicenter Prospective Trial Evaluating the Ability of Preoperative Computed Tomography Scan and Serum CA-125 to Predict Suboptimal Cytoreduction at Primary Debulking Surgery for Advanced Ovarian, Fallopian Tube, and Peritoneal Cancer. *Gynecol Oncol* (2014) 134 (3):455–61. doi: 10.1016/j.ygyno.2014.07.002
8. Gillies Robert J, Kinahan Paul E, Hricak H. Radiomics: Images Are More Than Pictures, They Are Data. *Radiology* (2016) 278(2):563–77. doi: 10.1148/radiol.2015151169
9. Drukker K, Giger ML, Joe BN, Kerlikowske K, Shepherd. Combined Benefit of Quantitative Three-Compartment Breast Image Analysis and Mammography Radiomics in the Classification of Breast Masses in a Clinical Data Set. *Radiology* (2018) 290(3):180608. doi: 10.1148/radiol.2018180608
10. Park Hyo J, Lee Seung S, Park B, Yun J, Sung YS, Shim WH, et al. Radiomics Analysis of Gadoxetic Acid-Enhanced MRI for Staging Liver Fibrosis. *Radiology* (2019) 292(1):269. doi: 10.1148/radiol.2019194012
11. Mattonen SA, Davidzon GA, Benson J, Leung ANC, Nair VS, Horng G, et al. Bone Marrow and Tumor Radiomics at 18F-FDG PET/CT: Impact on Outcome Prediction in Non–Small Cell Lung Cancer. *Radiology* (2019) 293 (2):190357. doi: 10.1148/radiol.2019190357
12. Sun R, Limkin EJ, Vakalopoulou M, Laurent D, Stéphane C, Rong HS, et al. A Radiomics Approach to Assess Tumour-Infiltrating CD8 Cells and Response to Anti-PD-1 or Anti-PD-L1 Immunotherapy: An Imaging Biomarker, Retrospective Multicohort Study. *Lancet Oncol* (2018) 19:S1470204518304133. doi: 10.1016/S1470-2045(18)30413-3
13. Dong D, Tang L, Li ZY, Fang MJ, Gao JB, Shan XH, et al. Development and Validation of an Individualized Nomogram to Identify Occult Peritoneal Metastasis in Patients With Advanced Gastric Cancer. *Ann Oncol* (2019) 30 (3):431–8. doi: 10.1093/annonc/mdz001

14. Liu S, He J, Liu S, Guan W, Chen L, Guan Y, et al. Radiomics Analysis Using Contrast-Enhanced CT for Preoperative Prediction of Occult Peritoneal Metastasis in Advanced Gastric Cancer. *Eur Radiol* (2020) 30:239–46. doi: 10.1007/s00330-019-06368-5
15. Guo HL, Ling H, Zhu YC, Wu K, Feng Y. Comparison Between Multi-Slice Spiral CT and Magnetic Resonance Imaging in the Diagnosis of Peritoneal Metastasis in Primary Ovarian Carcinoma. *Onco Targets Ther* (2018) 11:1087–94. doi: 10.2147/OTT.S147700
16. Thomassin-Naggara I, Aubert E, Rockall A, Jalaguier-Coudray A, Rouzier R, Darai E, et al. Adnexal Masses: Development and Preliminary Validation of an MR Imaging Scoring System. *Radiology* (2013) 267(2):432–43. doi: 10.1148/radiol.13121161
17. Qian LD, Ren JL, Liu AS, Gao Y, Niu GM. MR Imaging of Epithelial Ovarian Cancer: A Combined Model to Predict Histologic Subtypes. *Eur Radiol* (2020) 18(30):5815–25. doi: 10.1007/s00330-020-06993-5
18. Cohen MS, Dubois RM, Zeineh MM. Rapid and Effective Correction of RF Inhomogeneity for High Field Magnetic Resonance Imaging. *Hum Brain Mapp* (2015) 10(4):204–11. doi: 10.1002/1097-0193(200008)10:4<204::AID-HBM60>3.0.CO;2-2
19. Griethuysen J, Fedorov A, Parmar C, Hosny A, Aerts HJWL. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res* (2017) 77(21):e104–7. doi: 10.1158/0008-5472.CAN-17-0339
20. Parmar C, Grossmann P, Rietveld D, Rietbergen MM, Lambin P, Aerts Hugo JWL. Radiomic Machine Learning Classifiers for Prognostic Biomarkers of Head & Neck Cancer. *Front Oncol* (2015) 5(4):272. doi: 10.3389/fonc.2015.00272
21. Parmar C, Grossmann P, Bussink J, Lambin P, Aerts H. Machine Learning Methods for Quantitative Radiomic Biomarkers. *Rep* (2015) 5:13087. doi: 10.1038/srep13087
22. Peng H, Long F, Ding C. Feature Selection Based on Mutual Information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy. *IEEE Trans Pattern Anal Mach Intell* (2005) 27:1226–38. doi: 10.1109/TPAMI.2005.159
23. Huang Y-Q, Liang C-H, He L, Tian J, Liang CS, Chen X, et al. Development and Validation of a Radiomics Nomogram for Preoperative Prediction of Lymph Node Metastasis in Colorectal Cancer. *Sci Foundation China* (2016) 34 (18):2157–64. doi: 10.1200/JCO.2015.65.9128
24. Grabowska-Derlatka L, Derlatka P, Szeszkowski W, Cieszanowski A, Wang Y-X. Diffusion-Weighted Imaging of Small Peritoneal Implants in "Potentially" Early-Stage Ovarian Cancer. *BioMed Res Int* (2016) 2016:9254742–9254742. doi: 10.1155/2016/9254742
25. Rubini G, Altini C, Notaristefano A, Merenda N, Rubini D, Ianora AAS, et al. Role of 18F-FDG PET/CT in Diagnosing Peritoneal Carcinomatosis in the Restaging of Patient With Ovarian Cancer as Compared to Contrast Enhanced CT and Tumor Marker Ca-125. *Rev Esp Med Nucl Imagen Mol* (2014) 33(1):22–7. doi: 10.1016/j.remn.2013.06.008
26. Zhang B, Tian J, Dong D, Dongsheng G, Shuixing Z. Radiomics Features of Multiparametric MRI as Novel Prognostic Factors in Advanced Nasopharyngeal Carcinoma. *Clin Cancer Res* (2017) 23(15):4259–69. doi: 10.1158/1078-0432.CCR-16-2910
27. Chen, Qin, Nie, Ke, Niu, Tianye, et al. Rectal Cancer: Assessment of Neoadjuvant Chemoradiation Outcome Based on Radiomics of Multiparametric MRI. *Clin Cancer Res* (2016) 22(21):5256–64. doi: 10.1158/1078-0432.CCR-15-2997
28. Thrall Melissa M, DeLoia Julie A, Gallion H, Avril N. Clinical Use of Combined Positron Emission Tomography and Computed Tomography (FDG-PET/CT) in Recurrent Ovarian Cancer. *Gynecol Oncol* (2007) 105 (1):17–22. doi: 10.1016/j.ygyno.2006.10.060
29. Bilici A, Ustaalioglu B, Seker M, Nesrin C, Bulent T, Taflan S, et al. Clinical Value of FDG PET/CT in the Diagnosis of Suspected Recurrent Ovarian Cancer: Is There an Impact of FDG PET/CT on Patient Management? *Eur J Nucl Med Mol Imaging* (2010) 37(7):1259–69. doi: 10.1007/s00259-010-1416-2
30. Chen D, Liu Z, Liu W, Fu M, Jiang W, Xu S, et al. Predicting Postoperative Peritoneal Metastasis in Gastric Cancer With Serosal Invasion Using a Collagen Nomogram. *Nat Commun* (2021) 12(1):179. doi: 10.1038/s41467-020-20429-0
31. Arakawa, Keiichi, Kawai, Kazushige, Ishihara, Soichiro, et al. Prognostic Significance of Peritoneal Metastasis in Stage IV Colorectal Cancer Patients With R0 Resection: A Multicenter, Retrospective Study. *Dis Colon Rectum* (2017) 60l:1041–9. doi: 10.1097/DCR.0000000000000858
32. Werner ME, Karve S, Sukumar R, Cummings ND, Copp JA, Chen RC, et al. Folate-Targeted Nanoparticle Delivery of Chemo- and Radiotherapeutics for the Treatment of Ovarian Cancer Peritoneal Metastasis. *Biomaterials* (2011) 32(33):8548–54. doi: 10.1016/j.biomaterials.2011.07.067

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

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

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

frontiers in Oncology

ORIGINAL RESEARCH
published: 05 November 2021
doi: 10.3389/fonc.2021.755271

Integrative Radiogenomics Approach for Risk Assessment of Postoperative and Adjuvant Chemotherapy Benefits for Gastric Cancer Patients

Yin Jin[1†], *Yilun Xu*[1†], *Yanyan Li*[2], *Renpin Chen*[1*] *and Weiyang Cai*[1*]

[1] Department of Gastroenterology, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou, China, [2] Department of Urology, Second Affiliated Hospital & Yuying Children's Hospital of Wenzhou Medical University, Wenzhou, China

OPEN ACCESS

Edited by:
Jose Eduardo Villarreal Barajas, Royal Devon and Exeter Hospital, United Kingdom

Reviewed by:
Kujtim Latifi, Moffitt Cancer Center, United States
Weiwei Zong, Henry Ford Health System, United States

***Correspondence:**
Weiyang Cai
caiweiyang@sjtu.edu.cn
Renpin Chen
lightfish.8@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: 08 August 2021
Accepted: 27 September 2021
Published: 05 November 2021

Citation:
Jin Y, Xu Y, Li Y, Chen R and Cai W (2021) Integrative Radiogenomics Approach for Risk Assessment of Postoperative and Adjuvant Chemotherapy Benefits for Gastric Cancer Patients. Front. Oncol. 11:755271. doi: 10.3389/fonc.2021.755271

Gastric cancer (GC) is a typical heterogeneous malignant tumor, whose insensitivity to chemotherapy is a common cause of tumor recurrence and metastasis. There is no doubt regarding the effectiveness of adjuvant chemotherapy (ACT) for GC, but the population for whom it is indicated and the selection of specific options remain the focus of present research. The conventional pathological TNM prediction focuses on cancer cells to predict prognosis, while they do not provide sufficient prediction. Enhanced computed tomography (CT) scanning is a validated tool that assesses the involvement of careful identification of the tumor, lymph node involvement, and metastatic spread. Using the radiomics approach, we selected the least absolute shrinkage and selection operator (LASSO) Cox regression model to build a radiomics signature for predicting the overall survival (OS) and disease-free survival (DFS) of patients with complete postoperative gastric cancer and further identifying candidate benefits from ACT. The radiomics trait-associated genes captured clinically relevant molecular pathways and potential chemotherapeutic drug metabolism mechanisms. Our results of precise surrogates using radiogenomics can lead to additional benefit from adjuvant chemotherapy and then survival prediction in postoperative GC patients.

Keywords: radiogenomics, adjuvant chemotherapy benefit, postoperative gastric cancer, nomogram, survival prediction

INTRODUCTION

Gastric cancer (GC) is now the third most common cancer and the second leading cause of cancer-related mortality worldwide (1), of which nearly three-quarters occurred in Asia, and more than two-fifths occurred in China (2). Pathological staging according to the TNM (tumor, lymph node, and metastasis) system and histological subtype have been commonly recognized as the most used master for the prognostic definition and treatment strategy choice in GC. Complete surgical excision is generally conducted as the primary intervention for the majority of stage I–III and partial stage IV GC patients, but the 5-year recurrence and survival rate still spread over a broad range.

These findings reflected that GC is characteristic of biological heterogeneity with large variations in clinical outcomes even among those with the same stage. Thus, it is urgent to improve the prediction of GC prognosis by developing a novel signature to categorize patients and predict further survival.

Radiomics is a newly developing approach that transfers imaging data into a high-dimensional mineable feature space using a large number of automatically applied data-characterization algorithms (3, 4). Radiomics translates the genomic heterogeneity into expression in an intratumoral heterogeneity through imaging (5, 6). On the other hand, radiomics signature has the power to capture intratumoral heterogeneity through a noninvasive method. Previous studies have demonstrated that the prognosis and malignant degree of GC were closely related to imaging features. For example, metabolically active tumor volume (MATV) has been proven to be a prognostic factor in patients with GC (7); Li et al. constructed a radiomics signature of 18-F fluorodeoxyglucose PET/CT for prediction of GC survival (8); Jiang et al. selected 19 potential predictors from the 269 features identified, which provided a neoteric angle for individualized diagnosis and prediction of malignancy potential for GC patients (9); Jiang et al. developed machine learning for predicting the pathological stage for GC (8); and studies established a deep learning radiomics model for effectively predicting the lymph node metastasis of local GC (10). However, these radiomics studies did not show satisfied diagnostic efficiency, and many controversial results still existed. The principal underlying explanations might be that these previous studies were only based on dated imaging technology, which only extracted 269 features from the non-filtered segmented ROI, and that the results were easily influenced by different individuals and lacked a proper validation. More importantly, a radiomics nomogram research investigating the association of post-operative GC and candidates' selection for ACT has not yet been fully reported. Thus, more contributing factors should be offered for choice in the intended population and ideal regimens before therapy selection.

In this study, we adopted a quantitative radiomics approach and developed a multiple-feature-based radiomics signature, which function in predicting survival and assessing benefit from ACT for postoperative GC patients. Additionally, we also firstly explored the potential biological basis of radiomics with imaging and gene expression data. The radiomics trait-associated genes captured clinically relevant molecular pathways and potential chemotherapeutic drug metabolism.

MATERIAL AND METHODS

Study Population and Design

Figure 1 presents the workflow of the study. We utilized two independent datasets in this study that were re-collected from an institution in China and from open-access online repositories, respectively. This study was approved by the institutional research board of the First Affiliated Hospital of Wenzhou Medical University. The records and images of 428 persons diagnosed with GC between January 2014 and January 2017 were reviewed. All these patients satisfied the following inclusion criteria: 18 years or over; firstly diagnosed with primary gastric cancer; excluded other malignant tumors; with CT images within 1 month prior to therapy; complete resection of the tumor tissue; without serious heart, lung, or kidney dysfunction; complete pathology and laboratory; and able to provide informed consent. Simultaneously, a part of the chosen patients successively underwent surgery and 8–12 regular periods of standard chemotherapy (including S1 alone, XELOX and FOLFIRI/FOLFOX). All of these patients were followed up, and recurrent and dead patients were recorded during the follow-up. The cutoff time of the study was set in May 2021. After exclusion, 417 enrolled patients were divided into two datasets: 172 patients were assigned to the training set, whereas 245 patients were assigned to the validation set. On the other hand, a dataset comprised CT imaging data and matched RNA sequencing data of 47 resected GC were obtained from the TCGA database of The Cancer Imaging Archive (TCIA) to evaluate the biological process of radiomics signature. Disease-free survival (DFS) was defined as the duration from the date of diagnose to that of recurrence, death, or the last follow-up. Overall survival (OS) was defined as the duration from diagnosis to death or the last follow-up.

Feature Extraction and Selection

For each patient, the tumor region was contoured in a slice-by-slice manner on CT images by two experienced radiologists using MRIcroGL and 3D Slicer software (http://www.mccauslandcenter.sc.edu, www.slicer.org) (11, 12). We manually segmented the contour of the tumor in the arterial phase 5-mm CT images for image feature extraction. An open-source platform, PyRadiomics in Python (https://github.com/mvallieres/radiomics/), was utilized to extract 799 radiomics features from the selected segmented ROI (13), which include seven types of indexes, namely, Shape; First-Order Statistics; Gray Level Co-occurrence Matrix; Gray Level Run Length Matrix; Gray Level Size Zone Matrix; Neighboring Gray Tone Difference Matrix; and Gray Level Dependence Matrix. Feature selection is essential to avoiding overfitting, which was devised to reduce error to the high-dimensional radiomics features. At first, the intraclass correlation coefficients (ICCs) were calculated between the features extracted from the two radiologists, and features with either intra-observer or interobserver ICCs less than 0.75 were excluded. Secondly, features with significant differences between PD and non-PD (CR + PR + SD) groups were selected through a t-test (14).

Construct Radiomics Score and Combination Nomogram

The least absolute shrinkage and selection operator (LASSO) method was applied to select the most optimal radiomics feature subsets for predicting the radiomics score. A radiomics score was calculated for each patient *via* weighting by their LASSO Cox coefficients: wavelet-LLH glszm SizeZone NonUniformity Normalized*1.8977 – wavelet-LHL glcm Cluster Shade*0.0009 +

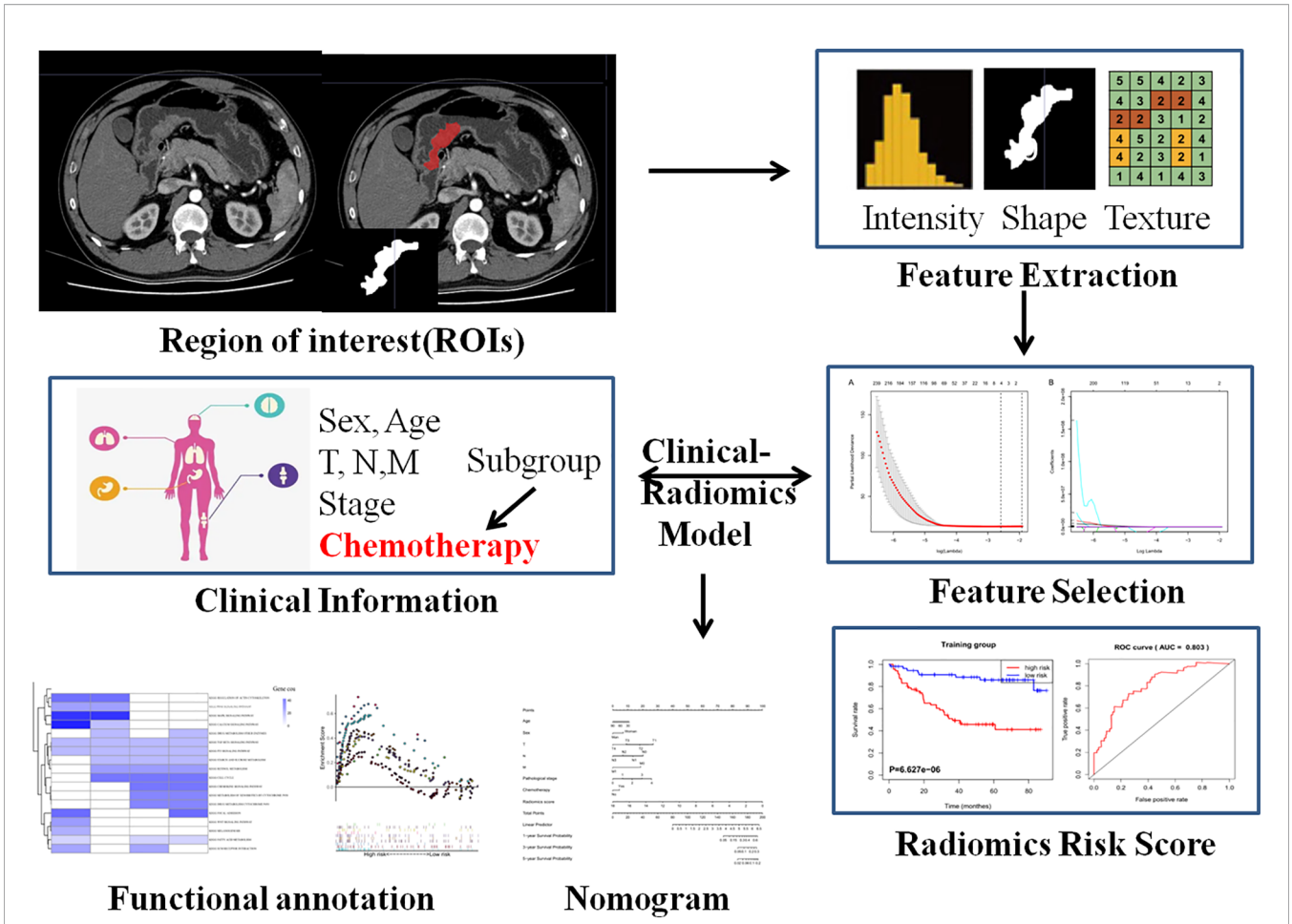


FIGURE 1 | Flowchart of the study design and main process.

wavelet-LHH first order Kurtosis*0.005364 + wavelet-HLH glszm Large Area Low Gray Level Emphasis*5.97E-06. Setting the median score as the cutoff line, the GC cohort was divided into low- and high-risk groups, then K–M survival analysis was employed. Furthermore, this calculation and the cutoff value were applied to the validation cohort.

The OS and DFS nomograms were constructed based on the main prognostic factors to predict the 1-, 3-, and 5-year survival for patients. Each patient could sum up variable scores and finally establish predictive measures of survival and relapse. The calibration curve for predicting the 1-, 3-, and 5-year OS and DFS indicated that the nomogram-predicted survival closely corresponded with actual survival outcomes. The survival analysis was conducted using rms, survival, and survcomp package. Hazard ratios (HRs) and 95% confidence intervals (CIs) were recorded.

Gene Set Enrichment Analysis and Functional Annotation

Gene set enrichment analysis (GSEA) was performed to explore the biological basis of the radiomics signature for prognosis prediction (15). The raw RNA-seq expression data were downloaded and normalized using the limma package. GO enrichment analysis of the signature genes was conducted using the R package clusterProfiler. Significantly enriched biological processes are summarized.

Statistical Analysis

The experimental data were analyzed by Prism 5.0 software (GraphPad) and R software (version 3.4.2, http://www.R-project.org). Student's t test and Wilcoxon's test were utilized to compare continuous variables and ordered categorical variables. Survival analysis was performed using the Kaplan–Meier method, and the survival of the clusters was compared using the log-rank test.

RESULTS

Baseline Clinic-Pathological Characteristics

At first, a total of 417 GC patients from Wenzhou Medical University met the inclusion criteria from January 1, 2014, to January 1, 2016. In this study, about 43.9% of the patients only

received surgery, while the others conducted both adjuvant postoperative therapy and surgery. As of March of 2021, 183 patients died during follow-up and none were lost to follow-up. Among the total patients, 281 patients (60.0%) achieved tumor control (CR + PR + SD), while 136 patients (40.0%) had progressive disease. The median follow-up time was 30.25 and 31.42 months in the training and validation cohorts, respectively. There was no significant difference between the two cohorts in terms of clinicopathologic factors or follow-up time. The baseline clinico-pathological parameters are summarized in **Supplementary Table 1**.

Radiomics Feature Selection and Signature Construction

A total of 799 features were extracted from the tumor volume, in which each sample quantified the intratumor heterogeneity by using the Gray Level Co-occurrence (GLCM), Gray Level Run Length Matrices (GLRLM), Gray Level Size Zone Matrix (GLSZM), Neighboring Gray Tone Difference Matrix (NGTDM), and Gray Level Dependence Matrix (GLDM) indexes. We firstly excluded redundant features with ICCs less than 0.75 and then conducted the t-test to select the features with statistically significant differences between PD and non-PD (CR + PR + SD) groups. Finally, four features were selected *via* LASSO regression (**Figure 2** and **Supplementary Figure 1**). The radiomics score of each GC case was calculated by a weighted linear combination of these four features and their corresponding coefficients. For low- and high-risk GC patients, the median radiomics score was selected as the cutoff line (**Figures 3A, B**). In either the total, training, or validation set, the median radiomics scores in the low-grade group were higher than those in the high-grade group ($p < 0.001$ for all patients, **Figure 3C**). We eliminated the interaction between the components of the original data by PCA analysis (**Figure 3D**). The overall survival distribution and ROC curves shown in **Figures 3E, F** yield significant prognosis outcome and AUCs of 0.803 for the training set and 0.753 for the validation set, demonstrating the discriminative power of the radiomics signature. Moreover, similar results were also obtained for the PFS analysis (**Supplementary Figure 2**).

To verify that the accuracy of the radiomics model was also important in additional GC cases, we further selected TCGA GC cohorts for validation. Consistent with the above results, the high radiomics subgroup had a worse prognosis than the low one. The distribution of radiomics score and survival information of patients were analyzed and are shown in **Supplementary Figures 3A, B**.

ACT Benefit Analysis Based on the Radiomics Score

Previous data suggested that image features are closely associated with chemotherapy efficacy; thus, we evaluated the benefit of chemotherapy according to the level of radiomics score in this study. As shown in **Figures 4A, B**, the adoption of ACT (n = 234/428) did not show significant OS and DFS survival benefit in all patients with complete postoperative GC patients ($p = 0.3362$ and 0.067, respectively). Using the median radiomics score as the cutoff line, we divided patients by chemotherapy therapy. As for the low radiomics subgroup, GC patients obtained a terrible response to ACT (**Figures 4C, D**), while patients with a high risk point showed no significant survival difference with or without ACT (**Figures 4E, F**). For patients with low radiomics scores, more effective systemic approaches to improve treatment outcomes need to be identified. Thus, the radiomics score was both a prognostic and predictive tool for post-operation GC patients.

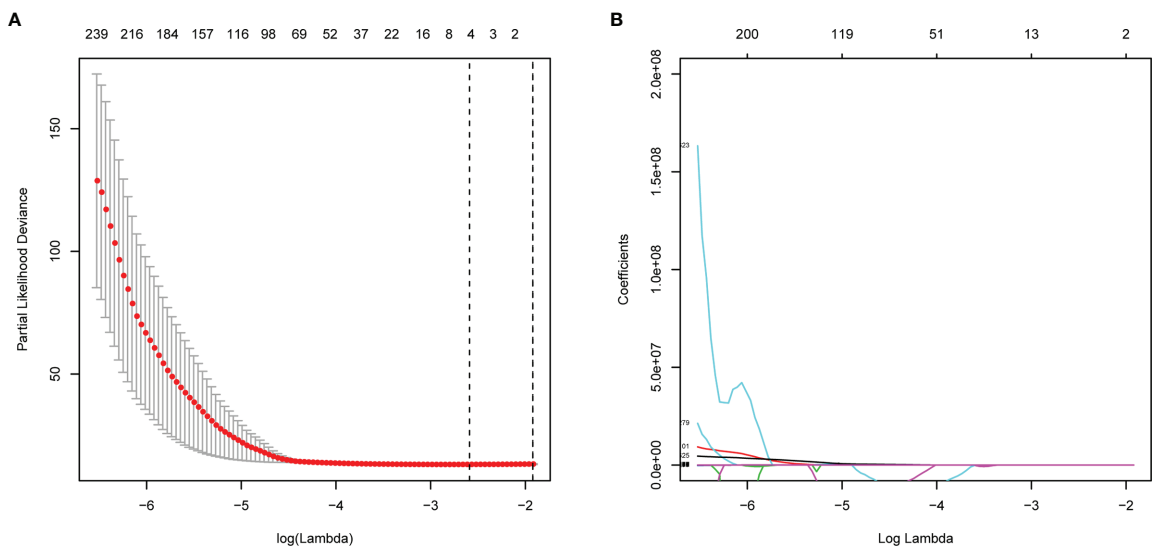


FIGURE 2 | Feature selections with LASSO Cox regression analysis. **(A)** λ selection in the LASSO model using 10-fold cross-validation. **(B)** LASSO coefficient profiles of all the radiomics features. Vertical black dashed line represents the optimal resulted in nine nonzero features.

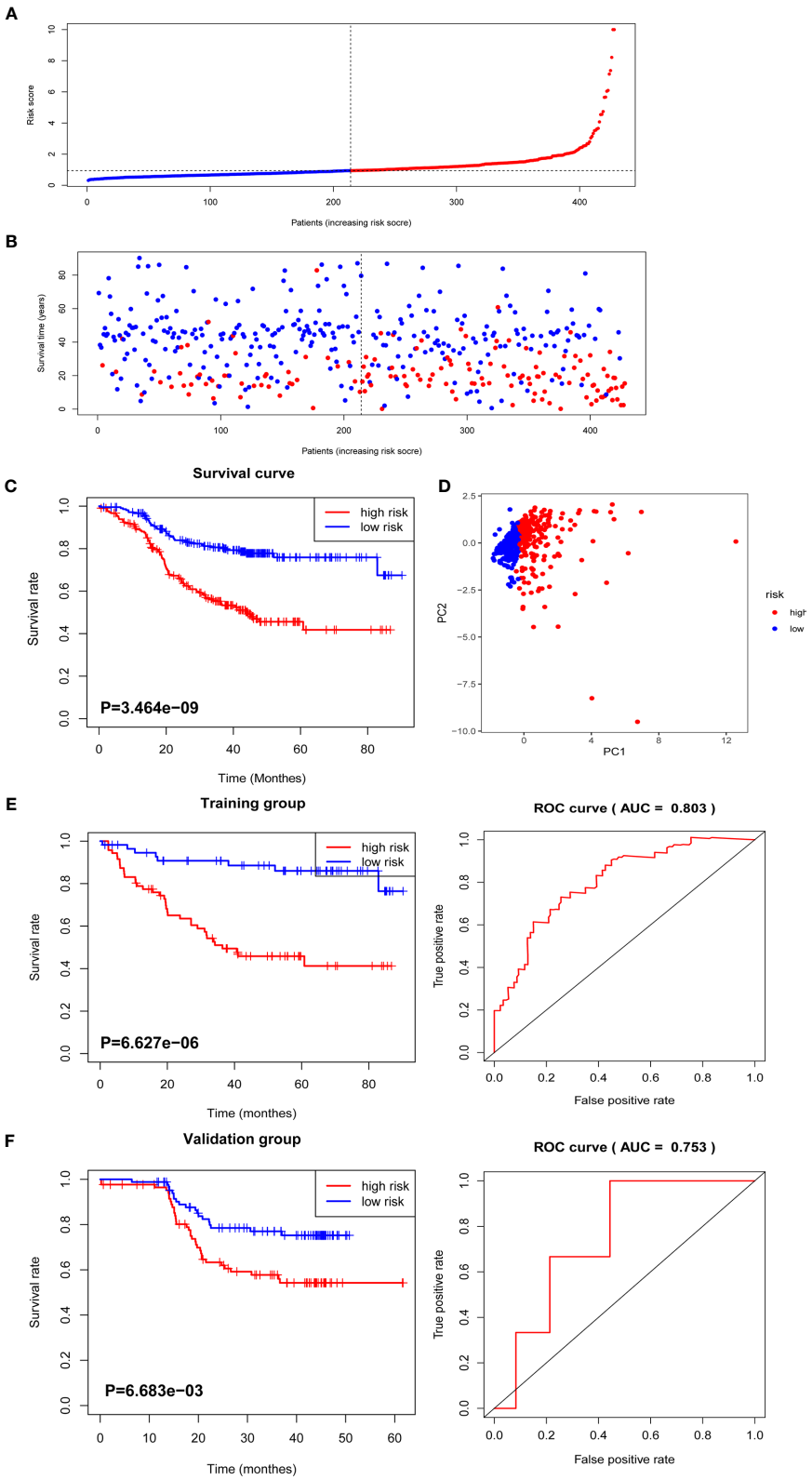


FIGURE 3 | The distribution of radiomics score and overall survival analysis. **(A)** The distribution of the radiomics score of GC patients. **(B)** The overall survival status for each GC patient. **(C)** Overall survival curve of total GC patients. **(D)** PCA depicts the variation of high- and low-radiomics scores. **(E)** Overall survival and ROC curves of the training group. **(F)** Overall survival and ROC curves of the validation group.

Radiomics Nomogram Construction

For OS and DFS, the multivariate analysis demonstrated that glszm Size Zone NonUniformity Normalized, glcm Cluster Shade, First-Order Kurtosis, and glszm Large Area Low Gray Level Emphasis were all significantly associated with survive ($p \leq 0.05$, **Supplementary Figure 1**). We further incorporated radiomics signature with other clinical factors, which has been proven to add prognostic information to better identify patients with different outcomes, and the radiomics nomogram was a good witness. As shown in **Figure 5**, the radiomics signature combined with TNM staging significantly reinforced the prognostic ability.

OS and DFS nomograms were constructed to predict the 1-, 3-, and 5-year overall survival and relapse of the GC (**Figure 5A** and **Supplementary Figure 4A**). Total score summations of each variable were based on the intersection of the vertical line. As shown in **Figure 5A**, radiomics score contributed the most risk points (ranged, 0–100), whereas other clinical information contributed much less (ranged, 0–40). By using this nomogram, we could convert each clinical index to the corresponding point and then calculate the total point, which was used to evaluate the 1-, 3-, and 5-year survival or relapse rate. Moreover, decision curve analysis showed a high accuracy of the predictive prognostic radiomics score (**Figures 5B–D**, and **Supplementary Figures 4B**, **D**). Decision curve analysis showed great predictive accuracy of prognostic nomograms for OS and DFS. In other words, the clinical-radiomics nomogram that incorporated both clinical risk factors and radiomics parameters showed excellent performance, with high 1-, 3-, and 5-year AUCs of 0.80, 0.816, and 0.965, respectively (**Figure 5E** and **Supplementary Figure 4E**).

Biological Basis of the Radiomics Signature

We further conducted genomic analysis to explore the molecular underpinning of the identified all-relevant features by evaluating possible radiogenomics links using the RNA-Sep technology. More than 70% of significantly different expression genes in the module of glszm Size Zone NonUniformity Normalized and glcm Cluster Shade were upregulated in the tumor tissues, while a small part of genes were negatively expressed (**Figure 6A**). When we examined the degree of overlap between radiomics feature genes to investigate the dependence between each feature, glcm Cluster Shade and glszm Large Area Low Gray Level Emphasis showed an overall high similarity (**Figure 6B**). The pre-ranked GSEA showed that the significant enriched pathways (FDR <0.1) among the top associations with these four radiomics factors were mostly correlated with drug metabolism and

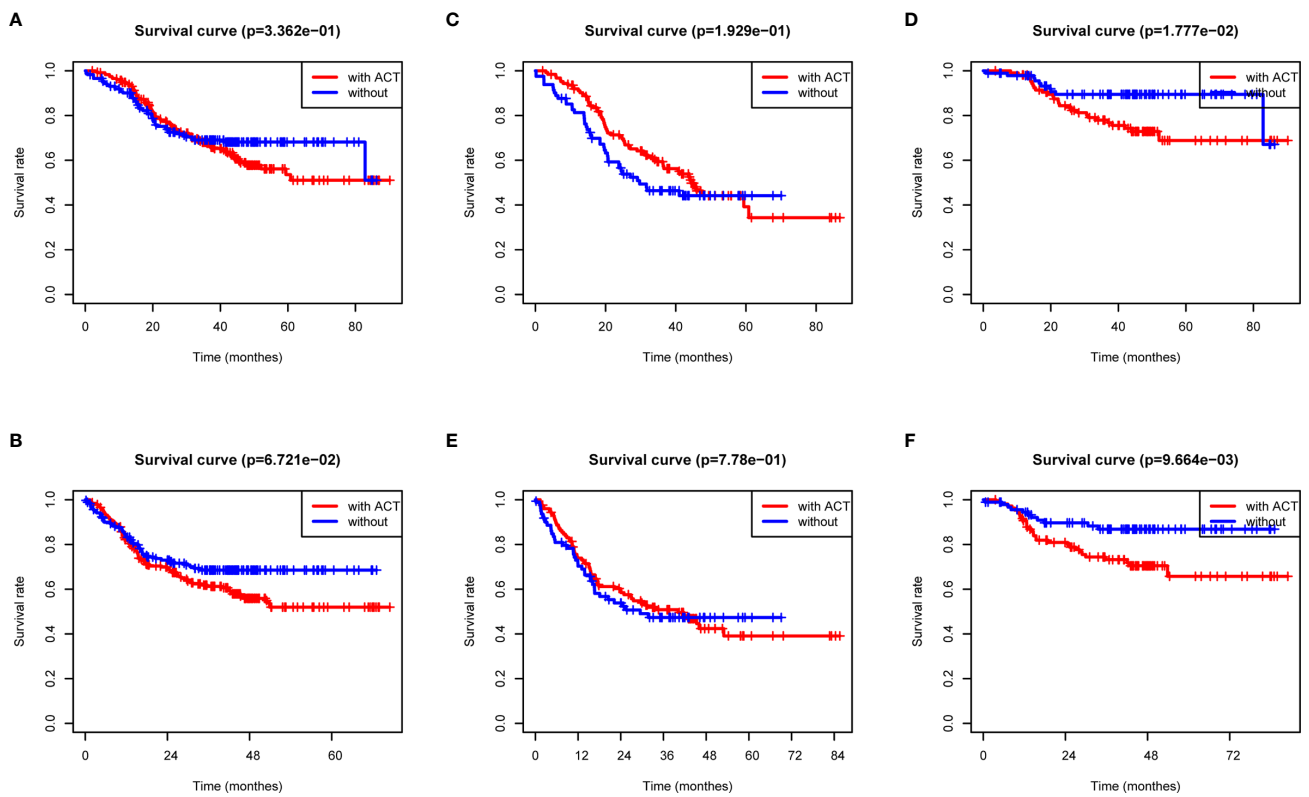


FIGURE 4 | ACT benefit analysis based on the radiomics score. **(A)** Overall survival analysis of all GC patients according to ACT. **(B)** Disease-free survival analysis of all GC patients according to ACT. **(C)** Overall survival analysis of low-radiomics score GC patients according to ACT. **(D)** Overall survival analysis of high-radiomics score GC patients according to ACT. **(E)** Disease-free survival analysis of low-radiomics score GC patients according to ACT. **(F)** Disease-free survival analysis of high-radiomics score GC patients according to ACT.

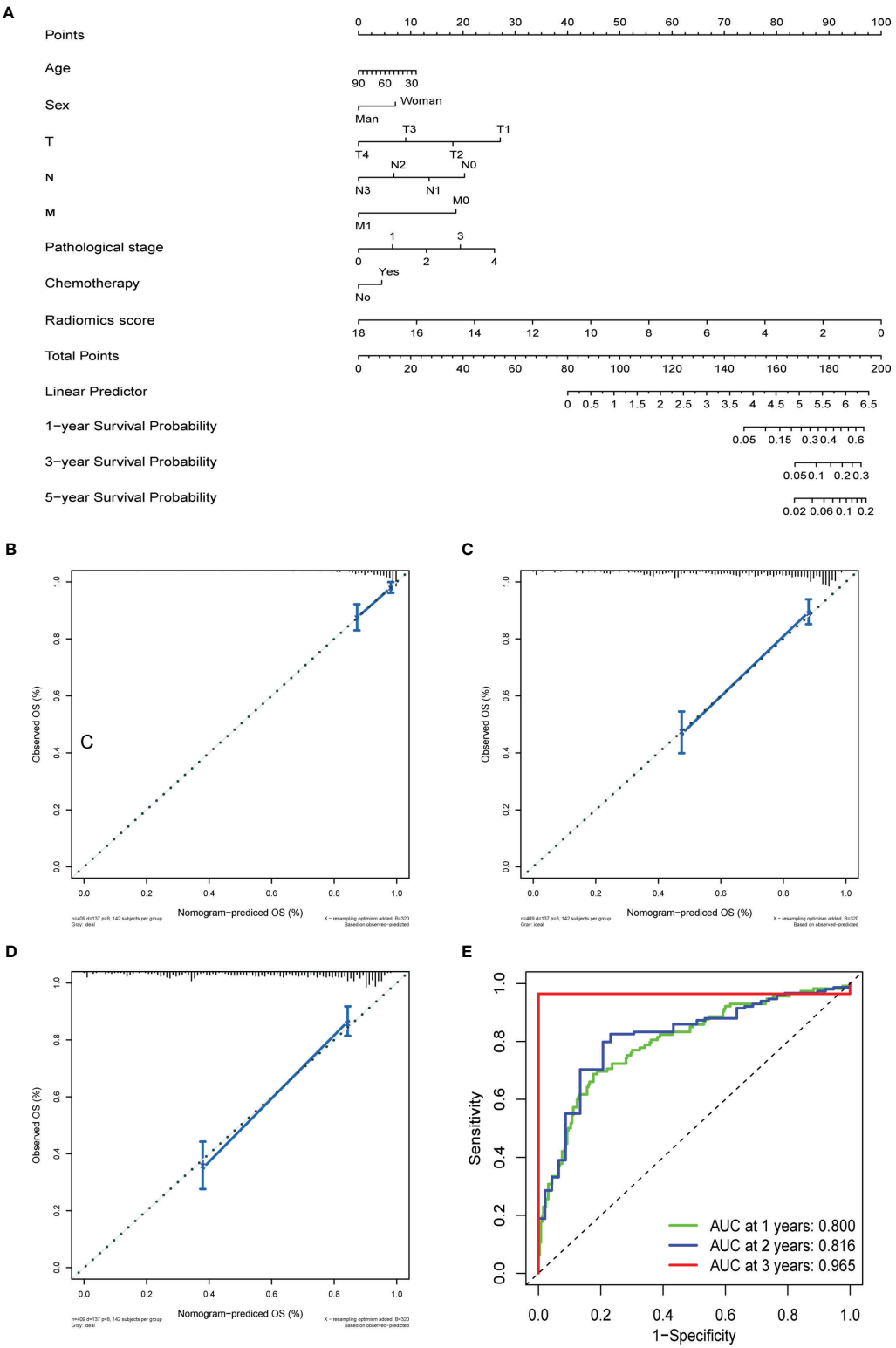


FIGURE 5 | The prediction performance analysis of overall survival. **(A)** The nomogram for predicting 1-, 3-, and 5-year OS after surgery. **(B)** Calibration curve for risk of 1-year overall survival. **(C)** Calibration curve for risk of 3-year overall survival. **(D)** Calibration curve for risk of 5-year overall survival. **(E)** Receiver operating characteristic (ROC) curves for 1-, 3-, and 5- year overall survival.

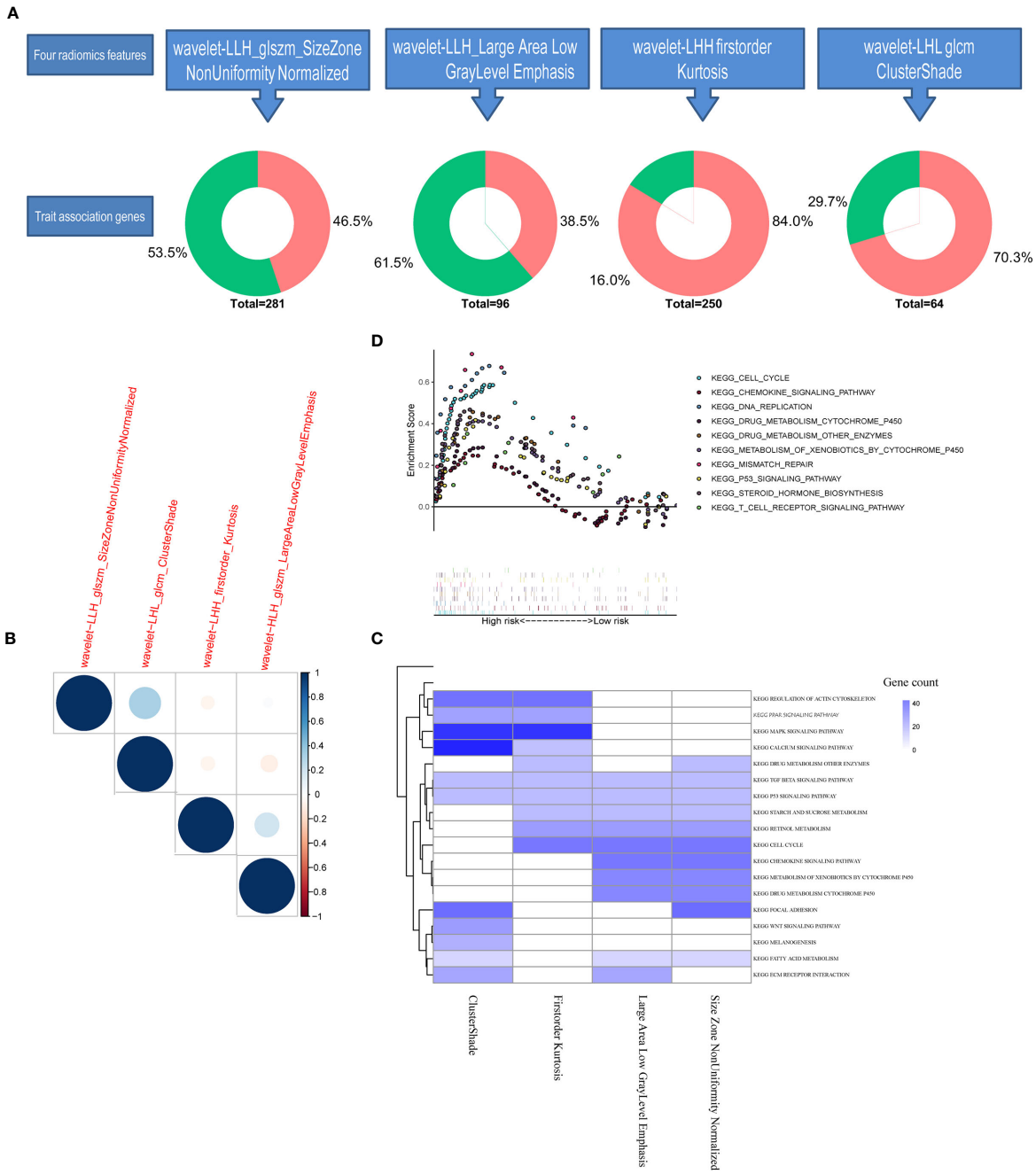


FIGURE 6 | Tarit-association genes and functional enrichment analysis. **(A)** Each differently expressed gene in four radiomics features in the genomic study. **(B)** Heatmap of the similarity between each train-associated gene calculated by the Spearman index. **(C)** Heatmap of GESA enrichment analysis. **(D)** GSEA enrichment analysis of the radiomics score model.

chemokine regulation (**Figure 6C**). Of the radiomics score signatures, the most enriched pathways were also gathered in the drug metabolism cytochrome P450 and other enzymes (**Figure 6D**). These results revealed that the developed imaging biomarker might reflect the different drug metabolic changes during cancer therapy, which could better stratify patients for more precise therapeutic care.

DISCUSSION

In consideration of the complexity of the heterogeneity of the morphologic, biologic, and clinical nature, conventional classification systems were no longer able to reflect the complex molecular nature of GC. As for proposed morphology-based classification systems, the World Health

Organization (WHO) (papillary, tubular, mucinous, and poorly cohesive (PCC-NOS)/signet ring) (16) and the Lauren (intestinal, diffuse, and mixed) classifications are the most commonly preferred (17). As for comprehensive genomic assays, the TCGA Consortium classified GC into four major genomic subtypes, including tumors positive for Epstein–Barr virus (EBV), microsatellite unstable tumors (MSI), genomically stable tumors (GS), and tumors with chromosomal instability (CIN) (18); the Asian Cancer Research Group (ACRG) divided GC into four subtypes: MSI, microsatellite stable (MSS)/epithelial–mesenchymal transition (EMT), MSS/tumor protein 53 (TP53) active, and MSS/TP53 inactive (19). Moreover, a plethora of genomic alterations have been identified, such as HER2, FGFR2, EGFR, PI3K, mTOR, and MET, which provided good identification of GC patients who derive therapeutic benefit from ACT (20). Therefore, the existence of GC intra-tumor heterogeneity affects the accuracy of clinical decisions and lead to a substantial discordance rate. Thus, researchers are now investigating auxiliary noninvasive approaches to precisely predict the therapy of GC preoperatively.

Radiomics is a promising tool which defines mathematical features from medical images using a series of data-characterization algorithms. This technique allows doctors to access standardized image texture information and to stand out informed inference. To date, radiomics has been proven to effectively predict biological characteristics of numerous types of cancers (7, 19, 20). In GC patients, although studies were preliminary, radiomics texture analyses have been proven to not only improve prediction of survival but also provide additional information in oncologic practice related to benign and malignant nodule differentiation, prediction of lymph node metastasis, histological subtype classification, response to chemotherapy assessment, and mutation type identification (9, 15, 17). In this study, a total of 799 two-dimensional features were extracted from each ROI by PyRadiomics. After dispelling redundancies, four selected radiomics features were extracted to construct a radiomics signature, which enabled more accurate identification of GC patients who might benefit from postoperation chemotherapy. As shown in the present study, the radiomics signature successfully identified high-risk GC patients with poor survival outcomes, for whom more intensified treatment was needed. Furthermore, for patients with a low radiomics score were more inclined to fall victim in the postoperation chemotherapy. By adding the clinical features, we also constructed the clinical-radiomics feature nomogram for predicting survival of GC after gastrectomy.

Surgical resection is the main curative method for GC, but the high rate of relapse in patients makes it important to consider adjuvant treatment selection. Current guidelines have strongly proven chemotherapy as a standard component for advanced GC therapies, whereas existing studies provided that a subgroup of patients does not benefit from the present ACT. At present, three major international GC guidelines guide the population indicated for ACT, which can be roughly divided into Europe (European Society for Medical Oncology, ESMO), United States (National Comprehensive Cancer Network, NCCN), and East Asia (21). Based on the MAGIC trial, ESMO recommended ACT for stage >T1N0 (22) and NCCN suggested clinical stage ≥ T2; yet, the Japanese guidelines still suggested all surgeries combined with postoperative chemotherapy (23). Although the efficacy of ACT for GC has been proven, there is no ideal measure for reasonable noninvasive selection, especially for early postoperative GC patients. The NCCN guidelines recommend ACT for pT1N1 gastric cancer patients after curative resection (24); on the contrary, the Japanese Gastric Cancer Treatment Guidelines did not show any ACT benefit with regard to pT1N1 gastric cancer patients after curative resection (24). By and large, the risk factors associated with postoperation chemotherapy selection include tumor invasion, lymph node metastases, tumor stage and Borrmann type, dMMR, gene mutant, family history, and physical condition. Although intensive protocols are promising, selecting the optimal adjuvant chemotherapy remains a difficult task that requires a balance between the therapeutic benefits and toxicity. The most common adverse events were asthenia/anorexia (33.3%), hematologic malignancies (29.6%), and infection (14.8%) (25). Thus, a biomarker study is urgently necessary for selecting the GC subgroups for which adjuvant treatment provides an oncological benefit postoperation. Our study provided a statistically robust approach to construct the radiomics signature for the administration of ACT in GC. The radiomics signature provides the incremental value for guiding the adoption of ACT in patients with a low radiomics score. The radiomics score elucidated the relationship between tumor characteristics and their imaging appearance as well as developed imaging biomarkers that can predict risk and outcomes, thereby better stratifying patients for more precise therapeutic care.

The radiogenomics analysis provided that a prognostic radiomics signature could capture tumor cell intratumor heterogeneity, which is also associated with underlying gene expression patterns. The radiogenomics analysis showed multiple associations between CT image features and trait-associated genes mostly correlated with various drug metabolisms and chemokine regulation. Given that radiomics signature provided the incremental value for the adoption of ACT, although the mechanism of the relationship between radiomics features and chemotherapy has not been shown thoroughly, we speculated that it may be associated with the strong correlation with cell cycling pathways, chemokine signaling, and chemotherapeutic drug metabolism (**Figure 6**). The present image-to-molecular feature associations could also be applied to assess therapeutic options based on biological pathway activity. The effects of whole-body chemotherapy for GC may be deeply influenced by drug metabolism and critical signaling pathways. First-Order Kurtosis, Large Area Low Gray Level Emphasis, and Size Zone NonUniformity Normalized could be effectively targeted by drug metabolism *via* cytochrome P450 and other enzymes, retinol metabolism, and sucrose metabolism. The gene pathway analysis indicated that radiogenomics may be suitable for predicting the efficacy of pathway–target therapies.

However, some limitations in this study should also be noticed. Firstly, the small sample size of GC patients and the

retrospective nature of the data collection possibly affect the statistical power. We need to increase the sample size and conduct a multicenter research to verify the accuracy and stability of the radiomics nomogram model. Secondly, the decision to treat or not to treat patients after surgery was made by the patients and/or clinicians, and the use of adjuvant chemotherapy was not within a randomized comparison. Chemotherapy drug side effects, types of chemotherapy selection, and irregular course of chemotherapy all existed in our follow-up study, which inevitably affected the outcome. Moreover, radiomics does indeed suffer from a closed-source nature, unharmonized acquisition settings, discordant reconstruction parameters, lack of interpretability, redundancy, and methodological bias. Complementary innovations in genetic and imaging-based studies that allow for the spatial quantification of tumor heterogeneity could provide a realization of precision oncology.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

WC and YJ conceived of and designed the experiments. YL, YX, and RC performed the data collection. WC and YX analyzed the data. WC, YX, and YJ wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was partially supported by grants from the Natural Science Foundation of Zhejiang Province of China (LY18H030008); Wu Jieping Medical Foundation (320.6750.17396), Wenzhou Science and Technology Bureau (No. 20190432), and the First Affiliated Hospital of Wenzhou Medical University (Grant No. FHY2019002).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2021.755271/full#supplementary-material

Supplementary Figure 1 | Forest plot showing the logistic regression analyses of the four radiomic features.

Supplementary Figure 2 | The distribution of radiomics score and disease-free survival. **(A)** The distribution of radiomics score of GC patients; **(B)** The disease-free survival status for each GC patients; **(C)** disease-free survival curve of total GC patients; **(D)** PCA depicts the variation of high- and low- radiomics score; **(E)** disease-free survival and ROC curves of training group; **(F)** disease-free survival and ROC curves of validation group.

Supplementary Figure 3 | Validation analysis of the radiomics score in TCGA cohort. **(A)** Overall survival curve of the TCGA radiomics score. **(A)** ROC curves of TCGA radiomics score.

Supplementary Figure 4 | The prediction performance analysis of disease-free survival. **(A)** The nomogram for predicting 1-, 3- and 5-year PFS after surgery; **(B)** Calibration curve for risk of 1- year disease-free survival; **(C)** Calibration curve for risk of 3-year disease-free survival; **(D)** Calibration curve for risk of 5-year disease-free survival; **(E)** Receiver operating characteristic (ROC) curves for 1-, 3- and 5-year disease-free survival.

REFERENCES

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: Cancer J Clin* (2018) 68 (6):394–424. doi: 10.3322/caac.21492
2. Wang YK, Wang YC, Shan F, Tang L, Li ZY, Ji JF. Exploration of Potential Beneficial People of Neoadjuvant Chemotherapy Based on Clinical Staging in Gastric Cancer: A Single Center Retrospective Study. *Zhonghua Wei Chang Wai Ke Za Zhi Chin J Gastrointestinal Surg* (2020) 23(2):152–7. doi: 10.3760/cma.j.issn.1671-0274.2020.02.010
3. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More Than Pictures, They Are Data. *Radiology* (2016) 278(2):563–77. doi: 10.1148/radiol.2015151169
4. 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
5. Diehn M, Nardini C, Wang DS, McGovern S, Jayaraman M, Liang Y, et al. Identification of Noninvasive Imaging Surrogates for Brain Tumor Gene-Expression Modules. *Proc Natl Acad Sci USA* (2008) 105(13):5213–8. doi: 10.1073/pnas.0801279105
6. Li SP, Padhani AR, Makris A. Dynamic Contrast-Enhanced Magnetic Resonance Imaging and Blood Oxygenation Level-Dependent Magnetic Resonance Imaging for the Assessment of Changes in Tumor Biology With Treatment. *J Natl Cancer Institute Monogr* (2011) 2011(43):103–7. doi: 10.1093/jncimonographs/lgr031
7. Choi BW, Kang S, Bae SU, Jeong WK, Baek SK, Song BI, et al. Prognostic Value of Metabolic Parameters on (18)F-Fluorodeoxyglucose Positron Tomography/Computed Tomography in Classical Rectal Adenocarcinoma. *Sci Rep* (2021) 11(1):12947. doi: 10.1038/s41598-021-92118-x
8. Jiang Y, Yuan Q, Lv W, Xi S, Huang W, Sun Z, et al. Radiomic Signature of (18)F Fluorodeoxyglucose PET/CT for Prediction of Gastric Cancer Survival and Chemotherapeutic Benefits. *Theranostics* (2018) 8(21):5915–28. doi: 10.7150/thno.28018
9. Jiang Y, Chen C, Xie J, Wang W, Zha X, Lv W, et al. Radiomics Signature of Computed Tomography Imaging for Prediction of Survival and Chemotherapeutic Benefits in Gastric Cancer. *EBioMedicine* (2018) 36:171–82. doi: 10.1016/j.ebiom.2018.09.007
10. Ao W, Cheng G, Lin B, Yang R, Liu X, Zhou S, et al. A Novel CT-Based Radiomic Nomogram for Predicting the Recurrence and Metastasis of Gastric Stromal Tumors. *Am J Cancer Res* (2021) 11(6):3123–34.
11. Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin JC, Pujol S, et al. 3d Slicer as an Image Computing Platform for the Quantitative Imaging

Network. *Magnetic Resonance Imaging* (2012) 30(9):1323–41. doi: 10.1016/j.mri.2012.05.001
12. Li X, Morgan PS, Ashburner J, Smith J, Rorden C. The First Step for Neuroimaging Data Analysis: DICOM to NIfTI Conversion. *J Neurosci Methods* (2016) 264:47–56. doi: 10.1016/j.jneumeth.2016.03.001
13. Rutherford M, Mun SK, Levine B, Bennett W, Smith K, Farmer P, et al. A DICOM Dataset for Evaluation of Medical Image De-Identification. *Sci Data* (2021) 8(1):183. doi: 10.1038/s41597-021-00967-y
14. Gstoettner M, Sekyra K, Walochnik N, Winter P, Wachter R, Bach CM. Inter- and Intraobserver Reliability Assessment of the Cobb Angle: Manual Versus Digital Measurement Tools. *Eur Spine J Off Publ Eur Spine Soc Eur Spinal Deformity Soc Eur Section Cervical Spine Res Soc* (2007) 16(10):1587–92. doi: 10.1007/s00586-007-0401-3
15. Varemo L, Nielsen J, Nookaew I. Enriching the Gene Set Analysis of Genome-Wide Data by Incorporating Directionality of Gene Expression and Combining Statistical Hypotheses and Methods. *Nucleic Acids Res* (2013) 41 (8):4378–91. doi: 10.1093/nar/gkt111
16. Nagtegaal ID, Odze RD, Klimstra D, Paradis V, Rugge M, Schirmacher P, et al. The 2019 WHO Classification of Tumours of the Digestive System. *Histopathology* (2020) 76(2):182–8. doi: 10.1111/his.13975
17. Lauren P. The Two Histological Main Types Of Gastric Carcinoma: Diffuse And So-Called Intestinal-Type Carcinoma. An Attempt at a Histo-Clinical Classification. *Acta Pathol Microbiol Scandinavica* (1965) 64:31–49. doi: 10.1111/apm.1965.64.1.31
18. Cancer Genome Atlas Research N. Comprehensive Molecular Characterization of Gastric Adenocarcinoma. *Nature* (2014) 513 (7517):202–9. doi: 10.1038/nature13480
19. Cristescu R, Lee J, Nebozhyn M, Kim KM, Ting JC, Wong SS, et al. Molecular Analysis of Gastric Cancer Identifies Subtypes Associated With Distinct Clinical Outcomes. *Nat Med* (2015) 21(5):449–56. doi: 10.1038/nm.3850
20. Kahraman S, Yalcin S. Recent Advances in Systemic Treatments for HER-2 Positive Advanced Gastric Cancer. *OncoTargets Ther* (2021) 14:4149–62. doi: 10.2147/OTT.S315252
21. Ao S, Wang Y, Song Q, Ye Y, Lyu G. Current Status and Future Perspectives on Neoadjuvant Therapy in Gastric Cancer. *Chin J Cancer Res Chung-kuo Yen Cheng Yen Chiu* (2021) 33(2):181–92. doi: 10.21147/j.issn.1000-9604.2021.02.06
22. Fitzgerald TL, Efird JT, Bellamy N, Russo N, Jindal N, Mosquera N, et al. Perioperative Chemotherapy Versus Postoperative Chemoradiotherapy in Patients With Resectable Gastric/Gastroesophageal Junction Adenocarcinomas: A Survival Analysis of 5058 Patients. *Cancer* (2017) 123 (15):2909–17. doi: 10.1002/cncr.30692
23. Park SH, Son T, Seo WJ, Lee JH, Choi YY, Kim HI, et al. Prognostic Impact of Extended Lymph Node Dissection Versus Limited Lymph Node Dissection on Pn0 Proximal Advanced Gastric Cancer: A Propensity Score Matching Analysis. *J Gastric Cancer* (2019) 19(2):212–24. doi: 10.5230/jgc.2019.19.e20
24. Park J, Kim YJ, Song MK, et al. Definitive Chemoradiotherapy Versus Radical Hysterectomy Followed by Tailored Adjuvant Therapy in Women With Early-Stage Cervical Cancer Presenting With Pelvic Lymph Node Metastasis on Pretreatment Evaluation: A Propensity Score Matching Analysis. *Cancers* (2021) 13(15). doi: 10.3390/cancers13153703
25. Mansouri H, Zemni I, Achouri L, Mahjoub N, Ayedi MA, Ben Safta I, et al. Chemoradiotherapy or Chemotherapy as Adjuvant Treatment for Resected Gastric Cancer: Should We Use Selection Criteria? *Rep Pract Oncol Radiother J Greatpoland Cancer Center Poznan Polish Soc Radiat Oncol* (2021) 26 (2):266–80. doi: 10.5603/RPOR.a2021.0040

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: 09 November 2021
doi: 10.3389/fonc.2021.725507

Geometric and Dosimetric Evaluation of Deep Learning-Based Automatic Delineation on CBCT-Synthesized CT and Planning CT for Breast Cancer Adaptive Radiotherapy: A Multi-Institutional Study

Zhenhui Dai[1†], Yiwen Zhang[2†], Lin Zhu[1], Junwen Tan[3], Geng Yang[1], Bailin Zhang[1], Chunya Cai[1], Huaizhi Jin[1], Haoyu Meng[1], Xiang Tan[1], Wanwei Jian[1], Wei Yang[2] and Xuetao Wang[1*]*

[1] Department of Radiation Therapy, The Second Affiliated Hospital, Guangzhou University of Chinese Medicine, Guangzhou, China, [2] School of Biomedical Engineering, Southern Medical University, Guangzhou, China, [3] Department of Oncology, The Fourth Affiliated Hospital, Guangxi Medical University, Liuzhou, China

OPEN ACCESS

Edited by:
Jose Eduardo Villarreal Barajas, Royal Devon and Exeter Hospital, United Kingdom

Reviewed by:
Sarah Quirk, University of Calgary, Canada
Xudong Xue, Hubei Cancer Hospital, China

***Correspondence:**
Wei Yang
weiyanggm@gmail.com
Xuetao Wang
wangxuetao0625@126.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: 15 June 2021
Accepted: 12 October 2021
Published: 09 November 2021

Citation:
Dai Z, Zhang Y, Zhu L, Tan J, Yang G, Zhang B, Cai C, Jin H, Meng H, Tan X, Jian W, Yang W and Wang X (2021) Geometric and Dosimetric Evaluation of Deep Learning-Based Automatic Delineation on CBCT-Synthesized CT and Planning CT for Breast Cancer Adaptive Radiotherapy: A Multi-Institutional Study. Front. Oncol. 11:725507. doi: 10.3389/fonc.2021.725507

Purpose: We developed a deep learning model to achieve automatic multitarget delineation on planning CT (pCT) and synthetic CT (sCT) images generated from cone-beam CT (CBCT) images. The geometric and dosimetric impact of the model was evaluated for breast cancer adaptive radiation therapy.

Methods: We retrospectively analyzed 1,127 patients treated with radiotherapy after breast-conserving surgery from two medical institutions. The CBCT images for patient setup acquired utilizing breath-hold guided by optical surface monitoring system were used to generate sCT with a generative adversarial network. Organs at risk (OARs), clinical target volume (CTV), and tumor bed (TB) were delineated automatically with a 3D U-Net model on pCT and sCT images. The geometric accuracy of the model was evaluated with metrics, including Dice similarity coefficient (DSC) and 95% Hausdorff distance (HD95). Dosimetric evaluation was performed by quick dose recalculation on sCT images relying on gamma analysis and dose-volume histogram (DVH) parameters. The relationship between ΔD95, ΔV95 and DSC-CTV was assessed to quantify the clinical impact of the geometric changes of CTV.

Results: The ranges of DSC and HD95 were 0.73–0.97 and 2.22–9.36 mm for pCT, 0.63–0.95 and 2.30–19.57 mm for sCT from institution A, 0.70–0.97 and 2.10–11.43 mm for pCT from institution B, respectively. The quality of sCT was excellent with an average mean absolute error (MAE) of 71.58 ± 8.78 HU. The mean gamma pass rate (3%/3 mm criterion) was 91.46 ± 4.63%. DSC-CTV down to 0.65 accounted for a variation of more than 6% of V95 and 3 Gy of D95. DSC-CTV up to 0.80 accounted for a variation of less than 4% of V95 and 2 Gy of D95. The mean ΔD90/ΔD95 of CTV and TB were less than

2Gy/4Gy, 4Gy/5Gy for all the patients. The cardiac dose difference in left breast cancer cases was larger than that in right breast cancer cases.

Conclusions: The accurate multitarget delineation is achievable on pCT and sCT *via* deep learning. The results show that dose distribution needs to be considered to evaluate the clinical impact of geometric variations during breast cancer radiotherapy.

Keywords: deep learning, automatic delineation, synthetic CT, dosimetric evaluation, adaptive radiotherapy

INTRODUCTION

Intensity-modulated radiotherapy (IMRT) after breast-conserving surgery significantly improves the survival of breast cancer patients (1). However, there are the patient setup error and anatomical structure changes during the interfractional radiotherapy (2). The variation range of mean central lung distance is 0.59–2.94 cm (3). The mean 3D displacement of patient setup is 7.3 and 7.6 mm by laser and port film setup, respectively (4). The deviation could lead to inconsistencies between the actual delivery dose and the planning dose (5). Large interfraction variation is observed, motivating the need for adaptive radiotherapy. Adaptive radiotherapy can automatically adjust the plan according to changes in the target volume (6, 7). When the patient is lying on the couch waiting for treatment, plan evaluation and adaptation need to be completed as quickly as possible. Online adaptation, which requires real-time delineation of the contours of the target volumes and organs at risk (OARs) for re-planning, is a promising technique (8). Some studies have been conducted for online adaptation, especially for prostate cancer as well as for head and neck cancer (9–13). Cone-beam CT (CBCT) is a common tool for location verification in radiotherapy and can be used for plan adaptation (14, 15). However, imaging artifacts caused by respiratory movement make CBCT-based adaptive radiotherapy for breast cancer infeasible. CBCT images cannot be directly used for dose calculation due to inaccurate HU values and needs to be converted into synthetic CT for dosimetric evaluation (16–19).

The delineation of target volumes and OARs is a prerequisite for adaptive radiotherapy. However, manual delineation is time-consuming and labor-intensive and cannot meet the requirements of real-time adaptive radiotherapy (20). It is necessary to build an automatic delineation model (21, 22). Some researchers used atlas-based segmentation software for delineation of target volumes on computed tomography (CT) images for radiotherapy. Dice score of segmentations with these commercial software is not high enough (23–26). CBCT-based delineation can be achieved by deformable image registration and direct delineation on CBCT images. Deformable image registration could transfer the contours to CBCT images from planning CT images (27). However, deformable registration relying on the image quality and algorithm cannot perform well for patients with large variations, leading to uncertainty in propagating contours (28). Direct CBCT-based delineation can reduce uncertainty from registration errors. Schreier et al. (29) investigated segmentation for the male pelvis using CBCT and CT images. Peroni et al. (11) developed an automatic strategy to generate online virtual CT and automatically segmented structures using CBCT and virtual CT images for head and neck cancer adaptive radiation therapy. Inter-observer variability is high in the delineation of target volumes and OARs on CT and CBCT scans of the chest. At present, most of the studies on the automatic segmentation of chest medical images do not perform well, and it is necessary to develop a model with better performance to delineate all the target volumes and OARs accurately at one time. Additionally, the geometric metrics do not fully indicate clinical quality. Therefore, it is necessary to evaluate the performance of the automatic delineation model in terms of clinical applicability (30). The geometric and dosimetric changes between planning CT (pCT) and synthetic CT (sCT) needs to evaluated due to high clinical significance for adaptive therapy.

In our study, we investigated the feasibility of automatically delineating multiple contours based on deep learning for breast cancer radiation therapy. The synthetic CT image was first generated from CBCT images with a cycle generative adversarial network (cycleGAN). Second, we developed an automatic delineation model using 3D U-net based on pCT and the radiotherapy structure of breast cancer patients to delineate the target volumes and OARs on planning CT and synthetic CT images, respectively. Third, the treatment plan was transferred to the synthetic CT image from the planning CT image. It could be verified quantitatively by quick dose recalculation for dosimetric evaluation. The Flowchart of the proposed method is shown as **Supplementary Figure 1** in the **Supplementary Material**. The clinical impact of geometric variations in target volumes and OARs was evaluated to provide the feasibility for breast cancer adaptive radiotherapy.

MATERIALS AND METHODS

Patient Datasets

Datasets obtained from two medical institutions in China between January 2014 and December 2020 were analyzed retrospectively. A total of 1,127 patients (institution A: 1,074/ institution B: 53) who received radiotherapy after breast-conserving surgery were included. The data of 75 patients from institution A including pCT and CBCT images with BH were split into 52 samples for training and 23 samples for testing on CBCT-to-CT synthesis. The pCT images and structures of 1,052 patients were randomly divided into a training set (700 patients),

validation set (100 patients), and test set (252 patients) for automatic delineation model training and evaluation. Among the test set, 199 patients were from institution A, and 53 patients were from institution B. The patient characteristics are summarized in **Table 1**.

Patients were immobilized on the breast bracket before obtaining a CT scan with 5 mm slices (SOMATOM Sensation Open, Siemens). The dimensions of the images were 512 × 512 voxels for each slice. CBCT is widely used for target position verification and setup error correction during breast patient radiation therapy. In this study, CBCT images were captured by a Varian Edge treatment machine (Varian Medical Systems, Palo Alto) in half-fan mode utilizing breath-hold (BH) guided by AlignRT (Vision RT Ltd, London, UK). The images with 512 × 512 voxels for each slice were reconstructed with a 1.99 mm slice thickness. The quality of CBCT with the BH technique is much higher than that with conventional scan. CBCT could be used to generate accurate synthetic CT for dose recalculation.

Clinical target volumes (CTV), tumor bed (TB), and organs at risk (OARs) were delineated by two radiation oncologists according to the ESTRO consensus guideline. The director with 20 years' experience of the radiation oncology department of the corresponding institution was consulted in cases of disagreement. Manually delineated contours were used as ground truth for training and testing. The radiotherapy plans were designed with the Pinnacle3 treatment planning system (Philips Radiation Oncology Systems, Philips Healthcare) by combining 3D conformal radiation therapy (CRT) and intensity modulated radiation therapy (IMRT). The prescribed dose was 52.2 Gy given to CTV and 63.8 Gy given to TB in 29 fractions. CRT achieved 80% of the total dose, and IMRT achieved 20% of the total dose. One hundred percent of the prescribed dose covered 95% of the volume of the target area. The physicist first designed the plan, and then the radiation oncologist and the physicist jointly evaluated the plan before implementation to ensure the quality of the plan.

TABLE 1 | Summary of patient characteristics.

Patient characteristics	Value
Age range (y)	22–72
Laterality	
Right	600
Left	527
Institution	
A	1,074
B	53
Stage, No.	
0	92
I	530
II	433
III	72

Automatic Delineation of Target Volumes and Organs at Risk

Our automatic delineation model is applied to 3D volume. Patch-based training makes the model cannot judge left and right lungs and target volumes. Therefore, through a series of preprocessing steps, as shown in **Figure 1A**, we feed the human body volume as completely as possible into the network, so that the network has a larger receptive field. First, the Hough transform line detection, the threshold method, and the morphological method were used to remove the bed and obtain a human body mask. Then, the minimum circumscribed cube is cropped from the body mask, which is the region of interest (ROI) of each subject. Finally, the spatial resolution of all ROIs is converted to 2 × 2 × 5 mm^3, and the intensity is normalized by z-score for training.

We use Res-SE-U-Net (31) as the automatic delineation network, which is a modified 3D U-Net (32). Res-SE-U-Net includes the down-sampling path, up-sampling path, and skip-connection layer, which can extract the multiscale features of images. In addition, the addition of Res-block and SE-block leads

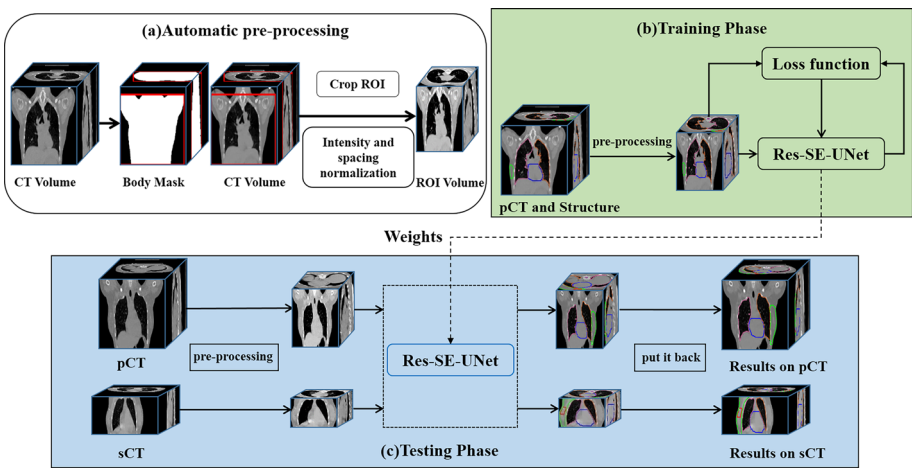


FIGURE 1 | The workflow of the automatic delineation model. **(A)** automatic preprocessing, **(B)** training phase, **(C)** testing phase.

to its stronger feature extraction ability than the original U-Net. The workflow of the automatic delineation model is shown in **Figure 1**. Training of a network took about 48 h, whereas all the target volumes and OARs for one patient were predicted in 5 s.

CBCT-to-CT Synthesis

We use a 2D cycleGAN (33), which is an unsupervised image-to-image translation deep learning framework, to generate sCT. The cycleGAN contains two generators ($G_{CBCT-CT}$ and $G_{CT-CBCT}$) and two discriminators (D_{CT} and D_{CBCT}). The generator $G_{CBCT-CT}$ takes CBCT as input and generates the sCT; in contrast, $G_{CT-CBCT}$ takes CT as input and generates the synthetic CBCT. The discriminator D_{CT} and D_{CBCT} discriminates whether the CT or CBCT images are real or synthesized, respectively. The sCT of one patient could be generated in 3–4 s by the trained model. The schematic flow is shown as **Supplementary Figure 2** in the **Supplementary Material**.

The optimization of the cycleGAN includes two objective functions: adversarial loss and cycle consistency loss. The antagonistic objectives of generators and discriminators are reflected in adversarial loss. We denote the data distribution as $I_{CT} \sim p_{\text{data}}(I_{CT})$ and $I_{CBCT} \sim p_{\text{data}}(I_{CBCT})$. The adversarial loss is expressed as:

$$\begin{aligned}\mathcal{L}_{GAN}(G_{CBCT-CT}, D_{CT}, I_{CT}, I_{CBCT}) = \mathbb{E}_{I_{CT} \sim p_{data}(I_{CT})}[\log D_{CT}(I_{CT})] \\ + \mathbb{E}_{I_{CBCT} \sim p_{data}(I_{CBCT})}[\log(1 - D_{CT}(G_{CBCT-CT}(I_{CBCT})))]\end{aligned} \quad (1)$$

where I_{CBCT} is the real CBCT and $G_{CBCT-CT}$ (I_{CBCT}) is the synthetic CT generated by $G_{CBCT-CT}$.

Geometric Evaluation

Automatic Delineation Performance Evaluation

The performance of the automatic delineation model was evaluated on pCT and sCT, respectively. To quantitatively assess the delineation accuracy, we used two metrics: Dice similarity coefficient (DSC) and 95% Hausdorff distance (HD95). DSC describes the spatial overlap between the automated delineation and the ground-truth. The metrics HD95 was used to evaluate the shape difference in the study. The equations are defined as **Supplementary Equations** (1, 2) in the **Supplementary Material**.

CBCT-Synthesized sCT Quality Evaluation

Mean error (ME), mean absolute error (MAE), peak signal to noise ratio (PSNR), structural similarity index (SSIM), and spatial non-uniformity (SNU) were used to evaluate the image quality of sCT and CBCT, respectively. The formulas for these metrics are defined as **Supplementary Equations (4–8)** in the **Supplementary Material**. We selected five regions of interest (ROIs) to calculate the SNU, as shown in **Supplementary Figure 3** of the **Supplementary Material**. ME and MAE are the magnitudes of the difference between the pCT and the sCT. The lower these values are, the better the image quality is. High PSNR and SSIM mean high image quality. In this study, deformable registration was performed on the sCT to align it with pCT, and the metrics were calculated within the body mask of the sCT.

Dosimetric Evaluation

Dosimetric accuracy was evaluated based on the sCT images using clinical breast cancer treatment plans. A quick dose recalculation on the sCT images was performed to verify the treatment plan. The treatment plan that was transferred to the sCT from the pCT kept the same parameters as the original pCT-based plan. The difference in dose distribution between pCT and sCT was evaluated with gamma analysis and dose-volume histogram (DVH) parameters. The difference in DVH metrics of target volumes and OARs between pCT and sCT were also assessed for quantitative dosimetric evaluation. The DVH metrics of the target volumes, including D90, D95, and V95, were analyzed. Target coverage was defined as the dose received by 90 and 95% of the target volume (D90, D95) and the percent volume receiving 95% of the prescribed dose (V95) for the TB and CTV. If the dose difference in the target volumes and OARs on the sCT exceeds the threshold, it needs to rescan the pCT for re-planning. $\Delta D90$, $\Delta D95$, and $\Delta V95$ are defined as:

$$\Delta D90 = |D90(pCT) - D90(sCT)| \quad (2)$$

$$\Delta D95 = |D95(pCT) - D95(sCT)| \quad (3)$$

$$\Delta V95 = |V95(pCT) - V95(sCT)| \quad (4)$$

We investigated the relationship between the DSC and the dose difference to evaluate the effect of anatomical changes on dose during radiotherapy. sCT images were rigidly registered to pCT by reference to the bony landmarks. DSC-CTV for automatically delineating CTV on sCT images compared with manually delineating target volumes on pCT images was correlated with dosimetric metrics.

Clinical Evaluation

The reproducibility and robustness of the automatic delineation model were evaluated using DSC and HD95 in a multi-institution study. The training set was from institution A. Of the 252 patients in the test set, 199 were from institution A and 53 were from institution B. The robustness of the model was validated by multi-institutional testing.

Clinical evaluation of the automatic delineation model was performed on an independent test set of 199 pCT scans and 23 sCT from institution A. The automatically delineating contours were checked by three groups a, b, and c in institution A based on their clinical experience. Each group consisted of two radiation oncologists. Three groups who were blinded to the ground truth contours reviewed the automatic delineating contours. The evaluation results were acceptable with no corrections, acceptable with minor corrections, and unacceptable respectively. Finally, the acceptable ratio of all the targets and OARs was analyzed.

Statistical Analysis

MATLAB (version 2018b, The MathWorks Inc, Natick, MA, USA) was used for statistical analysis. A t test was used to test the statistical significance of the absolute difference of the dosimetric metrics for both plans, and a Spearman's rank correlation test

was used for correlation testing between DSC and ΔV95 or ΔD95 of CTV. Two-sided p-values were provided, and p-values <0.05 were considered significant.

RESULTS

Geometric Evaluation

The ME, MAE, PSNR, and SSIM comparisons between the CBCT, synthetic CT (sCT), and planning CT (pCT) images are shown in **Table 2**. The average ME and MAE between CBCT and pCT images within the body was −37.71 ± 15.49 and 86.42 ± 10.12 HU, whereas the average ME and MAE between sCT and pCT images was 8.46 ± 11.88 and 71.58 ± 8.78 HU. The mean SNUs for CBCT, sCT, and pCT were 9.22 ± 3.89, 4.95 ± 4.13, 2.12 ± 0.85%, respectively. The HU value of sCT image is much closer to that of pCT image than that of CBCT image. The similarity increased obviously between sCT and pCT images with lower ME, MAE and higher PSNR, SSIM compared to CBCT and pCT images. The detailed comparison between pCT and sCT is shown as **Supplementary Figure 3** in the **Supplementary Material**.

Figure 2 shows examples of the ground truth and the contours of automatic delineation on pCT and sCT images. There is good consistency for CTV and OARs between automatic delineation and manual delineation from human experts on pCT. The concordance can be found to decrease at the upper and lower bounds of the CTV from 2D sagittal sCT images. The automatically drawn tumor bed (TB) on sCT is obviously larger than the manually drawn TB.

The testing results of automatic delineation of multiple institutions are shown in **Table 3**. The results of pCT were calculated among 199 patients from institution A and 53 patients from institution B, respectively. The results of sCT were calculated among 23 patients from institution A for CBCT-to-CT synthesis testing cohorts. Good DSC and HD95 scores were found for the most contours on pCT (DSC: 0.73–0.97, HD95: 2.22–9.36 mm). The performance was slightly lower for the contours on sCT from institution A (DSC: 0.63–0.95, HD95: 2.30–19.57 mm). The mean DSC of CTV was 0.88 ± 0.03 for pCT, and 0.83 ± 0.03 for sCT, respectively. The segmentation model was also effective for pCT from institution B (DSC: 0.70–0.97, HD95: 2.10–11.43 mm). The mean DSC of CTV on pCT from institution B was 0.80 ± 0.06. The accuracy of automatic delineation for the datasets from institution B was lower than that from institution A.

The clinical evaluation showed that the acceptable ratios of OARs, CTV, and TB were 76.38–100, 70.35–83.92, and 53.27–57.79% on pCT images and 73.91–82.61, 69.57–78.26, and 43.48–52.17% on sCT images, respectively, as shown in **Table 4**. Overall, the automatic delineation of CTV and OARs was clinically acceptable after minor corrections by the evaluation of medical group a, b, and c.

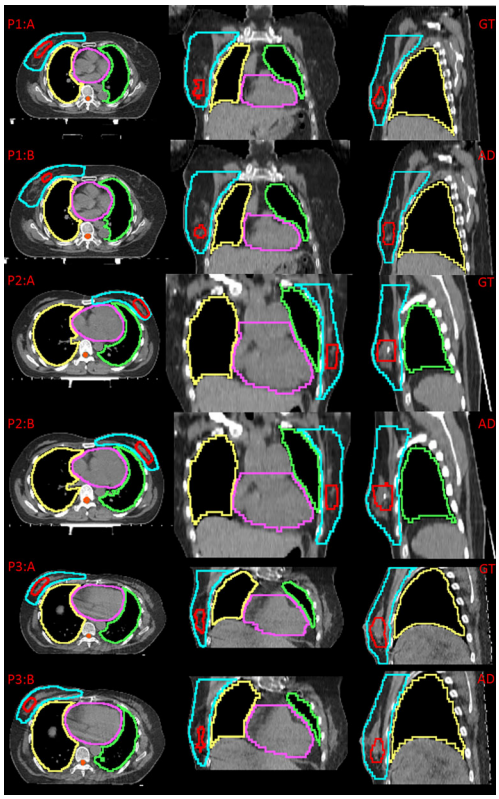


FIGURE 2 | Comparison between the ground truth (GT) and automatic delineation (AD) at axial plane, coronal plane, sagittal plane, P1:A indicates GT on pCT for patient 1, P1:B indicates AD on pCT for patient 1; P2:A indicates GT on pCT for patient 2, P2:B indicates AD on pCT for patient 2; P3:A indicates GT on sCT for patient 3, P3:B indicates AD on sCT for patient 3.

Dosimetric Evaluation

The dose distribution and dose-volume histogram (DVH) of the plan on synthetic CT (sCT) and planning CT (pCT) are shown in **Figure 3**. P1:A and P2:A indicate the dose distribution on pCT, and P1:B and P2:B indicate the dose distribution on sCT. The red line represents TB, and the blue shaded area represents the dose of 63.8 Gy. The wathet line represents CTV, and the yellow

TABLE 2 | Similarity analysis between CBCT and pCT images, sCT and pCT images with all the testing patient datasets.

Type	ME (HU)	MAE (HU)	PSNR (dB)	SSIM	SNU (%)
CBCT *vs* pCT	−37.71 ± 15.49	86.42 ± 10.12	20.19 ± 5.26	0.88 ± 0.04	9.22 ± 3.89 *vs* 2.12 ± 0.85
sCT *vs* pCT	8.46 ± 11.88	71.58 ± 8.78	23.34 ± 3.63	0.92 ± 0.02	4.95 ± 4.13 *vs* 2.12 ± 0.85

TABLE 3 | Quantitative results (Mean ± SD) of automatic delineation performance in multiple institutions.

Institution	Image	Metrics	TB	CTV	Heart	Left lung	Right lung	Spinal cord
A	pCT	DSC	0.73 ± 0.08	0.88 ± 0.03	0.93 ± 0.06	0.97 ± 0.01	0.97 ± 0.01	0.82 ± 0.05
		HD95/mm	9.36 ± 4.80	9.13 ± 4.04	7.63 ± 5.60	2.22 ± 1.37	2.59 ± 2.24	5.12 ± 5.90
	sCT	DSC	0.63 ± 0.08	0.83 ± 0.03	0.90 ± 0.02	0.94 ± 0.01	0.95 ± 0.01	0.81 ± 0.03
		HD95/mm	19.57 ± 17.01	10.81 ± 4.81	9.31 ± 2.60	5.75 ± 1.61	5.62 ± 1.72	2.30 ± 0.32
B	pCT	DSC	0.70 ± 0.09	0.80 ± 0.06	0.92 ± 0.02	0.97 ± 0.01	0.97 ± 0.00	0.73 ± 0.07
		HD95/mm	11.43 ± 6.17	18.22 ± 6.94	8.51 ± 3.71	2.10 ± 0.54	2.45 ± 1.00	8.38 ± 5.90

TABLE 4 | Acceptable ratio for automatic delineation among different groups.

Image	Group	Acceptable ratio/%					
		CTV	TB	Left lung	Right lung	Heart	Spinal cord
pCT	a	76.88	57.79	100	98.99	90.95	81.91
	b	83.92	54.77	100	98.49	76.88	79.40
	c	70.35	53.27	95.98	96.48	88.44	76.38
sCT	a	73.91	52.17	95.65	100	86.96	82.61
	b	78.26	47.83	91.30	95.65	73.91	78.26
	c	69.57	43.48	86.96	82.61	82.61	73.91

shaded area represents the dose of 52.2 Gy. P1:C and P2:C indicate the DVH of the two plans. The solid line represents the DVH on pCT, and the dotted line represents the DVH on sCT.

Within the body, the mean ± standard deviation with 2%/2 mm and 3%/3 mm pass rates for the sCT images were 85.09 ± 6.28 and 91.46 ± 4.63% respectively. There was a negative correlation between DSC and ΔV95 or ΔD95 (r= −0.52, p= 2.4075E-11 and r= −0.51, p= 4.5815E-11, respectively). DSC of CTV down to 0.65 accounted for a variation of more than 6% of V95 and 3 Gy of D95 for CTV. DSC of CTV up to 0.80 accounted for a variation of less than 4% of V95 and 2 Gy of D95 for CTV, as shown in **Figure 4**.

Dosimetry evaluation of the plans on sCT *versus* original plans for 29 fractions is summarized in **Table 5**. The mean ΔV95 of CTV is less than 6%. The mean ΔD90/ΔD95 of CTV is less than 2 Gy/4 Gy. The mean ΔD90/ΔD95 of tumor bed was less than 4 Gy/5 Gy for all patients. We observed an absolute difference of more than 8% of the ΔV95 of TB. The poor delineation accuracy of TB leads to large dosimetry errors. The mean ΔD95 of CTV and ΔV10 of the heart are 4.20 ± 1.45 Gy and 3.92 ± 3.29% in the left-sided patients, 2.84 ± 0.84 Gy and 1.60 ± 1.96% in the right-sided patients, respectively. The dosimetric difference of target volume and heart in left-sided patients is greater than that in right-sided patients. P-values of the dosimetric difference of the TB and CTV were below 0.05, and p-values of the dosimetric difference of the OAR were over 0.05.

DISCUSSION

Adaptive radiotherapy based on CBCT for patient setup is a promising approach for improving treatment accuracy (34). Liu et al. (35) developed a deep learning approach to generate CBCT-based synthetic CT images and validated the dose calculation accuracy for clinical use in CBCT-guided pancreatic adaptive radiotherapy. However, sCT-based segmentation was not involved, which was also the key factor in adaptive radiotherapy. In our study, the synthetic CT image quality analysis of the thorax yielded an ME/MAE of 8.46/71.58 HU, and our results were better than those reported by Eckl et al. (36), with 29.6/94.2 HU. The SNU in synthetic CT was close to the SNU in planning CT. It demonstrated that synthetic CT had enough quality for contour delineation and dose calculation. The automatic delineation model that was developed by 3D Res-SE-U-Net based on the planning CT and RT structures made full use of the 3D correlative information between image slices. The automatic delineation model performs well enough with 0.88 DSC for CTV on planning CT. The performance of the model was lower for synthetic CT because of the inaccurate HU and image artifacts caused by respiratory in synthetic CT. The clinically meaningful evaluation of the performance of the model should include not only geometric difference but also dosimetric assessment (37, 38). In this study, the geometric and dosimetric differences between the contours on planning CT and the contours on synthetic CT were analyzed to assess the clinical impact of the changes in target volumes and OARs during radiotherapy. The interobserver variability is large for the contours of breast cancer, resulting in difficulty delineating exactly (39). We improved the robustness of the model by enlarging the datasets to over 1,000. The automatic delineation of CTV and OARs was more consistent with the manual delineation due to their regular shapes and locations. Tumor bed was significantly different between automatic and manual contours because the position of the tumor bed varied greatly for each patient. The results of the multi-institutional test showed that the model is robust and accurate. The performance of the model on the datasets from institution B was worse than that from institution A. Our approach was effective in dosimetric verification based on synthetic CT from CBCT, and ΔV95 and ΔD95 of CTV could be used as dosimetric metrics for rescanning

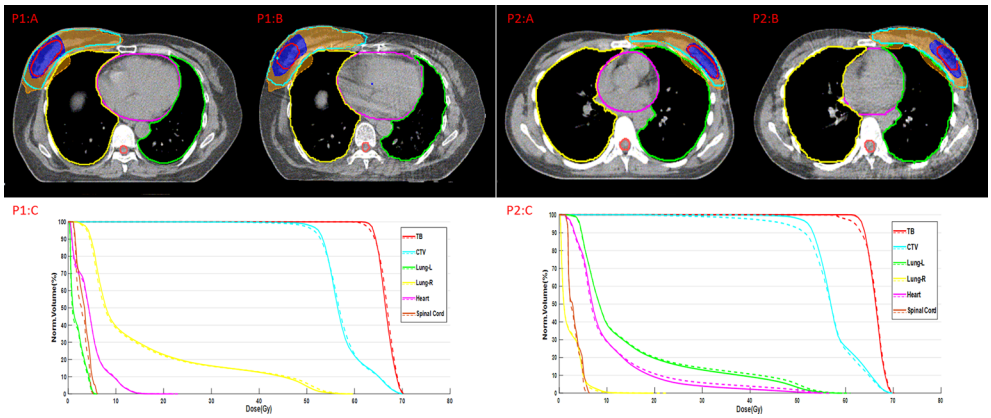


FIGURE 3 | Dose distribution and DVH of the plans on pCT and sCT.

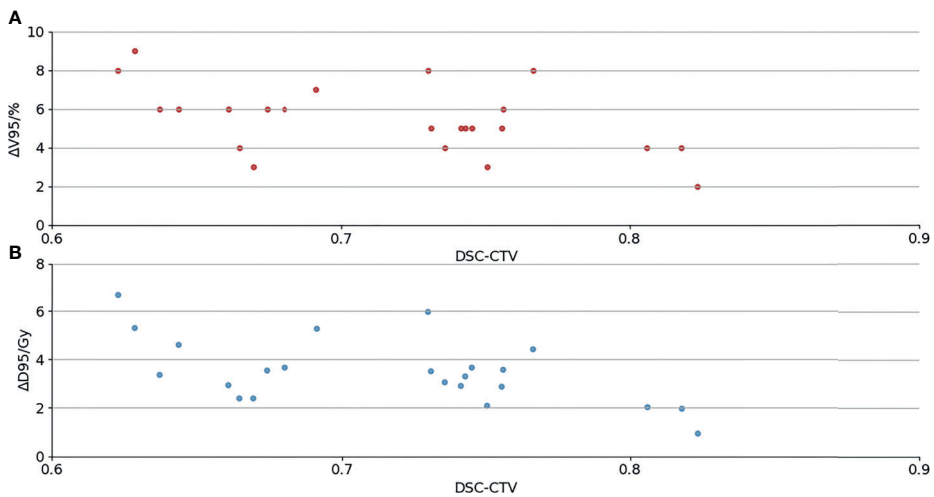


FIGURE 4 | **(A)** The relationship between DSC and ΔV95 of CTV; **(B)** the relationship between DSC and ΔD95 of CTV.

TABLE 5 | Absolute difference of the dosimetric metrics for both plans of 23 patients.

Structure	Metrics	Absolute difference (Mean ± SD)			p-value
		Right breast	Left breast	All patients	
Spinal Cord	ΔDmax[Gy]	0.35 ± 0.28	0.34 ± 0.57	0.34 ± 0.45	0.63
Ipsilateral Lung	ΔV20[%]	1.80 ± 1.55	1.83 ± 1.59	1.82 ± 1.53	0.87
	ΔDmean[Gy]	0.83 ± 0.75	1.03 ± 1.02	0.94 ± 0.89	0.34
Heart	ΔV10[%]	1.60 ± 1.96	3.92 ± 3.29	2.86 ± 2.95	0.91
	ΔV30[%]	0.10 ± 0.32	2.58 ± 3.00	1.45 ± 2.52	0.56
	ΔDmean[Gy]	0.19 ± 0.19	1.72 ± 2.17	1.03 ± 1.76	0.96
TB	ΔD90[Gy]	2.65 ± 1.35	4.33 ± 2.54	3.57 ± 2.21	8.18E-09
	ΔD95[Gy]	3.73 ± 1.62	5.58 ± 2.50	4.74 ± 2.30	2.62E-07
	ΔV95[%]	5.70 ± 3.56	10.25 ± 7.03	8.18 ± 6.06	1.32E-07
CTV	ΔD90[Gy]	1.68 ± 0.57	1.89 ± 0.65	1.80 ± 0.61	2.61E-13
	ΔD95[Gy]	2.84 ± 0.84	4.20 ± 1.45	3.58 ± 1.37	2.30E-12
	ΔV95[%]	4.70 ± 1.25	6.17 ± 1.95	5.50 ± 1.79	1.01E-15

pCT. There was a correlation between DSC and ΔD95 and ΔV95 for CTV; however, it was not an inversely proportional relationship. Poor DSC scores do not necessarily lead to inferior CTV dosimetry. The dose variation between the automatic delineation CTV on sCT images and the manual delineation CTV on planning CT images was not large if the DSC value was low due to the automatic delineation being too small. We evaluated the absolute difference of the dosimetric metrics for both plans. Geometric changes on synthetic CT have a greater impact on the cardiac dose difference in left breast cancer, and special attention needs to be paid to assess the cardiac dose for left breast cancer. ΔV95, ΔD95, and ΔD90 could be used as evaluation indicators for whether to re-plan.

Additional limitations include the following: (1) Deformable registration from planning CT to CBCT was performed because of the different slice thicknesses and scanned areas between the two images. Although deformable registration was used, it was difficult to align the anatomical structure in CBCT with the same structure in planning CT. The image quality of synthetic CT could be degraded due to the registration errors, which affect the delineation accuracy of the model to a certain extent. (2) DSC-CTV was computed by rigid registration between synthetic CT and planning CT. However, the limited registration accuracy could cause certain dosimetric uncertainties in CTV. (3) The synthetic CT was generated from any day's CBCT, not entire treatment course. The anatomical changes and dosimetric difference were not evaluated during entire treatment delivery course.

The ranges of clinically acceptable ratio for CTV delineation are between 70% and 83% among the different groups, showing no common objective evaluation of the delineation. Variability exists between observer groups, demonstrating that the difference between automatic and manual delineation depends not only on contouring routines and guidelines, but also on personal preference. In the future, we hope to develop a universal model that can not only meet the quality requirements of multiple clinical institutions, but also adapt to the personal preferences of each observer. The automatically generated contours could be carefully reviewed by the radiation oncologist and used for treatment planning.

CONCLUSION

This study demonstrated that the developed approaches are capable of reliably generating target and OAR contours on pCT and daily sCT images from CBCT images, which could greatly accelerate the re-planning process and meet the requirements of online plan adaptation. The automatic delineation model performed sufficiently well for most patients. The geometric and dosimetric differences between pCT and sCT images in fractional radiotherapy need to be evaluated due to the high clinical significance for breast cancer adaptive radiotherapy.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the medical ethics committee of the Second Affiliated Hospital of Guangzhou University of Chinese Medicine. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

ZD and YZ: design of methodology, development and implement of models, original drafting. LZ: data curation and preprocessing. GY, BZ, CC, and WJ: experimental results analysis, draft reviewing. JT, HJ, HM, and XT: data collection. WY and XW: design of methodology, review and editing. All authors contributed to the article and approved the submitted version.

FUNDING

This work is supported in part by the National Natural Science Foundation of China (81771916) and Guangzhou Science and Technology Plan (202102010264).

ACKNOWLEDGMENTS

The authors sincerely thank Dr. Jing Cai of the Hong Kong Polytechnic University for his draft reviewing.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2021.725507/full#supplementary-material

REFERENCES

1. Macchia G, Cilla S, Buwenge M, Zamagni A, Ammendolia I, Zamagni C, et al. Intensity-Modulated Radiotherapy With Concomitant Boost After Breast Conserving Surgery: A Phase I–II Trial. *Breast Cancer: Targets Ther* (2020) 12:243–9. doi: 10.2147/BCTT.S261587
2. Saliou MG, Giraud P, Simon L, Fournier-Bidoz N, Fourquet A, Dendale R, et al. Radiotherapy for Breast Cancer: Respiratory and Set-Up Uncertainties. *Cancer Radiother* (2005) 9:414–21. doi: 10.1016/j.canrad.2005.09.003
3. Smith RP, Bloch P, Harris EE, McDonough J, Sarkar A, Kassaee A, et al. Analysis of Interfraction and Intrafraction Variation During Tangential

Breast Irradiation With an Electronic Portal Imaging Device. *Int J Radiat Oncol Biol Phys* (2005) 62(2):373–8. doi: 10.1016/j.ijrobp.2004.10.022
4. Bert C, Metheany KG, Doppke KP, Taghian AG, Powell SN, Chen GT. Clinical Experience With a 3D Surface Patient Setup System for Alignment of Partial-Breast Irradiation Patients. *Int J Radiat Oncol Biol Phys* (2006) 64 (4):1265–74. doi: 10.1016/j.ijrobp.2005.11.008
5. Baroni G, Garibaldi C, Scabini M, Riboldi M, Catalano G, Tosi G, et al. Dosimetric Effects Within Target and Organs at Risk of Interfractional Patient Mispositioning in Left Breast Cancer Radiotherapy. *Int J Radiat Oncol Biol Phys* (2004) 59(3):861–71. doi: 10.1016/j.ijrobp.2004.02.043
6. Zegers CML, Baeza JA, van Elmpt W, Murrer LHP, Verhoeven K, Boersma L, et al. Three-Dimensional Dose Evaluation in Breast Cancer Patients to Define Decision Criteria for Adaptive Radiotherapy. *Acta Oncol* (2017) 56(11):1487–94. doi: 10.1080/0284186X.2017.1349334
7. Hurkmans CW, Dijckmans I, Reijnen M, van der Leer J, van Vliet-Vroegindeweij C, van der Sangen M. Adaptive Radiation Therapy for Breast IMRT Simultaneously Integrated Boost: Three-Year Clinical Experience. *Radiother Oncol* (2012) 103:183–7. doi: 10.1016/j.radonc.2011.12.014
8. Lim-Reinders S, Keller BM, Al-Ward S, Sahgal A, Kim A. Online Adaptive Radiation Therapy. *Int J Radiat Oncol Biol Phys* (2017) 99(4):994–1003. doi: 10.1016/j.ijrobp.2017.04.023
9. Zhang J, Ahunbay E, Li XA. Technical NoteAcceleration of Online Adaptive Replanning With Automation and Parallel Operations. *Med Phys* (2018) 45 (10):4370–6. doi: 10.1002/mp.13106
10. Liang X, Zhao W, Hristov D, Buyyounouski MK, Hancock SL, Bagshaw H, et al. A Deep Learning Framework for Prostate Localization in Cone Beam CT Guided Radiotherapy. *Med Phys* (2020) 47(9):4233–40. doi: 10.1002/mp.14355
11. Peroni M, Ciardo D, Spadea MF, Riboldi M, Comi S, Alterio D, et al. Automatic Segmentation and Online virtual CT in Head-And-Neck Adaptive Radiation Therapy. *Int J Radiat Oncol Biol Phys* (2012) 84(3):427–33. doi: 10.1016/j.ijrobp.2012.04.003
12. Peng Y, Chen S, Qin A, Chen M, Gao X, Liu Y, et al. Magnetic Resonance-Based Synthetic Computed Tomography Images Generated Using Generative Adversarial Networks for Nasopharyngeal Carcinoma Radiotherapy Treatment Planning. *Radiother Oncol* (2020) 150:217–24. doi: 10.1016/j.radonc.2020.06.049
13. Dai X, Lei Y, Wang T, Dhabaan AH, McDonald M, Beitler JJ, et al. Head-And-Neck Organs-at-Risk Auto-Delineation Using Dual Pyramid Networks for CBCT-Guided Adaptive Radiotherapy. *Phys Med Biol* (2021) 66(4):045021. doi: 10.1088/1361-6560/abd953
14. Stankiewicz M, Li W, Rosewall T, Tadic T, Dickie C, Velec M. A Patterns of Practice of Adaptive Re-Planning for Anatomic Variances During Cone-Beam CT Guided Radiotherapy. *Tech Innovations Patient Support Radiat Oncol* (2019) 12:50–5. doi: 10.1016/j.tipsro.2019.10.003
15. Qin Y, Zhang F, Yoo DS, Kelsey CR, Yin FF, Cai J, et al. Adaptive Stereotactic Body Radiation Therapy Planning for Lung Cancer. *Int J Radiat Oncol Biol Phys* (2013) 87(1):209–15. doi: 10.1016/j.ijrobp.2013.05.008
16. Almatani T, Hugtenburg RP, Lewis RD, Barley SE, Edwards MA. Automated Algorithm for CBCT-Based Dose Calculations of Prostate Radiotherapy With Bilateral Hip Prostheses. *Br J Radiol* (2016) 89:20160443. doi: 10.1259/bjr.20160443
17. Li Y, Zhu J, Liu Z, Teng J, Xie Q, Zhang L, et al. A Preliminary Study of Using a Deep Convolution Neural Network to Generate Synthesized CT Images Based on CBCT for Adaptive Radiotherapy of Nasopharyngeal Carcinoma. *Phys Med Biol* (2019) 64:145010. doi: 10.1088/1361-6560/ab2770
18. Maspero M, Houweling AC, Savenije MHF, van Heijst TCF, Verhoeff JJC, Kotte ANTJ, et al. A Single Neural Network for Cone-Beam Computed Tomography-Based Radiotherapy of Head-and-Neck, Lung and Breast Cancer. *Phys Imaging Radiat Oncol* (2020) 14:24–31. doi: 10.1016/j.phro.2020.04.002
19. Sun H, Fan R, Li C, Lu Z, Xie K, Ni X, et al. Imaging Study of Pseudo-CT Synthesized From Cone-Beam CT Based on 3D CycleGAN in Radiotherapy. *Front Oncol* (2021) 11:603844. doi: 10.3389/fonc.2021.603844
20. van Dijk LV, Van den Bosch L, Aljabar P, Peressutti D, Both S, Steenbakkers JHM, 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
21. Eldesoky AR, Francolini G, Thomsen MS, Yates ES, Nyeng TB, Kirkove C, et al. Dosimetric Assessment of an Atlas Based Automated Segmentation for Loco-Regional Radiation Therapy of Early Breast Cancer in the Skagen Trial 1: A Multi-Institutional Study. *Clin Trans Radiat Oncol* (2017) 2:36–40. doi: 10.1016/j.ctro.2017.01.004
22. Simões R, Wortel G, Wiersma TG, Janssen TM, van der Heide UA, Remeijer P. Geometrical and Dosimetric Evaluation of Breast Target Volume Autocontouring. *Phys Imaging Radiat Oncol* (2019) 12:38–43. doi: 10.1016/j.phro.2019.11.003
23. Anders LC, Stieler F, Siebenlist K, Schäfer J, Lohr F, Wenz F. Performance of an Atlas-Based Autosegmentation Software for Delineation of Target Volumes for Radiotherapy of Breast and Anorectal Cancer. *Radiother Oncol* (2012) 102:68–73. doi: 10.1016/j.radonc.2011.08.043
24. 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) 7:160. doi: 10.1186/1748-717X-7-160
25. Kaderka R, Gillespie EF, Mundt RC, Bryant AK, Sanudo-Thomas CB, Harrison AL, et al. Geometric and Dosimetric Evaluation of Atlas Based Auto-Segmentation of Cardiac Structures in Breast Cancer Patients. *Radiother Oncol* (2019) 131:215–20. doi: 10.1016/j.radonc.2018.07.013
26. Frederick A, Roumeliotis M, Grendarova P, Craighead P, Abedin T, Watt E, et al. A Framework for Clinical Validation of Automatic Contour Propagation: Standardizing Geometric and Dosimetric Evaluation. *Pract Radiat Oncol* (2019) 9:448–55. doi: 10.1016/j.prro.2019.06.017
27. Hou J, Guerrero M, Chen W, D'Souza WD. Deformable Planning Ct to Cone-Beam Ct Image Registration in Head-and-Neck Cancer. *Med Phys* (2011) 38 (4):2088–94. doi: 10.1118/1.3554647
28. Wu RY, Liu AY, Williamson TD, Yang J, Wisdom PG, Zhu XR, et al. Quantifying the Accuracy of Deformable Image Registration for Cone-Beam Computed Tomography With a Physical Phantom. *J Appl Clin Med Phys* (2019) 20(10):92–100. doi: 10.1002/acm2.12717
29. Schreier J, Genghi A, Laaksonen H, Morgas T, Haas B. Clinical Evaluation of a Full-Image Deep Segmentation Algorithm for the Male Pelvis on Cone-Beam CT and CT. *Radiother Oncol* (2020) 145:1–6. doi: 10.1016/j.radonc.2019.11.021
30. van Rooij W, Dahele M, Ribeiro Brandao H, Delaney AR, Slotman BJ, Verbakel WF. Deep Learning-Based Delineation of Head and Neck Organs at Risk: Geometric and Dosimetric Evaluation. *Int J Radiat Oncol Biol Phys* (2019) 104(3):677–84. MBChB. doi: 10.1016/j.ijrobp.2019.02.040
31. Wang X, Yang G, Zhang Y, Zhu L, Xue X, Zhang B, et al. Automated Delineation of Nasopharynx Gross Tumor Volume for Nasopharyngeal Carcinoma by Plain CT Combining Contrast-Enhanced CT Using Deep Learning. *J Radiat Res Appl Sci* (2020) 13(1):568–77. doi: 10.1080/16878507.2020.1795565
32. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: *Presented at the Int Conf* Munich, Germany: Medical Image Computing and Computer-Assisted Intervention (2015). p. 234–41.
33. Zhu J-Y, Park T, Isola P, Efros A. *Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks*. Venice, Italy: IEEE International Conference on Computer Vision (ICCV) (2017).
34. Kurz C, Nijhuis R, Reiner M, Ganswindt U, Thieke C, Belka C, et al. Feasibility of Automated Proton Therapy Plan Adaptation for Head and Neck Tumors Using Cone Beam Ct Images. *Radiat Oncol* (2016) 11:64. doi: 10.1186/s13014-016-0641-7
35. Liu Y, Lei Y, Wang T, Fu Y, Tang X, Curran WJ, et al. CBCT-Based Synthetic CT Generation Using Deep-Attention cycleGAN for Pancreatic Adaptive Radiotherapy. *Med Phys* (2020) 47(6):2472–83. doi: 10.1002/mp.14121
36. Eckl M, Hoppen L, Sarria GR, Boda-Heggemann J, Simeonova-Chergou A, Steil V, et al. Evaluation of a Cycle-Generative Adversarial Network-Based Cone-Beam CT to Synthetic CT Conversion Algorithm for Adaptive Radiation Therapy. *Physica Med* (2020) 80:308–16. doi: 10.1016/j.ejmp.2020.11.007
37. Tsuji SY, Hwang A, Weinberg V, Yom SS, Quivey JM, Xia P. Dosimetric Evaluation of Automatic Segmentation for Adaptive IMRT for Head-and-Neck Cancer. *Int J Radiat Oncol Biol Phys* (2010) 77(3):707–14. doi: 10.1016/j.ijrobp.2009.06.012
38. Cao M, Stiehl B, Yu V, Sheng K, Kishan AU, Chin RK, et al. Analysis of Geometric Performance and Dosimetric Impact of Using Automatic Contour Segmentation for Radiotherapy Planning. *Front Oncol* (2020) 10:1762. doi: 10.3389/fonc.2020.01762

39. Men K, Zhang T, Chen X, Chen B, Tang Y, Wang S, et al. Fully Automatic and Robust Segmentation of the Clinical Target Volume for Radiotherapy of Breast Cancer Using Big Data and Deep Learning. *Physica Med* (2018) 50:13–9. doi: 10.1016/j.ejmp.2018.05.006

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: 16 December 2021
doi: 10.3389/fonc.2021.789014

Evaluation of PD-L1 Expression Level in Patients With Non-Small Cell Lung Cancer by ^{18}F-FDG PET/CT Radiomics and Clinicopathological Characteristics

Jihui Li[1†], Shushan Ge[1†], Shibiao Sang[1†], Chunhong Hu[2] and Shengming Deng[1,3,4*]*

[1] Department of Nuclear Medicine, The First Affiliated Hospital of Soochow University, Suzhou, China, [2] Department of Radiology, The First Affiliated Hospital of Soochow University, Suzhou, China, [3] Department of Nuclear Medicine, Suqian First Hospital, Suqian, China, [4] State Key Laboratory of Radiation Medicine and Protection, Soochow University, Suzhou, China

OPEN ACCESS

Edited by:
Savino Cilla,
Gemalli Molise Hospital, Italy

Reviewed by:
Liyuan Zhang,
Second Affiliated Hospital of Soochow University, China
Joon Young Choi,
Sungkyunkwan University, South Korea

***Correspondence:**
Shengming Deng
dshming@163.com
Chunhong Hu
sdhuchunhong@sina.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: *04 October 2021*
Accepted: *30 November 2021*
Published: *16 December 2021*

Citation:
Li J, Ge S, Sang S, Hu C and Deng S (2021) Evaluation of PD-L1 Expression Level in Patients With Non-Small Cell Lung Cancer by ^{18}F-FDG PET/CT Radiomics and Clinicopathological Characteristics. Front. Oncol. 11:789014. doi: 10.3389/fonc.2021.789014

Purpose: In the present study, we aimed to evaluate the expression of programmed death-ligand 1 (PD-L1) in patients with non-small cell lung cancer (NSCLC) by radiomic features of ^{18}F-FDG PET/CT and clinicopathological characteristics.

Methods: A total 255 NSCLC patients (training cohort: n = 170; validation cohort: n = 85) were retrospectively enrolled in the present study. A total of 80 radiomic features were extracted from pretreatment ^{18}F-FDG PET/CT images. Clinicopathologic features were compared between the two cohorts. The least absolute shrinkage and selection operator (LASSO) regression was used to select the most useful prognostic features in the training cohort. Radiomics signature and clinicopathologic risk factors were incorporated to develop a prediction model by using multivariable logistic regression analysis. The receiver operating characteristic (ROC) curve was used to assess the prognostic factors.

Results: A total of 80 radiomic features were extracted in the training dataset. In the univariate analysis, the expression of PD-L1 in lung tumors was significantly correlated with the radiomic signature, histologic type, Ki-67, SUV_{max}, MTV, and TLG ($p < 0.05$, respectively). However, the expression of PD-L1 was not correlated with age, TNM stage, and history of smoking ($p > 0.05$). Moreover, the prediction model for PD-L1 expression level over 1% and 50% that combined the radiomic signature and clinicopathologic features resulted in an area under the curve (AUC) of 0.762 and 0.814, respectively.

Conclusions: A prediction model based on PET/CT images and clinicopathological characteristics provided a novel strategy for clinicians to screen the NSCLC patients who could benefit from the anti-PD-L1 immunotherapy.

Keywords: non-small cell lung cancer, PD-L1 immunotherapy, radiomics, ^{18}F-FDG PET/CT, clinicopathological

INTRODUCTION

Lung cancer is a malignant tumor with the highest morbidity and mortality in the world, and its average 5-year survival rate is only 15% (1). Non-small cell lung cancer (NSCLC) accounts for 80% ~ 85% of all lung cancer cases (2). Early diagnosis and treatment play a key role in improving the 5-year survival rate of lung cancer. In recent years, with the further study of tumor immune microenvironment, immunotherapy has developed rapidly, attracted more and more oncologists' attention, and become an important research field of tumor therapy, including lung cancer (3). The immunotherapy against programmed cell death protein 1 (programmed death-1, PD-1) and its ligand 1 (programmed death ligand-1, PD-L1) has been used in NSCLC, and good results have been achieved in patients, especially in individuals with high expression of PD-L1 (4, 5).

At present, the expression of PD-L1 in clinical practice is usually detected through the "gold standard" of immunohistochemistry (IHC) (6, 7). It is difficult to obtain a clear expression level of PD-L1 in high-quality tissue samples. Moreover, small tissue samples, such as biopsies, may not be representative of tumors because of the intratumor heterogeneity (8). Some studies have demonstrated that PD-1/PD-L1 inhibitors can benefit patients who have failed first-line chemotherapy when the PD-L1 expression rate is higher than 1%. Moreover, PD-1/PD-L1 inhibitors can even be used as a preferred treatment for patients when the PD-L1 expression exceeds 50% (9). Therefore, it is urgently necessary to find a new approach to assess the expression level of PD-L1.

Radiomics is an emerging field with great development potential, which was first proposed by Dutch scholar Lambin in 2012 (10). In recent years, image omics has developed rapidly, and optimistic results have been achieved in the diagnosis and differential diagnosis of diseases, tumor staging and grading, gene-phenotype prediction, treatment plan decision-making, efficacy evaluation, and prognostic prediction (11–13). In particular, it shows great superiority in lung tumors (14). However, biopsies capture heterogeneity within only a small portion of a tumor and usually at just a single anatomic site, while radiomics captures heterogeneity across the entire tumor.

Combination of functional-metabolic and morphological imaging and F18-fluorodeoxyglucose-positron emission tomography/computed tomography (^{18}F-FDG PET/CT) is the most advanced non-invasive imaging technology at present, which can reflect the glucose uptake level of tissues to a certain extent. It has important application value in the diagnosis, staging, curative effect, and prognostic evaluation of lung cancer (15–17). Several studies have reported the role of radiomics in various malignancies (18, 19). However, research using radiomics based on ^{18}F-FDG PET/CT in combination with clinical risk factors for NSCLC is relatively limited.

In the present study, we aimed to develop a prediction model that incorporated both the radiomic signature and clinicopathologic risk factors for individual prediction of PD-L1 expression in NSCLC patients. Our findings could be helpful to identify the patients who could benefit from the immunotherapy.

MATERIALS AND METHODS

Patients

A retrospective study consisting of NSCLC patients who underwent a combined imaging protocol of ^{18}F-FDG PET/CT between January 2019 and March 2021 was conducted. Ethical approval was obtained for this retrospective analysis, and the informed consent requirement was waived. A total of 255 patients were randomly divided into the training (n = 170) and validation (n = 85) cohorts following a ratio of 7:3 (20). Inclusion criteria were as follows: (a) patients who underwent biopsy or surgery of lung tumor; (b) patients with IHC examination of PD-L1 performed; (c) histological type and grade were pathologically proven; and (d) standard ^{18}F-FDG PET/CT was performed before biopsy or surgery. Exclusion criteria were as follows: (a) therapy (radiotherapy, chemotherapy, or chemoradiotherapy) was performed before ^{18}F-FDG PET/CT and IHC; (b) patients with unknown histological grade; (c) the size of the primary lesion was too small for segmentation; and (d) patients with other types of cancers or with incomplete clinical and imaging datasets.

Detection of PD-L1 Expression

The biopsy and surgery specimens of lung tumors through hematoxylin-eosin (H&E) staining were pathologically examined to confirm the histologic type and grade under the microscope. Furthermore, the expression of PD-L1 was determined through the IHC assay in our study. PD-L1 test kit (22C3 pharmDx) was obtained from the Dako company. The back-to-back interpretation of PD-L1 expression was performed by pathologists, and further reanalysis would be implemented when there was an inconsistency compared with previous results. The data of patients were divided with a PD-L1 cutoff value of 1% and 50%.

PET/CT Image Acquisition and Reconstruction

Patients were recommended to fast for at least 4 h before the FDG-PET/CT scan (4.07-5.55 MBq/kg). Blood glucose levels were maintained at less than 11 mmol/L. A whole-body scan was acquired at 60 ± 10 min after intravenous injection of ^{18}F-FDG using an integrated PET/CT scanner (Discovery STE; General Electric Medical Systems, Milwaukee WI, USA). First, low-dose CT images were performed, with parameters as follows: 140 kV, 120 mA, transaxial field of view (FOV) of 70 cm, pitch of 1.75, rotation time of 0.8 s, and slice thickness of 3.75 mm, followed by PET images, with 2-3 min per bed position and 7-8 bed positions per patient.

Feature Segmentation and Extraction

Tumor segmentation was performed to select primary lesions of NSCLC cases after image acquisition. **Figure 1** shows the workflow of radiomics analysis in this study. PET and CT images of the DICOM format were transferred to LIFEx freeware and automatically fused by the freeware. The LIFEx freeware was used to do quantitative PET/CT analyses (v7.0.0 https://www.lifexsoft.org/) (21). Two experienced nuclear

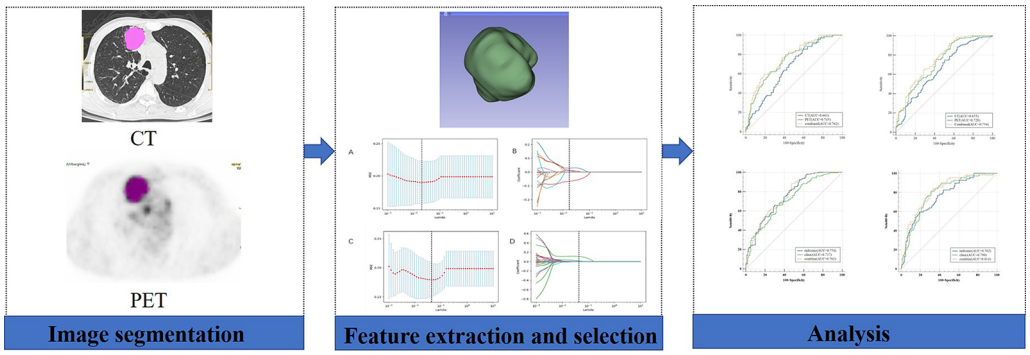


FIGURE 1 | Workflow of the radiomic analysis. A 58-year-old man underwent ^{18}F-FDG PET/CT for staging workup of NSCLC patients with a SUV_{max} of 11.4. The VOI of the lesion was manually delineated, and 41% of SUV_{max} was applied as a threshold to optimize the VOI.

medicine physicians manually segmented the three-dimensional volume of interest (VOI) on each slice, and a threshold of 41% of the maximum standardized uptake value (SUV_{max}) was used to define VOI, including metabolic tumor volume (MTV) and total focal glycolysis (TLG) of lesions. TLG is the MTV multiplied by the mean SUV of the tumor. The voxel size for spatial resampling was 2 × 2 × 2 mm. For CT data, intensity discretization was done with 400 gray levels and absolute scale boundaries between -1,000 and 3,000 HU, whereas for PET data, it was done with 64 bins between 0 and 20. The intraclass correlation coefficient (ICC) was used to determine the repeatability/reproducibility of features in our research, and ICC >0.75 was selected. Subsequently, the least absolute shrinkage and selection operator (LASSO) COX regression model was used to select the most useful prognostic features with 10-fold cross-validation for selecting the parameter Lambda in the training cohort (**Figure 2**) (22, 23).

Prediction Model

Univariable logistic regression analysis began with the following clinical candidate predictors: age, gender, tumor location, histology type and grade, CEA level, smoking history, Ki-67. The radiomic signature and clinicopathologic risk factors with statistically significant differences were incorporated to develop a prediction model by using multivariable logistic regression analysis in a training cohort consisting of 170 consecutive

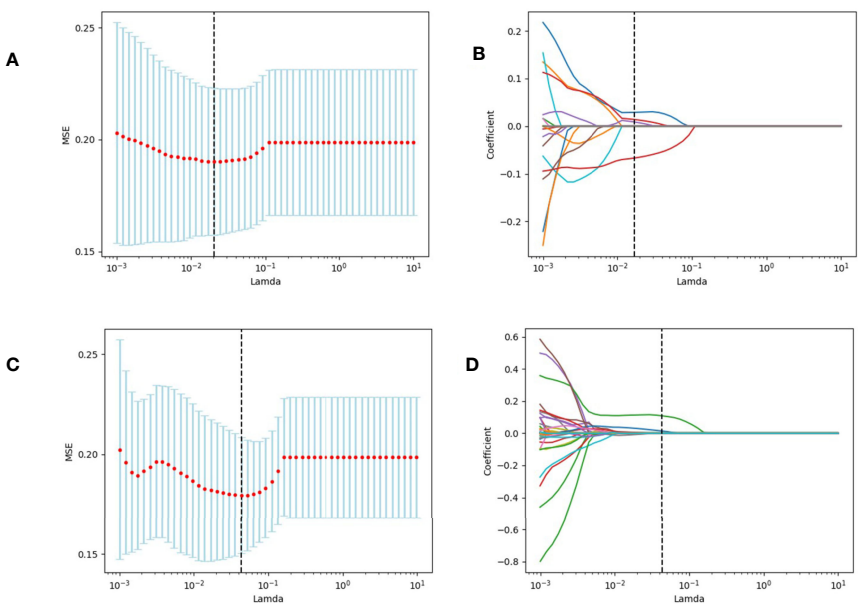


FIGURE 2 | We used two feature selection methods, maximum relevancy and minimum redundancy (MRMR) and the LASSO to select the radiomic feature of CT and PET. At first, MRMR was performed to eliminate the redundant and irrelevant features. Then LASSO was conducted to choose the optimized subset of features to construct the final model. **(A, C)** Tuning parameter Lambda (λ) selection in the LASSO model used 10-fold cross-validation *via* minimum criteria. **(B, D)** LASSO coefficient profiles of the retained features. A vertical line was drawn at the value selected using 10-fold cross-validation.

patients, and the receiver operating characteristic (ROC) curve and the corresponding area under the curve (AUC) were reckoned for the prediction model in the training cohort and validation cohort, respectively.

Statistical Analysis

All statistical analyses were operated with SPSS software version 26.0 (SPSS Inc., Chicago, IL, USA) and python 3.8.0 (https://www.python.org). The differences in patients' characteristics between the training and validation cohorts were compared using the Chi-square test. The spearman rank-order correlation was calculated to analyze the relevance between the expression of PD-L1 and selected features. AUC of the ROC was calculated to evaluate the performance of our prediction model. A $p < 0.05$ was considered statistically significant.

RESULTS

Patients' Characteristics

Table 1 summarizes the clinical characteristics of patients in the training and validation cohorts. A total of 255 patients were enrolled in this study. Among the patients selected, 188 cases (73.7%) were adenocarcinoma, and 67 (26.3%) were squamous cell carcinoma. We demonstrated that several clinicopathologic characteristics might be associated with the expression of PD-L1. Of these patients, there were not any statistically significant differences in the clinical characteristics between the training and validation cohorts (**Table 1**). In our univariate analysis, the expression of PD-L1 was significantly correlated with gender, histologic type, tumor location, and Ki-67 ($p< 0.05$, respectively). However, it was not correlated with age, TNM stage, and history of smoking ($p> 0.05$, respectively) (**Table 2**). Based on the ROC analysis, the optimal cutoff values of SUV_{max}, MTV, and TLG for the PD-L1 1% group were 5.21, 123.94, and 216.62, respectively. Moreover, the optimal cutoff values of SUV_{max}, MTV, and TLG for the PD-L1 50% group were 6.82, 137.57, and 191.68, respectively (**Figure 3**).

Feature Selection in the Training Cohort

A total of 80 radiomic features were extracted in the training dataset (**Table 3**). For the prediction of PD-L1 expression level over 1%, 18 features considered valuable for predicting the PD-L1 expression were extracted, including six features from the CT dataset and 12 features from the PET dataset. For the prediction of PD-L1 expression level over 50%, seven features considered valuable for predicting the PD-L1 expression were extracted, including four features from the CT dataset and three features from the PET dataset. The ICC of the radiomic features was all above 0.75.

Diagnostic Validation of Radiomic Signature and Clinical Features

The model evaluation was conducted in the testing cohort. **Figure 4** shows the AUCs of ROC for the three models (CT, PET, and the combined model) in predicting the PD-L1 expression ≥ 1% and 50%. The AUC scores for predicting the PD-L1 expression over 1% were 0.655 (95% confidence interval [CI]: 0.593-0.713) and 0.728 (95% CI: 0.699-0.782) for features derived from CT and PET only, respectively, and it became 0.754 (95% CI: 0.696-0.805) for combined features. For the prediction of PD-L1 expression over 50%, the AUC scores were 0.661 (95% CI: 0.599-0.719), 0.745 (95% CI: 0.687-0.797), and 0.762 (95% CI: 0.705-0.813) for features derived from CT, PET, and combined model, respectively. **Figure 5** shows the AUCs of ROC for the three models (radiomics, clinics, and the combined model) in predicting PD-L1 expression ≥ 1% and 50%. Using LIFEx, the region of interest (ROI) was initially identified around the tumor outline on the CT and PET images (**Figure 6**).

For the prediction of PD-L1 expression over 1%, the AUC scores were 0.754 (95% CI: 0.696-0.805), 0.636 (95% CI: 0.574-0.695), and 0.757 (95% CI: 0.699-0.808) for features derived from radiomics, clinics, and combined model, respectively. For the

TABLE 1 | Characteristics of the training and validation cohorts.

Characteristics	Total (n = 255)	Training (n = 170)	Validation (n = 85)	t/χ^2	p
Sex				0.009	0.925
Male	170	113	57		
Female	85	57	28		
Age, median ± SD, years	64.22 ± 9.51	64 ± 9.07	64.66 ± 10.37	-.0659	0.603
Tumor location				0.133	0.715
Left lung	97	66	31		
Right lung	158	104	54		
Histologic type, No. (%)				0.04	0.841
Squamous cell carcinoma	67	44	23		
Adenocarcinoma	188	126	62		
TNM stage, No. (%)				2.758	0.097
I-II	194	124	70		
III- IV	61	46	15		
Smoking history				0.314	0.575
Smoker	168	110	58		
Never	87	60	27		
Ki67					
<20%	68	39	29	3.62	0.057
≥20%	187	131	56		

TABLE 2 | Characteristics of NSCLC patients with different PD-L1 expression levels.

Characteristics	PD-L1 < 1%(n = 101)	PD-L1≥1% (n = 154)	t/χ^2	p	PD-L1 < 50% (n = 186)	PD-L1≥50% (n = 69)	t/χ^2	p
Sex			3.967	0.046*			7.242	0.001*
Male	60	110			115	55		
Female	41	44			71	14		
Age, median ± SD, years	63.26 ± 9.46	64.85 ± 9.52			64.15 ± 9.44	64.41 ± 9.75	-0.255	0.849
Tumor location			2.867	0.09			7.996	0.005*
Left lung	32	65			66	38		
Right lung	69	89			120	31		
Histologic type, No. (%)			6.169	0.013*			4.842	0.028*
Squamous cell carcinoma	18	49			42	25		
Adenocarcinoma	83	105			144	44		
TNM stage, No. (%)			27.617	0.000*			2.634	0.105
I-II	86	82			128	40		
III- IV	15	72			58	29		
Smoking history			5.321	0.021*			1.108	0.292
Smoker	58	110			119	49		
Never	43	44			67	20		
Ki67			11.721	0.001*			5.564	0.018*
<20%	41	32			57	11		
≥20%	60	122			129	58		
SUV_{max}			30.304	0.000*			29.694	0.000*
High	58	135			105	64		
Low	43	19			81	5		
MTV			10.924	0.001*			9.998	0.002*
High	37	89			72	42		
Low	64	65			114	27		
TLG			17.149	0.001*			20.118	0.000*
High	41	103			93	56		
Low	60	51			93	13		

**Statistically significant, p < 0.05.*

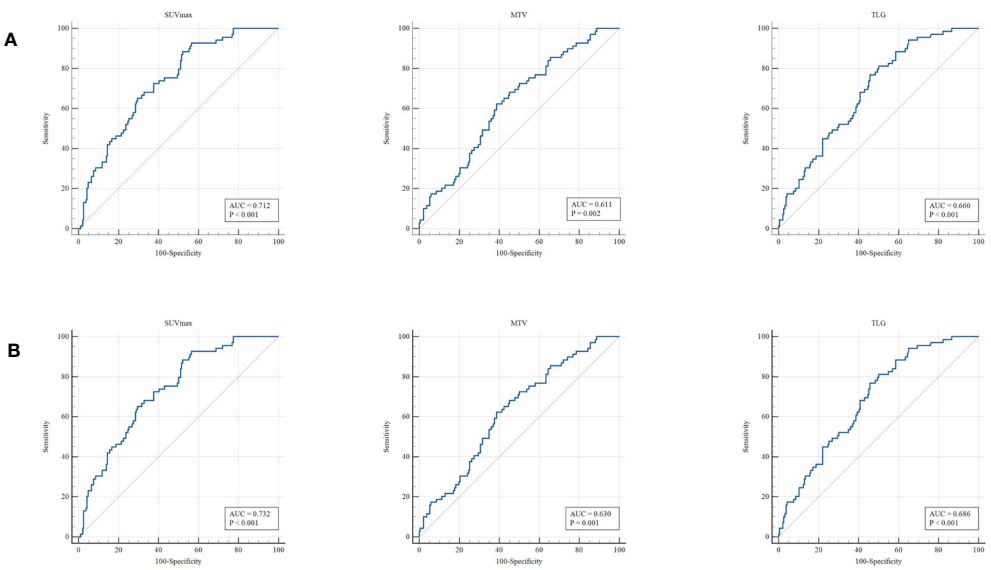


FIGURE 3 | ROC curve for the determination of the most discriminative cutoff points for SUV_{max}, MTV, and TLG in primary tumors. The optimal cutoff values of SUVmax, MTV, and TLG for the PD-L1 1% group **(A)** were 5.21, 123.94, and 216.62, respectively. The optimal cutoff values of SUV_{max}, MTV, and TLG for the PD-L1 50% group **(B)** were 6.82, 137.57, and 191.68, respectively.

TABLE 3 | Radiomic parameters.

Conventional textural features	First-order textural features
SUVmin	HISTO_Skewness
SUVmean	HISTO_Kurtosis
SUVstd	HISTO_Entropy_log10
SUVmax	HISTO_Entropy_log2
SUVpeak*	HISTO_Energy
TLG*	SHAPE_Sphericity
	SHAPE_Compacity
	SHAPE_Volume (mL)
	SHAPE_Volume(vx)
Higher-order textural features	
GLZLM	**GLRLM**
GLZLM_SZE (Short-Zone Emphasis)	GLRLM_SRE (Short-Run Emphasis)
GLZLM_LZE (Long-Zone Emphasis)	GLRLM_LRE (Long-Run Emphasis)
GLZLM_LGZE (Low Gray-level Zone Emphasis)	GLRLM_LGRE (Low Gray-level Run Emphasis)
GLZLM_HGZE (High Gray-level Zone Emphasis)	GLRLM_HGRE (High Gray-level Run Emphasis)
GLZLM_SZLGE (Short-Zone Low Gray-level Emphasis)	GLRLM_SRLGE (Short-Run Low Gray-level Emphasis)
GLZLM_SZHGE (Short-Zone High Gray-level Emphasis)	GLRLM_SRHGE (Short-Run High Gray-level Emphasis)
GLZLM_LZLGE (Long-Zone Low Gray-level Emphasis)	GLRLM_LRLGE (Long-Run Low Gray-level Emphasis)
GLZLM_LZHGE (Long-Zone High Gray-level Emphasis)	GLRLM_LRHGE (Long-Run High Gray-level Emphasis)
GLZLM_GLNU (Gray-Level Non-Uniformity for zone)	GLRLM_GLNU (Gray-Level Non-Uniformity for run)
GLZLM_ZLNU (Zone Length Non-Uniformity)	GLRLM_RLNU (Run Length Non-Uniformity)
GLZLM_ZP (Zone Percentage)	GLRLM_RP (Run Percentage)
GLCM	**NGLDM**
GLCM_Homogeneity	NGLDM_Coarseness
GLCM_Energy	NGLDM_Contrast
GLCM_Contrast	NGLDM_Busyness
GLCM_Correlation	
GLCM_Entropy_log10	
GLCM_Entropy_log2	
GLCM_Dissimilarity	

**Calculated only for PET.*

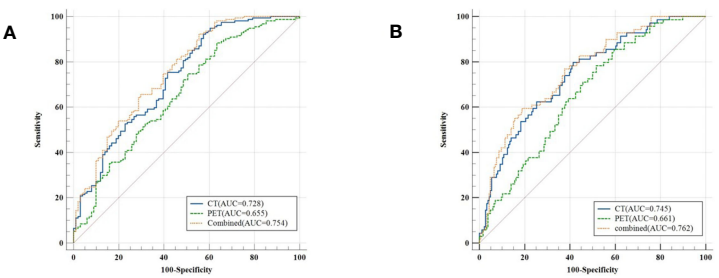


FIGURE 4 | Classifiers' performance on predicting the expression status of PD-L1 based on three models (CT, PET, and the combined model). Classifiers' performance on predicting 1% level of PD-L1 **(A)**. Classifiers' performance on predicting 50% level of PD-L1 **(B)**.

prediction of PD-L1 expression over 50%, the AUC scores were 0.762 (95% CI: 0.705-0.813), 0.672 (95% CI: 0.611-0.730), and 0.814 (95% CI: 0.761-0.860) for features derived from radiomics, clinics, and combined model, respectively.

DISCUSSION

Our present study demonstrated that ^{18}F-FDG PET/CT radiomic signature was useful for assessing the expression rate of PD-L1 through radiomic features in NSCLC patients. Radiomic signature successfully stratified patients according to the PD-L1 expression rate threshold of 1% and 50%. The combination of the radiomic signature and clinicopathologic risk factors presented a better diagnostic efficacy compared with the simple radiomic signature or clinical feature model.

Studies have shown that PD-L1 is highly expressed on the surface of a variety of tumor cells (including lung cancer) (24–27). PD-1/PD-L1 inhibitors can exert the immune effect of T cells against tumors in a variety of ways and inhibit tumor development. More and more clinical evidence supports the effect of PD1/PD-L1 inhibitors in the treatment of lung cancer (28).

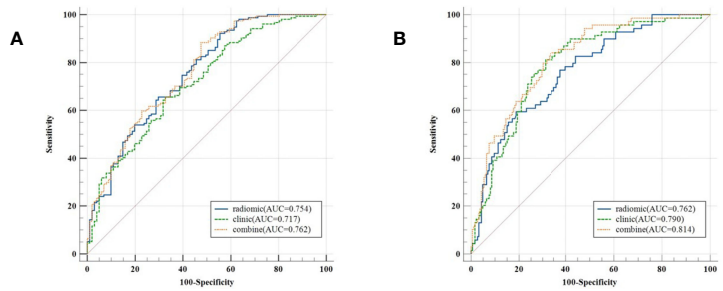


FIGURE 5 | Classifiers' performance on predicting the expression status of PD-L1 based on three models (radiomics, clinics, and the combined model). Classifiers' performance on predicting 1% level of PD-L1 **(A)**. Classifiers' performance on predicting 50% level of PD-L1 **(B)**.

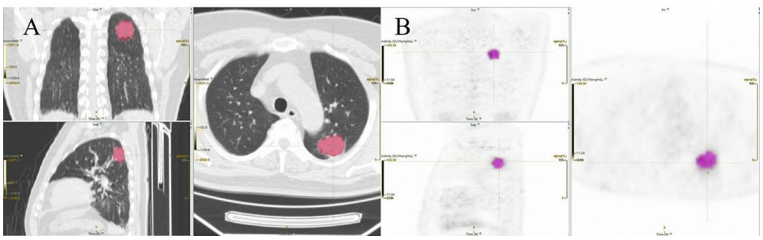


FIGURE 6 | An example of segmentation coronary image, sagittal image, and transaxial images of CT **(A)** and PET **(B)** showing an example of VOI for measuring imaging features of NSCLC.

Varying degrees of survival benefit and delay of disease progression in patients with lung adenocarcinoma can be achieved no matter PD-1/PD-L1 inhibitors are used alone or in combination with chemotherapy and molecular targeted therapy. At present, IHC is the main method utilized in detecting the PD-L1 expression rate. Several preclinical PET studies have also demonstrated non-invasive imaging of PD-L1 expression in tumors (29, 30).

At present, radiomics has been widely used in lung cancer patients, while it is rarely used to predict the expression of PD-L1 in NSCLC based on PET/CT images and clinicopathologic risk. Cancer cells within the same tumor are now recognized to be diverse in many ways. Many cell features, such as shape or phenotypic expression, display of inherent or acquired treatment resistance, and ability to initiate new tumor development, show heterogeneity. Intratumor heterogeneity is an important factor in determining tumor treatment response and patient prognosis (31). The blood perfusion, hypoxia, cell proliferation, necrosis, and other factors within the tumor cause significant internal biological differences (32). Because it gives an observer-independent measurement, SUV_{max} is a widely used parameter. MTV and TLG have been developed to measure the metabolic activity of the entire tumor mass. These parameters are designed to measure the overall changes in tumor glycolysis. Preliminary research has found that compared with SUV_{max} and SUV_{mean}, ^{18}F-FDG PET/CT image texture analysis can capture heterogeneity across the entire tumor and provide more valuable information in the diagnosis, staging, curative effect prediction, and prognosis of NSCLC (33). Preoperative SUV_{max} is correlated with PD-L1 expression in NSCLC patients. In the univariate analysis, the expression of PD-L1 in lung tumors was significantly correlated with SUV_{max}, MTV, and TLG ($p< 0.05$). The features extracted from CT performed better than those of PET in assessing the expression status of PD-L1 both in the PD-L1 1% and 50% groups. The reason was attributed to that the density resolution of the PET image was not so good as the CT image, which could have a great effect on extracting and selecting the meaningful radiomic features. When combined with CT features, the model showed improvement in distinguishing the PD-L1 expression level between in the PD-L1 1% and 50% groups. A recent study has assessed the expression of PD-L1 by radiomic features from PET/CT images in NSCLC patients, showing that radiomic signatures of PD-L1 expression over 1% and 50% reach an AUC score of 0.85 and 0.880, respectively (34). However, they do not combine radiomic features with clinical risk factors in the prediction model. Sun *et al.* have assessed the expression of PD-L1 in tumor cells in NSCLC patients by using a radiomic study based on CT images and clinicopathologic features, and the score of AUC is 0.848 (35), which is consistent with our results. Another study has investigated the association between PD-L1 expression and textural features of PET images in 53 patients with

oropharyngeal or hypopharyngeal cancer, while the sample number is too small, and the constructed prediction model of PD-L1 expression by the radiomics cannot be robust (36). A study has shown that PD-L1 is more common in patients with the following clinical characteristics: larger tumor size, more positive lymph node involvement, greater historically tumor grade, and higher Ki-67 index (37). Another study has pointed out that positive Ki-67 expression is strongly associated with positive PD-L1 expression (38), which is consistent with our results. In the present study, we classified NSCLC patients according to their PD-L1 expression levels and found that PD-L1 expression levels were associated with differences in gender, pathological type, and Ki-67 levels of patients. Wu *et al.* have found that the expression of PD-L1 is significantly associated with the advanced N stage but not with T and M stages (39). Subsequently, we incorporated the radiomic signature and clinicopathological factors into a combined model, which presented a better diagnostic efficacy (AUC=0.814) compared with the simple radiomic signature or clinical feature model both in the PD-L1 1% and 50% groups. In our present study, we found that for the prediction of PD-L1 expression over 1% and 50%, the AUC scores were 0.757 and 0.814 for features derived from the combined model, respectively. Our findings were consistent with the previous studies, indicating that PET radiomic features were useful to screen the NSCLC patients who could benefit from the anti-PD-L1 immunotherapy. However, only very few studies have investigated the sensitivity and specificity of PD-L1 in NSCLC patients. Many sources may cause these differences, such as small sample size, image segmentation, acquisition and reconstruction parameters, and feature extraction software. Further investigations in a larger cohort population are required to validate our conclusions.

Repeatability is a basic requirement in radiomic analysis (40, 41). In the present study, all ^{18}F-FDG PET/CT images were realized in the same center using the same acquisition and reconstruction protocols. To reduce the impact of discretization values on robustness, a reliable discretization using a fixed size of bins was adopted (42, 43).

The present study has several limitations. First, this was a single-center retrospective study, and the sample size was small. This indicated that the variability among image characteristics from various localities was not completely captured, and potential selection bias might exist. Therefore, our results need to be confirmed by studies with larger sample sizes. Second, the expression of PD-L1 in the tumor had inherent instability in individual patients. Third, manual drawing ROI and manual image segmentation were adopted, which had poor reproducibility and high technical requirements for operators.

CONCLUSIONS

In the present study, we established a prediction model based on PET/CT images and clinicopathological characteristics to predict the expression of PD-L1 in NSCLC patients and provided a novel strategy for clinicians to screen the patients who could benefit from anti-PD-L1 immunotherapy.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

JL, SG, and SS contributed equally to this work. All authors contributed to the article and approved the submitted version.

FUNDING

The present study was supported by the National Natural Science Foundation of China (grant no. 81601522), Medical Youth Talent Project of Jiangsu Province (grant no. QNRC2016749), Gusu Health Talent Program (grant no. GSWS2020013), Suzhou People's Livelihood Science and Technology Project (grant no. SYS2019038), and Project of State Key Laboratory of Radiation Medicine and Protection, Soochow University, (No. GZK1202127).

REFERENCES

1. Muller DC, Johansson M, Brennan P. Lung Cancer Risk Prediction Model Incorporating Lung Function: Development and Validation in the UK Biobank Prospective Cohort Study. *J Clin Oncol* (2017) 35(8):861–9. doi: 10.1200/JCO.2016.69.2467
2. Collins LG, Haines C, Perkel R, Enck RE. Lung Cancer: Diagnosis and Management. *Am Fam Physician* (2007) 75(1):56–63.
3. Osmani L, Askin F, Gabrielson E, Li QK. Current WHO Guidelines and the Critical Role of Immunohistochemical Markers in the Subclassification of Non-Small Cell Lung Carcinoma (NSCLC): Moving From Targeted Therapy to Immunotherapy. *Semin Cancer Biol* (2018) 52(Pt 1):103–9. doi: 10.1016/j.semcancer.2017.11.019
4. Lastwika KJ, Wilson W 3rd, Li QK, Norris J, Xu H, Ghazarian SR, et al. Control of PD-L1 Expression by Oncogenic Activation of the AKT-Mtor Pathway in non-Small Cell Lung Cancer. *Cancer Res* (2016) 76(2):227–38. doi: 10.1158/0008-5472.CAN-14-3362
5. Brahmer JR, Tykodi SS, Chow LQ, Hwu WJ, Topalian SL, Hwu P, et al. Safety and Activity of Anti-PD-L1 Antibody in Patients With Advanced Cancer. *N Engl J Med* (2012) 366(26):2455–65. doi: 10.1056/NEJMoa1200694
6. Gniadek TJ, Li QK, Tully E, Chatterjee S, Nimmagadda S, Gabrielson E. Heterogeneous Expression of PD-L1 in Pulmonary Squamous Cell

Carcinoma and Adenocarcinoma: Implications for Assessment by Small Biopsy. *Mod Pathol* (2017) 30(4):530–8. doi: 10.1038/modpathol.2016.213
7. Rimm DL, Han G, Taube JM, Yi ES, Bridge JA, Flieder DB, et al. A Prospective, Multi-Institutional, Pathologist-Based Assessment of 4 Immunohistochemistry Assays for PD-L1 Expression in Non-Small Cell Lung Cancer. *JAMA Oncol* (2017) 3(8):1051–8. doi: 10.1001/jamaoncol.2017.0013
8. Senosain MF, Massion PP. Intratumor Heterogeneity in Early Lung Adenocarcinoma. *Front Oncol* (2020) 10:349. doi: 10.3389/fonc.2020.00349
9. Zhang M, Li G, Wang Y, Wang Y, Zhao S, Haihong P, et al. PD-L1 Expression in Lung Cancer and its Correlation With Driver Mutations: A Meta-Analysis. *Sci Rep* (2017) 7(1):10255. doi: 10.1038/s41598-017-10925-7
10. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, et al. Radiomics: Extracting More Information From Medical Images Using Advanced Feature Analysis. *Eur J Cancer* (2012) 48(4):441–6. doi: 10.1016/j.ejca.2011.11.036
11. Zheng X, Yao Z, Huang Y, Yu Y, Wang Y, Liu Y, et al. Deep Learning Radiomics can Predict Axillary Lymph Node Status in Early-Stage Breast Cancer. *Nat Commun* (2020) 11(1):1236. doi: 10.1038/s41467-020-15027-z [Published Correction Appears in Nat Commun. 2021 Jul 12;12(1):4370.
12. Lao J, Chen Y, Li ZC, Li Q, Zhang J, Liu J, et al. A Deep Learning-Based Radiomics Model for Prediction of Survival in Glioblastoma Multiforme. *Sci Rep* (2017) 7(1):10353. doi: 10.1038/s41598-017-10649-8
13. Sun R, Limkin EJ, Vakalopoulou M, Dercle L, Champiat S, Han SR, et al. A Radiomics Approach to Assess Tumour-Infiltrating CD8 Cells and Response to Anti-PD-1 or Anti-PD-L1 Immunotherapy: An Imaging Biomarker, Retrospective Multicohort Study. *Lancet Oncol* (2018) 19(9):1180–91. doi: 10.1016/S1470-2045(18)30413-3
14. Chen B, Yang L, Zhang R, Luo W, Li W. Radiomics: An Overview in Lung Cancer Management-a Narrative Review. *Ann Transl Med* (2020) 8(18):1191. doi: 10.21037/atm-20-4589
15. Kandathil A, Kay FU, Butt YM, Wachsmann JW, Subramaniam RM. Role of FDG PET/CT in the Eighth Edition of TNM Staging of Non-Small Cell Lung Cancer. *Radiographics* (2018) 38(7):2134–49. doi: 10.1148/rg.2018180060
16. Farsad M. FDG PET/CT in the Staging of Lung Cancer. *Curr Radiopharm* (2020) 13(3):195–203. doi: 10.2174/1874471013666191223153755
17. Kandathil A, Sibley RC III, Subramaniam RM. Lung Cancer Recurrence: ^{18}F-FDG PET/CT in Clinical Practice. *AJR Am J Roentgenol* (2019) 213(5):1136–44. doi: 10.2214/AJR.19.21227
18. Delgadillo R, Ford JC, Abramowitz MC, Dal Pra A, Pollack A, Stoyanova R. The Role of Radiomics in Prostate Cancer Radiotherapy. *Strahlenther Onkol* (2020) 196(10):900–12. doi: 10.1007/s00066-020-01679-9
19. Sah BR, Owczarczyk K, Siddique M, Cook GJR, Goh V. Radiomics in Esophageal and Gastric Cancer. *Abdom Radiol (NY)* (2019) 44(6):2048–58. doi: 10.1007/s00261-018-1724-8
20. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images are More Than Pictures, They are Data. *Radiology* (2016) 278(2):563–77. doi: 10.1148/radiol.2015151169
21. Nioche C, Orlhac F, Boughdad S, Reuzé S, Goya-Outi J, Robert C, et al. Lifex: A Freeware for Radiomic Feature Calculation in Multimodality Imaging to Accelerate Advances in the Characterization of Tumor Heterogeneity. *Cancer Res* (2018) 78(16):4786–9. doi: 10.1158/0008-5472.CAN-18-0125
22. Zhang B, Tian J, Dong D, Gu D, Dong Y, Zhang L, et al. Radiomics Features of Multiparametric MRI as Novel Prognostic Factors in Advanced Nasopharyngeal Carcinoma. *Clin Cancer Res* (2017) 23(15):4259–69. doi: 10.1158/1078-0432.CCR-16-2910
23. Tibshirani R. The Lasso Method for Variable Selection in the Cox Model. *Stat Med* (1997) 16(4):385–95. doi: 10.1002/(sici)1097-0258(19970228)16:4<385::aid-sim380>3.0.co 2-3.
24. Masugi Y, Nishihara R, Yang J, Mima K, da Silva A, Shi Y, et al. Tumour CD274 (PD-L1) Expression and T Cells in Colorectal Cancer. *Gut* (2017) 66 (8):1463–73. doi: 10.1136/gutjnl-2016-311421
25. Mittendorf EA, Philips AV, Meric-Bernstam F, Qiao N, Wu Y, Harrington S, et al. PD-L1 Expression in Triple-Negative Breast Cancer. *Cancer Immunol Res* (2014) 2(4):361–70. doi: 10.1158/2326-6066.CIR-13-0127
26. Gu L, Chen M, Guo D, Zhu H, Zhang W, Pan J, et al. PD-L1 and Gastric Cancer Prognosis: A Systematic Review and Meta-Analysis. *PloS One* (2017) 12(8):e0182692. doi: 10.1371/journal.pone.0182692
27. Yu H, Boyle TA, Zhou C, Rimm DL, Hirsch FR. PD-L1 Expression in Lung Cancer [Published Correction Appears in J Thorac Oncol. *J Thorac Oncol* (20172016) 1211(17):157–9964–75. doi: 10.1016/j.jtho.2016.04.014
28. Li JX, Huang JM, Jiang ZB, Li RZ, Sun A, Lai-Han Leung E, et al. Current Clinical Progress of PD-1/PD-L1 Immunotherapy and Potential Combination Treatment in Non-Small Cell Lung Cancer. *Integr Cancer Ther* (2019) 18:1534735419890020. doi: 10.1177/1534735419890020
29. Wu X, Huang Y, Zhao Q, Wang Q, Song Q, Li Q, et al. PD-L1 Expression Correlation With Metabolic Parameters of FDG PET/CT and Clinicopathological Characteristics in Non-Small Cell Lung Cancer. *EJNMMI Res* (2020) 10(1):51. doi: 10.1186/s13550-020-00639-9
30. Cui Y, Li X, Du B, Diao Y, Li Y. PD-L1 in Lung Adenocarcinoma: Insights Into the Role of ^{18}F-FDG PET/CT. *Cancer Manag Res* (2020) 12:6385–95. doi: 10.2147/CMAR.S256871
31. Denison TA, Bae YH. Tumor Heterogeneity and its Implication for Drug Delivery. *J Control Release* (2012) 164(2):187–91. doi: 10.1016/j.jconrel.2012.04.014
32. Bashir U, Siddique MM, Mclean E, Goh V, Cook GJ. Imaging Heterogeneity in Lung Cancer: Techniques, Applications, and Challenges. *AJR Am J Roentgenol* (2016) 207(3):534–43. doi: 10.2214/AJR.15.15864
33. Moon SH, Hyun SH, Choi JY. Prognostic Significance of Volume-Based PET Parameters in Cancer Patients. *Korean J Radiol* (2013) 14(1):1–12. doi: 10.3348/kjr.2013.14.1.1
34. Jiang M, Sun D, Guo Y, Guo Y, Xiao J, Wang L, et al. Assessing PD-L1 Expression Level by Radiomic Features From PET/CT in Nonsmall Cell Lung Cancer Patients: An Initial Result. *Acad Radiol* (2020) 27(2):171–9. doi: 10.1016/j.acra.2019.04.016
35. Sun Z, Hu S, Ge Y, Wang J, Duan S, Song J, et al. Radiomics Study for Predicting the Expression of PD-L1 in Non-Small Cell Lung Cancer Based on CT Images and Clinicopathologic Features. *J Xray Sci Technol* (2020) 28 (3):449–59. doi: 10.3233/XST-200642
36. Chen RY, Lin YC, Shen WC, Hsieh TC, Yen KY, Chen SW, et al. Associations of Tumor PD-1 Ligands, Immunohistochemical Studies, and Textural Features in ^{18}F-FDG PET in Squamous Cell Carcinoma of the Head and Neck. *Sci Rep* (2018) 8(1):105. doi: 10.1038/s41598-017-18489-2
37. Huang W, Ran R, Shao B, Li H. Prognostic and Clinicopathological Value of PD-L1 Expression in Primary Breast Cancer: A Meta-Analysis. *Breast Cancer Res Treat* (2019) 178(1):17–33. doi: 10.1007/s10549-019-05371-0
38. Rubino S, Kim Y, Zhou J, Dhilon J, Li R, Spiess P, et al. Positive Ki-67 and PD-L1 Expression in Post-Neoadjuvant Chemotherapy Muscle-Invasive Bladder Cancer is Associated With Shorter Overall Survival: A Retrospective Study. *World J Urol* (2021) 39(5):1539–47. doi: 10.1007/s00345-020-03342-5
39. Wu CE, Chang CF, Kou-Sheng L, Chiang J, Lee SW, Chiu YC. PD-L1 Immunohistochemistry Comparability and Their Correlation With Clinical Characteristics in NSCLC. *Anal Cell Pathol (Amst)* (2020) 2020:3286139. doi: 10.1155/2020/3286139
40. Schwier M, van Griethuysen J, Vangel MG, Pieper S, Peled S, Tempany C, et al. Repeatability of Multiparametric Prostate MRI Radiomics Features. *Sci Rep* (2019) 9(1):9441. doi: 10.1038/s41598-019-45766-z
41. Vallières M, Zwanenburg A, Badic B, Cheze Le Rest C, Visvikis D, Hatt M. Responsible Radiomics Research for Faster Clinical Translation. *J Nucl Med* (2018) 59(2):189–93. doi: 10.2967/jnumed.117.200501
42. Liberini V, De Santi B, Rampado O, Gallio E, Dionisi B, Ceci F, et al. Impact of Segmentation and Discretization on Radiomic Features in ^{68}Ga-DOTA-TOC PET/CT Images of Neuroendocrine Tumor. *EJNMMI Phys* (2021) 8(1):21. doi: 10.1186/s40658-021-00367-6
43. Lu L, Lv W, Jiang J, Ma J, Feng Q, Rahmim A, et al. Robustness of Radiomic Features in [11C] Choline and [18F] FDG PET/CT Imaging of Nasopharyngeal Carcinoma: Impact of Segmentation and Discretization. *Mol Imaging Biol* (2016) 18(6):935–45. doi: 10.1007/s11307-016-0973-6

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

The reviewer LZ declared a shared parent affiliation with the authors to the handling editor at the time of the review.

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

this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Oncology

ORIGINAL RESEARCH
published: 31 January 2022
doi: 10.3389/fonc.2021.756503

Comprehensive Output Estimation of Double Scattering Proton System With Analytical and Machine Learning Models

*Jiahua Zhu[1,2], Taoran Cui[1], Yin Zhang[1], Yang Zhang[1], Chi Ma[1], Bo Liu[1,3], Ke Nie[1], Ning J. Yue[1] and Xiao Wang[1]**

[1] Department of Radiation Oncology, Rutgers-Cancer Institute of New Jersey, Rutgers-Robert Wood Johnson Medical School, New Brunswick, NJ, United States, [2] Department of Radiation Oncology, Reading Hospital, Tower Health, West Reading, PA, United States, [3] Department of Radiation Oncology, City of Hope National Medical Center, Duarte, CA, United States

OPEN ACCESS

Edited by:
Sonali Rudra,
MedStar Georgetown University Hospital, United States

Reviewed by:
Guillermo Cabrera-Guerrero,
Pontificia Universidad Católica de Valparaíso, Chile
Yingzi Liu,
Emory University, United States

***Correspondence:**
Xiao Wang
xw240@cinj.rutgers.edu

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

Received: 10 August 2021
Accepted: 29 December 2021
Published: 31 January 2022

Citation:
Zhu J, Cui T, Zhang Y, Zhang Y, Ma C, Liu B, Nie K, Yue NJ and Wang X (2022) Comprehensive Output Estimation of Double Scattering Proton System With Analytical and Machine Learning Models. Front. Oncol. 11:756503. doi: 10.3389/fonc.2021.756503

Objectives: The beam output of a double scattering proton system varies for each combination of beam option, range, and modulation and therefore is difficult to be accurately modeled by the treatment planning system (TPS). This study aims to design an empirical method using the analytical and machine learning (ML) models to estimate proton output in a double scattering proton system.

Materials and Methods: Three analytical models using polynomial, linear, and logarithm–polynomial equations were generated on a training dataset consisting of 1,544 clinical measurements to estimate proton output for each option. Meanwhile, three ML models using Gaussian process regression (GPR) with exponential kernel, squared exponential kernel, and rational quadratic kernel were also created for all options combined. The accuracy of each model was validated against 241 additional clinical measurements as the testing dataset. Two most robust models were selected, and the minimum number of samples needed for either model to achieve sufficient accuracy (± 3%) was determined by evaluating the mean average percentage error (MAPE) with increasing sample number. The differences between the estimated outputs using the two models were also compared for 1,000 proton beams with a randomly generated range, and modulation for each option.

Results: The polynomial model and the ML GPR model with exponential kernel yielded the most accurate estimations with less than 3% deviation from the measured outputs. At least 20 samples of each option were needed to build the polynomial model with less than 1% MAPE, whereas at least a total of 400 samples were needed for all beam options to build the ML GPR model with exponential kernel to achieve comparable accuracy. The two independent models agreed with less than 2% deviation using the testing dataset.

Conclusion: The polynomial model and the ML GPR model with exponential kernel were built for proton output estimation with less than 3% deviations from the measurements. They can be used as an independent output prediction tool for a double scattering proton beam and a secondary output check tool for a cross check between themselves.

Keywords: output model, analytical model, machine learning, Gaussian process regression, double scattering proton system

INTRODUCTION

Proton therapy is rapidly becoming one of the primary cancer treatment modalities in the recent decade. The utilization of the Bragg peak plays a pivotal role in delivering the prescription dose to the target, while sparing the normal tissues by stopping the proton beam at the distal end of the target (1–4). In order to cover the entire target with a desired dose, the pristine Bragg peak has to be modulated to the spread-out Bragg peak (SOBP) in terms of target size and depth (5–7). Due to the complexity of proton beamline to form various SOBPs in a double scattering proton machine, it is hard to model the output accurately. Therefore, most proton centers with a double scattering beam system have to measure the output of patient-specific proton beams in a water phantom to determine the required machine output, mostly in terms of monitor unit (MU).

In order to obtain the output of a proton beam conveniently and verify the output measurement, Kooy et al. proposed a semi-empirical analytical method to estimate the output as a function of r = (R - M)/M, where R and M denote the beam range and modulation, respectively (8, 9). This formula implements a basic model as a function of r and also corrects for the effective source position based on the inverse square law. However, this model was sensitive to the definition of range and modulation (10). A variation of 18% in output was observed at beam data with small modulation (10, 11). Therefore, Lin et al. proposed a parameterized linear quadratic model which defined r with a limited length of modulation (11). With this correction, the relative errors of predicted outputs compared to measured values were less than 3%. Besides, the basic model of output in Kooy's method was also fitted by the fourth-order Taylor polynomial multiplied by a range-related factor, which was close to unity (12). In terms of a comparison between Kooy's original method and the Taylor series approach, the predicted values from the Taylor series approach were closer to the measurements. Sahoo et al. comprehensively analyzed the determination of output from the proton machine beamline, where the relative output factor, SOBP factor, and range shifter factor were the primary factors to determine the output (13). The result also showed a good agreement to the measurement within 2% for 99% of those fields. However, this method required a large amount of measurements to verify the conversion from the SOBP factor and range shifter factor to the output, which was time-consuming and complicated. Machine learning (ML) models have also been used in output prediction (14). Sun et al. compared the accuracy of output from machine learning and Kooy's method (10). Up to 7.7% of relative error from Kooy's method was reduced to 3.17% by machine learning.

We propose three analytical models and three machine learning algorithms for output estimation. The analytical models include a polynomial fitting model, a linear fitting model, and a logarithm–polynomial fitting model, all with different equations for different options. The machine learning algorithms utilize the Gaussian process regression (GPR) model with different kernels to test the accuracy of output estimations, with one single model for all options. The definition of R and M is consistent with the machine vendor's definition, and the data are from our clinical beam measurement. The comparison between predicted and measured outputs was performed. In addition, the minimum number of beam data measurements needed for building a robust model is discussed, which can provide some insights for clinical implementation.

MATERIALS AND METHODS

Introduction of the Proton Machine

Mevion (Mevion Medical Systems, Inc., Littleton, MA) S250 utilizes a double scatter system to broaden the pencil beam and creates a uniform dose distribution with a beam shaping system. The beam shaping system includes primary and secondary scatters, one absorber, and one range modulator, which spread out the Bragg peak. There are two types of nozzles on the inner gantry, a large applicator (maximum 25 cm in diameter) and a small applicator (maximum 14 cm in diameter), respectively. A brass aperture mounted on the applicator shapes the proton beam to cover the target. A compensator mounted at the end of the applicator modulates the distal end of the proton beam. There are 24 options with different beam ranges, beam modulations, and field sizes, as listed in **Table 1**. The first 12 options are large options to be used with the large applicator. The other 12 options are deep/small options to be used with the small applicator.

Output Measurement

Due to the complexity of the proton beamline in a double scattering system, the Varian Eclipse treatment planning system (TPS) (Varian Medical Systems, Palo Alto, CA) does not provide MU directly for a proton beam. Instead, the output has to be determined manually for each clinical proton beam.

TABLE 1 | The statistics of all options.

	Max range (cm)	Min range (cm)	Max modulation (cm)	Training field	Testing field
Option 1	25.0	22.6	20.0	55	8
Option 2	22.5	20.9	20.0	40	8
Option 3	20.8	18.8	20.0	76	20
Option 4	18.7	16.8	18.7	99	22
Option 5	16.7	14.9	16.7	68	18
Option 6	14.8	13.2	14.8	81	13
Option 7	13.1	11.5	13.1	90	20
Option 8	11.4	10.0	11.4	98	22
Option 9	9.9	8.6	9.9	90	19
Option 10	8.5	7.3	8.5	86	9
Option 11	7.2	6.1	7.2	45	3
Option 12	6.0	5.0	6.0	21	2
Option 13	32.0	29.6	10.0	3	0
Option 14	29.5	27.1	10.0	12	0
Option 15	27.0	24.6	10.0	37	8
Option 16	24.5	22.1	10.0	49	2
Option 17	22.0	20.1	10.0	7	1
Option 18	20.0	17.8	20.0	19	5
Option 19	17.7	15.4	17.7	52	6
Option 20	15.3	13.3	15.3	105	14
Option 21	13.2	11.2	13.2	173	11
Option 22	11.1	9.1	11.1	123	12
Option 23	9.0	7.0	9.0	65	9
Option 24	6.9	5.0	6.9	50	7

To determine the MU for a clinical proton beam, a verification plan was generated by copying the original clinical proton beam to a water phantom with the same proton energy fluence. Regardless of the setup in the original clinical plan, a consistent setup with SSD = 190 cm was used in the verification plan. The compensator in the original clinical plan was removed in the verification plan to reduce measurement uncertainty. A reference point was added to determine the dose at the mid-SOBP of the beam, and the output measurements were conducted in the water phantom at the mid-SOBP of the same proton beam with a Farmer chamber (IBA Dosimetry America Inc., Memphis, TN) at SSD = 190 cm. Attention was paid to the in-plane location of the reference point to ensure lateral charge particle equilibrium for accurate dose prediction. Sun et al. and Sahoo et al. demonstrated that the field size effect is negligible with a field opening of at least 5-cm diameter (10, 13). If the verification point was blocked, it was shifted. The same setup was then applied in a water phantom for absolute dose measurement. The absolute point dose at the verification point of 100 MU was measured following the IAEA TRS398 protocol (15). The output was essentially d/MU, where d was the absolute point dose measured, and MU was 100. Given the outputs were the same for both clinical beam and verification beam, the MU of the patient-specific beam would be calculated by taking the ratio of the verification point dose from TPS to the measured output.

Analytical Model-Based Output Estimation

In order to validate and verify and eventually replace the manual measurement, output models were built based on previous measurements. Analytical models using an empirical formula to convert from range/modulation to output were built for each option, based on 1,785 proton clinical field measurements. 1,544 clinical proton fields from 2015 to 2019 were categorized as training dataset and the rest (241 fields) as the testing dataset. Three analytical models were employed to estimate the output and compared to the measurement as reference. The specific workflow of data modeling and accuracy verification is shown in **Figure 1**.

Polynomial Fitting Model

The polynomial fitting model is an adaptation of Kooy's empirical formula. A simple demonstration is shown for better understanding. In Kooy's formula (9), output is a function of r = (R - M)/M. According to the vendor definition, R is defined as the depth at distal 90% of the normalized percent depth dose and M is defined as the length between proximal 95% and the distal 90% of the normalized percent depth dose. The basic model of Kooy's formula is expressed in Ferguson et al. (12) as

$$\begin{aligned} &d/MU(r(R,M)) \\ &= \frac{CF \times \psi_c \times D_c}{100/(1+a_0 r^{a1})} \times [s_0 + s_1(R - R_L)] \\ &\times \left(\frac{ESAD(r) - \Delta z_p}{ESAD(r) - \Delta z_p - \Delta z} \right)^2 \end{aligned} \quad (1)$$

The first term of Eq. 1 is the basic output prediction; the second term corrects the variation of output related to the virtual source position; and the third term is inverse square related.

A polynomial equation of each option was fitted to replace the basic model in Eq. 2 (12):

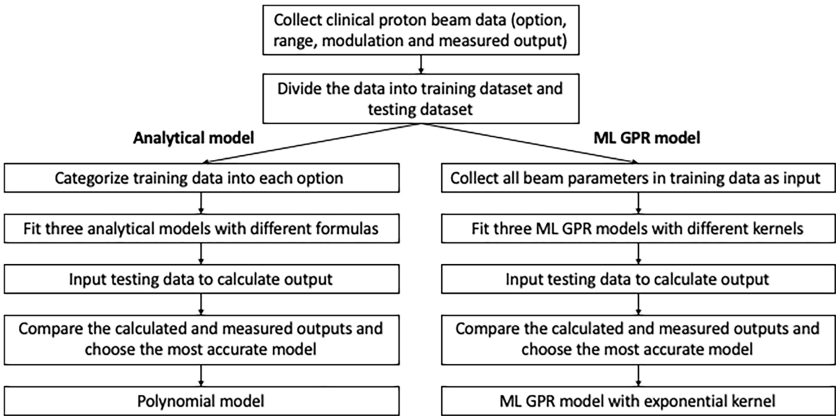


FIGURE 1 | Workflow of model fitting and testing for analytical/ML models.

$$d/MU(r(R,M)) = (p_0 + p_1 r + p_2 r^2 + p_3 r^3 + p_4 r^4) \times [s_2 + s_3(R - R_L)] \times \left(\frac{ESAD(r) - \Delta z_p}{ESAD(r) - \Delta z_p - \Delta z}\right)^2 \quad (2)$$

where s_2 and s_3 are the option-specific fitting parameters.

In terms of the fitting data, Ferguson et al. listed the values of s_0 and s_2 in different options and those are very close to unity (12). s_1 and s_3 were found to be much less than s_0 and s_2; therefore, the variation of second terms from unity could be negligible. The third term is only to correct the measurement position if the effective source is not located at the middle SOBP.

To simplify the calculation of output, we replaced the SAD setup with the SSD setup in our method, in which Δz is always zero and the third term equals unity.

Therefore, the equation of output estimation can be approximated by a quadratic Taylor polynomial in Eq. 3.

$$^{d}/_{MU} = a \times r^2 + b \times r + c, \quad (3)$$

where $r = (R - M)/M$, d/MU denotes the output and a, b, and c are the fitting parameters.

Linear Fitting Model

The linear fitting model estimates output as the function of the logarithm of R/M (Eq. 4). The rationale of choosing this model was to space out data points clustered in the low R/M region, as observed from the polynomial model. From the polynomial fitting graph, it was observed that the output variations in the low R/M region (full modulation) were larger, with a lot more data points than the high R/M region (**Figure 2**). This finding is consistent with what Sun et al. and Kim et al. reported (10, 16). This model can be expressed as

$$\log\left(^{d}/_{MU}\right) = k \times \log\left(^{R}/_{M}\right) + b \quad (4)$$

where d/MU, R, and M are the variables, and k and b are the fitting parameters.

Logarithm–Polynomial Fitting Model

The logarithm–polynomial fitting algorithm is an independent model from the previous two models, since the variables in previous models are both related to R/M. To build a model with a different variable, while still keeping the model accurate, different approaches were made and the most accurate one was selected. In this model, the output is the function of logarithm(R)/logarithm(M) in Eq. 5.

$$^{d}/_{MU} = a' \times r'^2 + b \times r' + c', \quad (5)$$

where $r' = \log(R)/\log(M)$, and a′, b′, and c′ are the fitting parameters.

Machine Learning-Based Output Estimation

Different from analytical methods, machine learning (ML) methods do not need to model option by option. Instead, they use option number, beam range, and modulation to predict the output. To test the efficacy and accuracy of ML modeling, three ML GPR models with different kernels, including exponential kernel, squared exponential kernel, and rational quadratic kernel, were used for the output calculation (17). The model is shown in Eq. 6.

$$y = h(x)\beta + f(x) \quad (6)$$

where h(x) is a set of basic functions that transform the original feature vector x into a new feature vector h(x), and f(x) models the uncertainties from the system.

GPR is a non-parametric Bayesian approach toward regression problems that can be utilized in exploration and exploitation scenarios (17, 18). It predicts the output data by incorporating prior knowledge and fit a function to the data. The Gaussian process is a set of random variables, such that any finite number of them has a joint Gaussian distribution. The mean

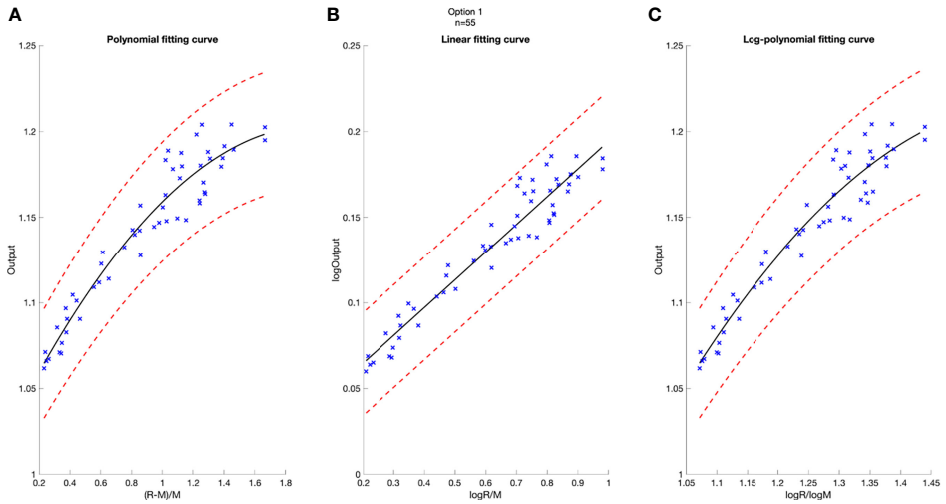


FIGURE 2 | Model-based fitting curves for Option 5, including the polynomial fitting curve **(A)**, the linear fitting curve **(B)**, and the log-polynomial fitting curve **(C)**. 3% confidence level in red dashed line. Number of data points n = 68.

function of f(x) is 0, and the covariance function is k(x, x′), that is, f(x)~GP(0, k(x, x′)).

The probability distribution of y is

$$P(y_i|f(x_i),x_i) \sim N(y_i|h(x_i)^T\beta + f(x_i),\sigma^2) \tag{7}$$

which can be written in matrix form as

$$P(y|f,X) \sim N(y_i|H\beta + f,\sigma^2 I) \tag{8}$$

where

$$X = \begin{pmatrix} x_1^T \\ \vdots \\ x_n^T \end{pmatrix}, y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, H = \begin{pmatrix} h(x_1^T) \\ \vdots \\ h(x_n^T) \end{pmatrix}, f = \begin{pmatrix} f(x_1) \\ \vdots \\ f(x_n) \end{pmatrix} \tag{9}$$

Then,

$$P(f|X) \sim N(f|0,K(X,X)) \tag{10}$$

K is the covariance matrix

$$K = \begin{pmatrix} k(x_1,x_1) & \cdots & k(x_1,x_n) \\ \vdots & \ddots & \vdots \\ k(x_n,x_1) & \cdots & k(x_n,x_n) \end{pmatrix}, \tag{11}$$

where k is the kennel function.

The kernels play very significant roles in the regression modeling and can map the features from the original values to the featuring spaces by involving the latent variables. In this model, three kernel functions were used, including exponential kernel, squared exponential kernel, and rational quadratic kernel.

After the training process, 5-fold cross-validation was performed to prevent overfitting. Then, the obtained models were evaluated using the parameters from the testing sets.

Robustness of Models Related to Sampling Numbers

The robustness of output models could be impacted by the number of data fed into the model. The evaluation of minimum number of data necessary for a robust model was conducted by comparing different model outputs with increasing number of inputs. Models were built with different sampling numbers, randomly selected from the original training dataset. The sampling numbers ranged from 10 to the number of training datasets. Each time a new model was generated, and the mean average percentage error (MAPE) was calculated to evaluate the differences between predicted outputs and the corresponding measurements. The comparisons were performed per option for polynomial models.

Difference Between Analytical and ML Models

Analytical fitting models and ML models play independently in output estimation, as can be seen in **Figure 1**. Although both polynomial model and ML GPR model with exponential kernel may be robust and accurate enough to be within clinical tolerance, the predicted output from the two models can be different. Also, a cross check of output is essential to verify the accuracy and effectiveness of the two methods. To assess the difference between the two models, 1,000 random points within the range and modulation of each option were generated to estimate output by using these two models, and the MAPE of predicted outputs and the corresponding measurements for the two models were calculated for each point. The estimated output with different combinations of range and modulation were compared between these two models and also to the measured output.

RESULTS

The total clinical fields including training data and testing data were categorized into 24 options (**Table 1**). In this table, Options 4, 6, 7, 8, 9, 10, 20, 21, and 22 were mostly used in clinic with sample numbers larger than 80.

Accuracy Analysis of Output Estimation

A deviation of 3% was used as tolerance in clinical output estimation. The analytical fitting curve of Option 5 is presented as an example in **Figure 2** to show the absolute error of modeling output relative to the measured value. The output is plotted as a function of R and M, with polynomial fit in **Figure 2A**, linear fit in **Figure 2B**, and logarithm–polynomial fit in **Figure 2C**. The red dashed line represents ±3% from the predicted output. The blue scattered marks representing the real measurements are all within ±3% of the predicted value in Option 5, indicating accurate prediction for three models. The coefficient of determination of each fitting curve is provided on **Figure 2**.

The histograms of the relative deviation of all 24 options are categorized in **Figure 3**. Compared to the other two analytical models, the polynomial fitting model provided a better agreement with measurement data, with all deviations within ±3%. In ML GPR models, the exponential kernel showed a more accurate output estimation than the other two with less than ±2% deviation from the measurement. The 5-fold cross-validation results are shown in **Figure 4**, which showed a similar performance as the training model. In addition, the testing data were imported into analytical and ML GPR models to verify the effectiveness and accuracy of output estimation (**Figure 5**). It was observed that the polynomial fitting model still provided a good output estimation within 3% deviation, and all the ML models also exhibited deviation within ±3%. To summarize, the polynomial model and ML GPR model with exponential kernel showed the best performance among all 6 models, with less than ±3% deviation from all measurement data.

Minimum Number of Fields Needed for Polynomial Model and ML GPR Model With Exponential Kernel Model

The trend of MAPE of models compared to measurements is shown in **Figure 6**. For the polynomial model, since it is specific to each option, Option 9 and Option 22 were chosen as the representatives of the polynomial model because of higher sample numbers available, as shown in **Figure 6A**. The trend of MAPE for the ML GPR model with the exponential kernel is shown in **Figure 6B**, with data from all options. As observed in this figure, the relative error in Options 9 and 22 both converged to be around 1% or less once 20 data points were used for building the polynomial model. For the ML GPR model with the exponential kernel, the convergence of MAPE was reached at around 400 data points, regardless of the option.

Evaluation of Difference Between Analytical Models and ML Models

Comparisons of output estimation between the polynomial fitting model and the ML GPR model with the exponential kernel for Options 8 and 22 are shown in **Figure 7**. MAPE for 1,000 randomly generated points between the two models are shown with corresponding R and M. Measurement data are marked as pink scattered points overlaid on the figure. It is shown in **Figure 7** that the two models agreed well in the regions where there were measurement data, with MAPE less than 2%. Considerable differences in the outputs were observed beyond the measurement region. More intuitive figures are shown in **Figure 7B** where the general trend of outputs splits in between polynomial and ML models with the decrease of modulation, and this split in 3D graphs illustrates the trend of difference between the two models.

DISCUSSION

Patient-specific output estimation in a double scattering proton machine has to rely on manual measurement, which can be a labor-intensive and error-prone process, so it is valuable to build models to second-check manual measurement, and the ultimate goal of the study is to build an automatic output estimation and MU determination process.

The output estimation derived from three analytical fitting models and three ML GPR models with different kernels was demonstrated and compared to the measurements. The polynomial fitting model and ML GPR model with the exponential kernel with the best performance were chosen. In terms of the distribution in the histogram, the polynomial fitting method provided the most accurate output estimation in analytical methods and the ML GPR model with the exponential kernel could provide more accurate output prediction than the other two ML GPR models. Also, the relative errors between the estimated and measured output for the polynomial fitting model and ML GPR model with exponential kernel were always within ±3% in both training data and testing data. Therefore, it is proof that those two models could be adopted as the output estimation models. Also, since they are two independent models, it is suggested that they can be used as second-check tools for clinical measurement, and also a cross-check tool for each other.

Compared with other models reported in literature, our models are more stable and accurate in output estimation. It has been reported by Kooy et al. (8, 9) that there was a large deviation between calculated and measured output in full range and full modulation. Sun et al. (10) also reported an apparent difference (>3%) even using their ML models. Comparatively, the advantage of our polynomial model and the ML GPR model is that the difference between the measured and calculated output is within 3%, which satisfies the clinical requirement and is thereby reliable for clinical use.

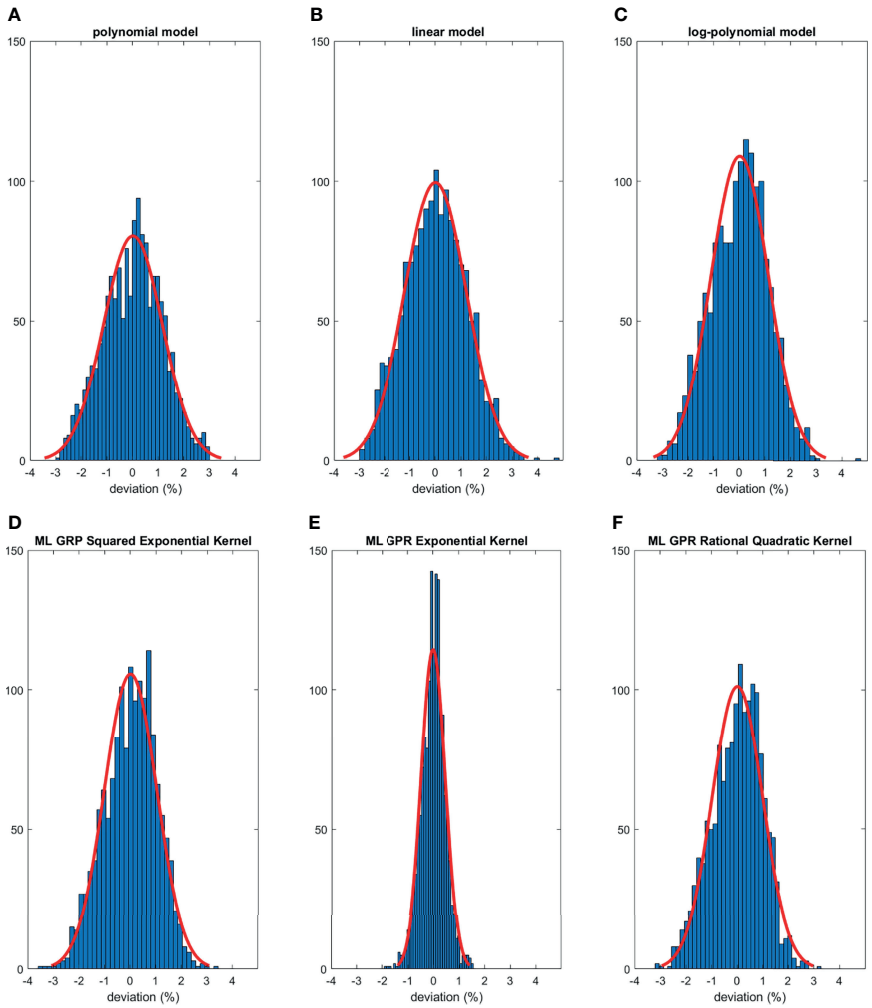


FIGURE 3 | Histograms of percent difference between analytical/ML GPR models and measurements using training data.

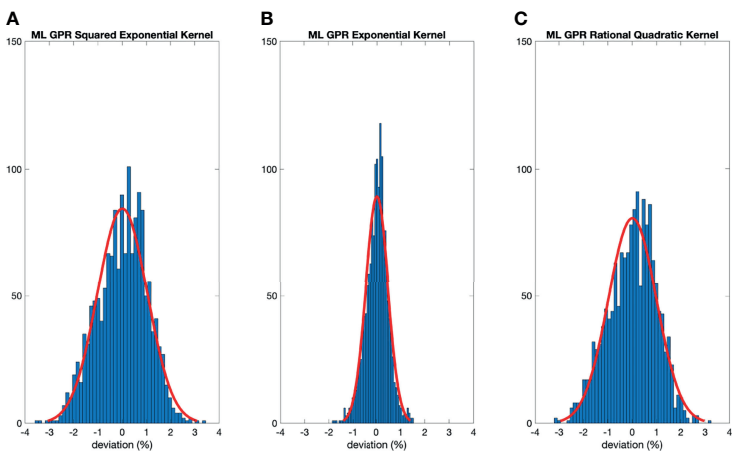


FIGURE 4 | Histograms of percent difference between the ML GPR models and measurements using 5-fold cross-validation.

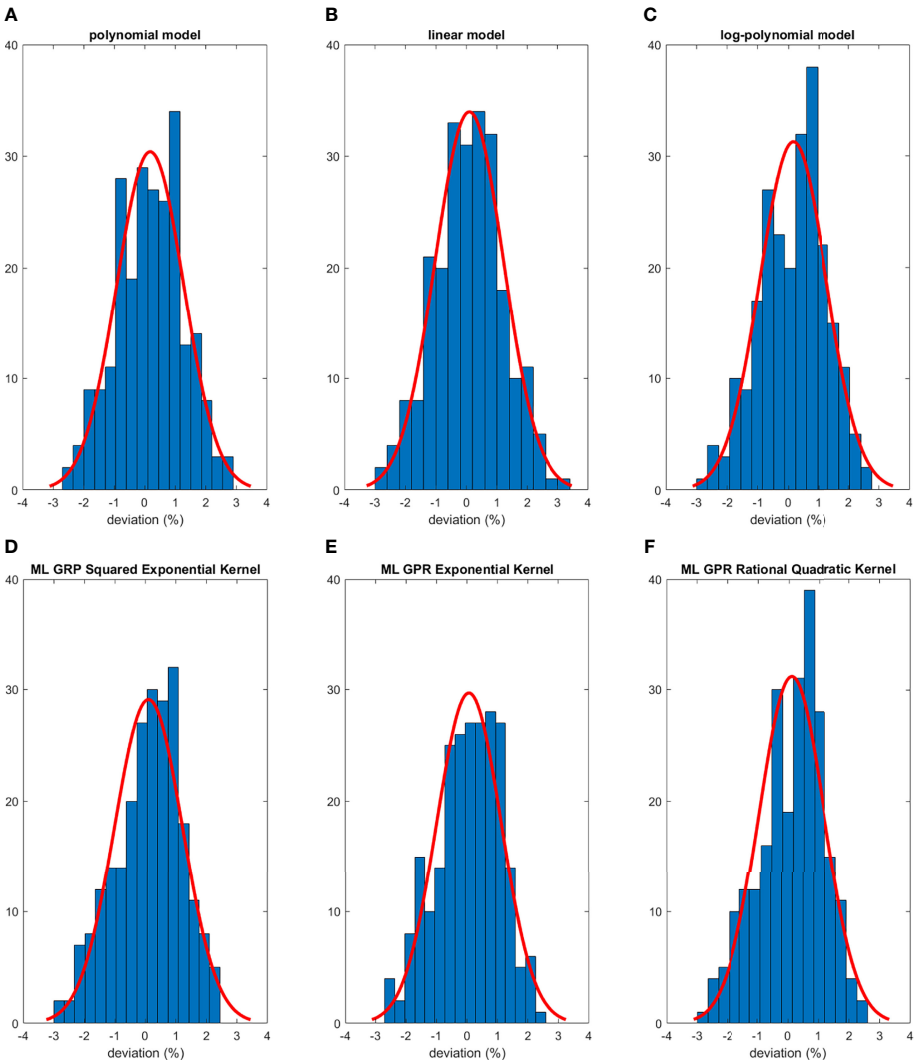


FIGURE 5 | Histograms of percent difference between analytical/ML models and measurements using testing data.

The polynomial model is an expansion of Kooy's empirical formula using Taylor series. This approach is similar to the equation developed by Ferguson et al. (12) but using a lower order of polynomials, thus simplifying the equation. As shown in the results, the performance of our quadratic polynomial model is comparable to their quartic polynomial model, as both models can achieve an accuracy of ±3%. The reason for the polynomial model to outperform the logarithm–polynomial model is probably because the variable of the polynomial model is $r = (R–M)/M$, same as that in Kooy's empirical model (8), which is a theoretical equation derived from physical properties of proton beam lines, while the variable of the logarithm–polynomial model is $r' = \log(R)/\log(M)$. The result of better performance of the polynomial model proves that the proton output is truly related to R/M. The intention of building the linear model was to space out data points clustered in the low R/M region, as observed from the polynomial model, to avoid overfitting. However, as the results show, the linear curve cannot simulate the trend of training data, thus resulting in a larger deviation. As a result, the other two analytical models cannot predict the output with similar accuracy as the polynomial model.

The reason to choose the GPR model for the output prediction is that the kernel functions can be used. Prior knowledge and specifications about the shape of the model can be added by selecting different kernel functions. Meanwhile, the Gaussian process directly captures the model uncertainty. In this way, the output model can be described as a distribution rather than approximated values. The exponential kernel was based on the assumption that the Euclidean distances between different data points were Laplacian distributed. The squared exponential kernel used another assumption that the Euclidean distances were normally distributed. The rational quadratic kernel can be seen as a scale mixture of squared exponential kernels with different characteristic length-scales (17). In this paper, the

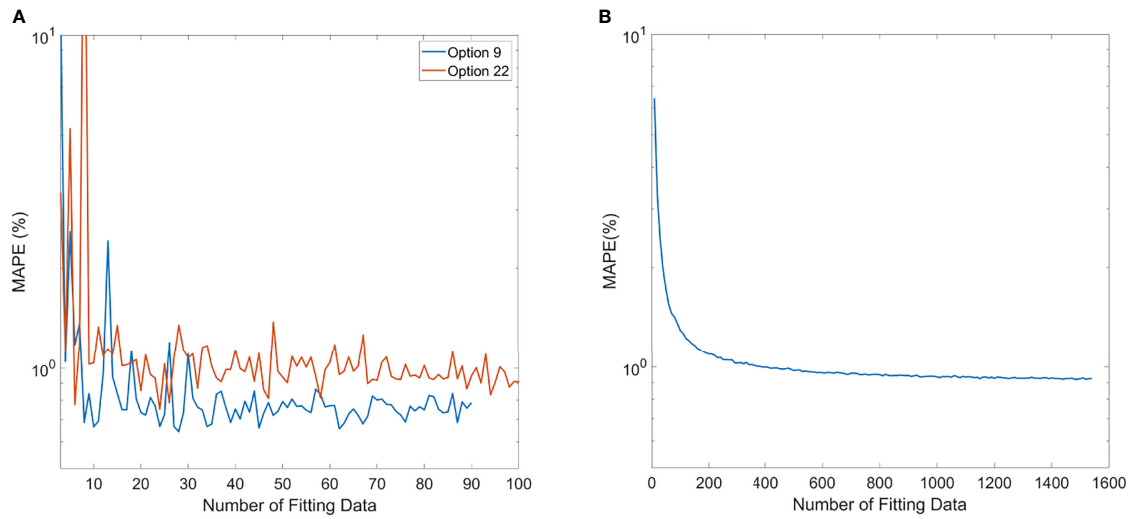


FIGURE 6 | Mean average percentage error (MAPE) of polynomial model **(A)** and ML exponential kernel model **(B)** with the increase in fitting data.

distributions of modulations and ranges were close to sparse, so the Laplacian distribution might be a better option, resulting in that the exponential kernel exhibited better performance.

The number of data needed for establishing a robust polynomial model was estimated by the MAPE trend with increasing data points. Options 9 and 22 were shown as an example that the MAPE converged once the number of training data increased to 20. This gives a simple guidance on the number of data necessary to build an accurate polynomial model for an option. Among all options, some of them were rarely used, especially the deep options (option 13 with 3 beams, option 14 with 12 beams, option 15 with 37 beams, option 17 with 7 beams, and option 18 with 19 beams). This is because clinically we tend to plan the proton beam to penetrate through a shorter path if possible, leading to lower usage of deep-ranged options. For those options with fewer data points, a polynomial-based output model would not be recommended. Instead, manual measurement would be required, until enough data points are accumulated.

For the ML GPR model with exponential kernel, convergence of MAPE to 1% was observed after the input of 400 fields. This needs to be clarified that when building the ML models, range and modulation as well as the option number were inputs to the model. Sun et al. also estimated the minimum number of fields needed for the ML cubist model (10). In their study, the mean absolute error converged to 0.7% after 1,200 data points. Their learning curves also showed a mean absolute error around 1% at 400 samples. Since the ML model does not discriminate different options, and some options may have fewer data samples than others, validation of accuracy of the model in all options is needed before clinic implementation.

The polynomial fitting model and ML method could be used as independent secondary check, and eventually the primary output estimation, replacing measurements. This requires the assessment of the agreement between two models. From **Figure 7**, the MAPE between two models was less than 2% if the data points lay within the region where there existed measurement data. Beyond this region (e.g., M = 2–6 cm in Option 4), the two models showed obvious different trends with increasing differences, which indicated that the user must evaluate the accuracy of output prediction with extra measurements; otherwise, the model cannot be used beyond the region with real measurements. It is suggested that the models should only be trusted to replace measurements with judgment that the beamline (R and M) falls within the region with enough measurement data.

Even though the polynomial fitting model and ML GPR model with the exponential kernel proved their feasibility for output estimation, it is still essential to pay attention to MU determination, as not only is the MU related to output but also its accuracy is related to the verification point dose. The accurate selection of the verification point is pivotal in MU determination. It is recommended to perform a sanity check on the MU of a clinical plan. A simple sanity check is to compare calculated MU to the prescription dose multiplied by the field weighting. The rationale of this sanity check is because the output is always close to 1. Future work includes automatic MU determination with Eclipse Scripting, to help get rid of the uncertainty of manual selection of verification points. Nevertheless, whether the output is measured or modeled, the MU must be verified prior to clinical treatment.

CONCLUSIONS

MU determination including output measurement is one of the most time-consuming and complicated works in patient QA for double-scattering proton machines. Compared with the output

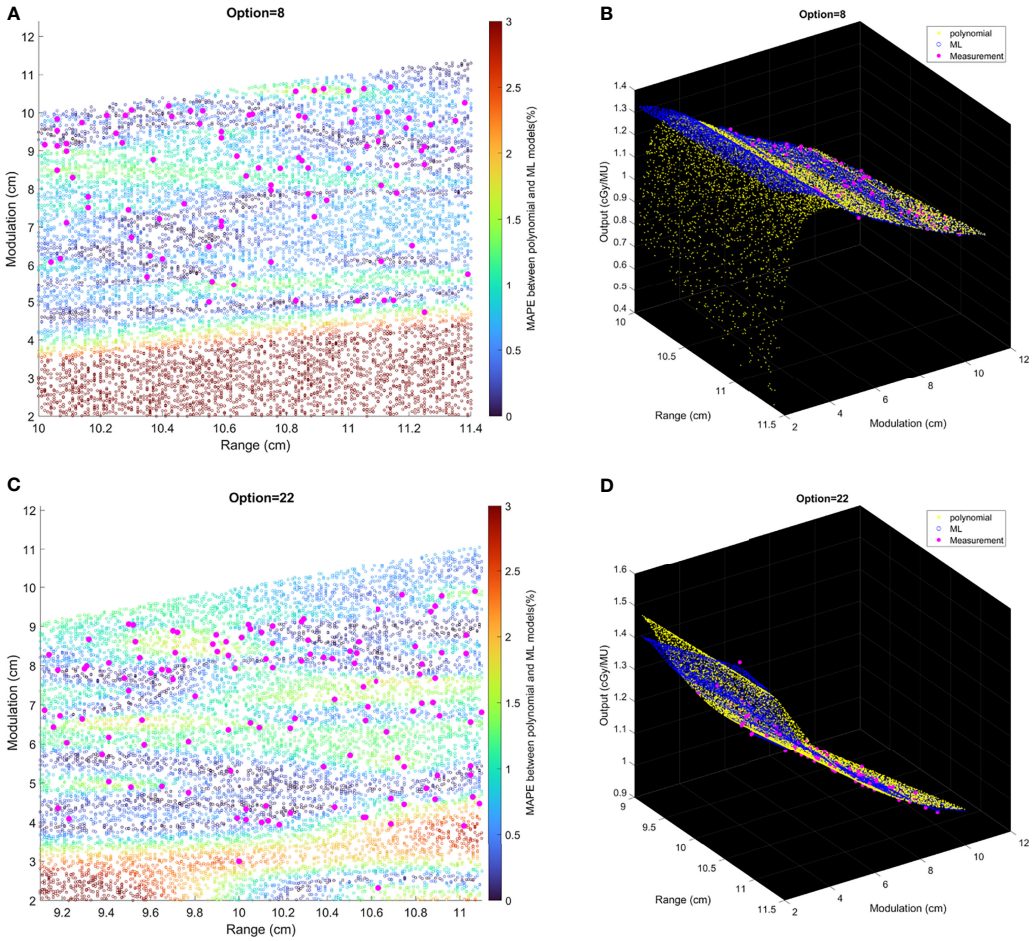


FIGURE 7 | The relative difference of output estimation with R and M. **(A, C)** show the differences of two models. The solid pink points are the measurement data, and the circles are MAPE between two models from the random data. **(B, D)** are the differences in 3D graphs that illustrate the trend of difference between two models.

measurement, analytical fitting models or ML models are more efficient to provide output estimation. Out of the six models presented in this paper, the polynomial model and ML GPR model with the exponential kernel both show accurate estimation, and the accuracy meets the clinical requirement (within ±3%). The minimum number of data needed to build a robust model is provided, although it is suggested that the validation of accuracy of the model is needed before clinical implementation. These independent output estimations can serve as second-check tools for measurements and have potential to replace the measurement as part of the standard MU determination procedure. The models exhibit robustness within the region where there exist measurement data, and the accuracy beyond the region with real measurements must be evaluated with extra measurements.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

Ethical review and approval were not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

XW, YiZ, and KN led the conception and design of the study. JZ, XW, BL, YiZ, TC, and KN contributed to the acquisition of data. JZ, XW, TC, YaZ, and CM contributed to the analysis, data modelling, and interpretation of data. JZ, XW, TC, and NY drafted and revised the article. All authors contributed to the article and approved the submitted version.

REFERENCES

1. Wilson RR. Radiological Use of Fast Protons. *Radiology* (1946) 47(5):487–91. doi: 10.1148/47.5.487
2. Weber DC, Trofimov AV, Delaney TF, Bortfeld T. A Treatment Planning Comparison of Intensity Modulated Photon and Proton Therapy for Paraspinal Sarcomas. *Int J Radiat Oncol Biol Phys* (2004) 58(5):1596–606. doi: 10.1016/j.ijrobp.2003.11.028
3. Kim TH, Lee WJ, Woo SM, Oh ES, Youn SH, Jang HY, et al. Efficacy and Feasibility of Proton Beam Radiotherapy Using the Simultaneous Integrated Boost Technique for Locally Advanced Pancreatic Cancer. *Sci Rep* (2020) 10 (1):21712. doi: 10.1038/s41598-020-78875-1
4. Rana S, Simpson H, Larson G, Zheng Y. Dosimetric Impact of Number of Treatment Fields in Uniform Scanning Proton Therapy Planning of Lung Cancer. *J Med Phys* (2014) 39(4):212–8. doi: 10.4103/0971-6203.144483
5. Paganetti H. *Proton Therapy Physics*. Boca Raton, FL:CRC press (2018).
6. Farr JB, Mascia AE, Hsi WC, Allgower CE, Jesseph F, Schreuder AN, et al. Clinical Characterization of a Proton Beam Continuous Uniform Scanning System With Dose Layer Stacking. *Med Phys* (2008) 35(11):4945–54. doi: 10.1118/1.2982248
7. Lansonneur P, Mammar H, Nauraye C, Patriarca A, Hierso E, Dendale R, et al. First Proton Minibeam Radiation Therapy Treatment Plan Evaluation. *Sci Rep* (2020) 10(1):7025–32. doi: 10.1038/s41598-020-63975-9
8. Kooy HM, Schaefer M, Rosenthal S, Bortfeld T. Monitor Unit Calculations for Range-Modulated Spread-Out Bragg Peak Fields. *Phys Med Biol* (2003) 48 (17):2797–808. doi: 10.1088/0031-9155/48/17/305
9. Kooy HM, Rosenthal SJ, Engelsman M, Mazal A, Slopsema RL, Paganetti H, et al. The Prediction of Output Factors for Spread-Out Proton Bragg Peak Fields in Clinical Practice. *Phys Med Biol* (2005) 50(24):5847–56. doi: 10.1088/0031-9155/50/24/006
10. Sun B, Lam D, Yang D, Grantham K, Zhang T, Mutic S, et al. A Machine Learning Approach to the Accurate Prediction of Monitor Units for a Compact Proton Machine. *Med Phys* (2018) 45(5):2243–51. doi: 10.1002/mp.12842
11. Lin L, Shen J, Ainsley CG, Solberg TD, McDonough JE. Implementation of an Improved Dose-Per-MU Model for Double-Scattered Proton Beams to Address Interbeamline Modulation Width Variability. *J Appl Clin Med Phys* (2014) 15(3):4748. doi: 10.1120/jacmp.v15i3.4748
12. Ferguson S, Ahmad S, Jin H. Implementation of Output Prediction Models for a Passively Double-Scattered Proton Therapy System. *Med Phys* (2016) 43 (11):6089. doi: 10.1118/1.4965046
13. Sahoo N, Zhu XR, Arjomandy B, Ciangaru G, Lii M, Amos R, et al. A Procedure for Calculation of Monitor Units for Passively Scattered Proton Radiotherapy Beams. *Med Phys* (2008) 35(11):5088–97. doi: 10.1118/1.2992055
14. Kang J, Schwartz R, Flickinger J, Beriwal S. Machine Learning Approaches for Predicting Radiation Therapy Outcomes: A Clinician's Perspective. *Int J Radiat Oncol Biol Phys* (2015) 93(5):1127–35. doi: 10.1016/j.ijrobp.2015.07.2286
15. Musolino SV. *"Absorbed Dose Determination in External Beam Radiotherapy: An International Code of Practice for Dosimetry Based on Standards of Absorbed Dose to Water; Technical Reports Series No. 398"*. Vienna, Austria: LWW (2001).
16. Kim DW, Lim YK, Ahn SH, Shin J, Shin D, Yoon M, et al. Prediction of Output Factor, Range, and Spread-Out Bragg Peak for Proton Therapy. *Med Dosim* (2011) 36(2):145–52. doi: 10.1016/j.meddos.2010.02.006
17. Rasmussen CE. Gaussian Processes in Machine Learning. In: *Summer School on Machine Learning*. Berlin, Heidelberg:Springer. (2003) 63–71.
18. Barber D. *Bayesian Reasoning and Machine Learning*. New York: Cambridge University Press (2012).

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: 16 February 2022
doi: 10.3389/fonc.2022.768152

Radiomic and Dosiomic Features for the Prediction of Radiation Pneumonitis Across Esophageal Cancer and Lung Cancer

Chanon Puttanawarut[1,2*], *Nat Sirirutbunkajorn*[3], *Narisara Tawong*[3], *Chuleeporn Jiarpinitnun*[3], *Suphalak Khachonkham*[3], *Poompis Pattaranutaporn*[3] *and Yodchanan Wongsawat*[2*]

[1] Chakri Naruebodindra Medical Institute, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Samut Prakan, Thailand, [2] Brain-Computer Interface Laboratory, Department of Biomedical Engineering, Faculty of Engineering, Mahidol University, Nakhorn Pathom, Thailand, [3] Department of Diagnostic and Therapeutic Radiology, Faculty of Medicine, Ramathibodi Hospital, Mahidol University, Bangkok, Thailand

OPEN ACCESS

Edited by:
Jose Eduardo Villarreal Barajas, Royal Devon and Exeter Hospital, United Kingdom

Reviewed by:
Alessandra Retico, Ministry of Education, Universities and Research, Italy
Valerio Nardone, University of Campania Luigi Vanvitelli, Italy

***Correspondence:**
Chanon Puttanawarut
chanon.puw@student.mahidol.edu
Yodchanan Wongsawat
yodchanan.won@mahidol.ac.th

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

Received: 31 August 2021
Accepted: 13 January 2022
Published: 16 February 2022

Citation:
Puttanawarut C, Sirirutbunkajorn N, Tawong N, Jiarpinitnun C, Khachonkham S, Pattaranutaporn P and Wongsawat Y (2022) Radiomic and Dosiomic Features for the Prediction of Radiation Pneumonitis Across Esophageal Cancer and Lung Cancer. Front. Oncol. 12:768152. doi: 10.3389/fonc.2022.768152

Purpose: The aim was to investigate the advantages of dosiomic and radiomic features over traditional dose-volume histogram (DVH) features for predicting the development of radiation pneumonitis (RP), to validate the generalizability of dosiomic and radiomic features by using features selected from an esophageal cancer dataset and to use these features with a lung cancer dataset.

Materials and Methods: A dataset containing 101 patients with esophageal cancer and 93 patients with lung cancer was included in this study. DVH and dosiomic features were extracted from 3D dose distributions. Radiomic features were extracted from pretreatment CT images. Feature selection was performed using only the esophageal cancer dataset. Four predictive models for RP (DVH, dosiomic, radiomic and dosiomic + radiomic models) were compared on the esophageal cancer dataset. We further used a lung cancer dataset for the external validation of the selected dosiomic and radiomic features from the esophageal cancer dataset. The performance of the predictive models was evaluated by the area under the curve (AUC) of the receiver operating characteristic curve (ROCAUC) and the AUC of the precision recall curve (PRAUC) metrics.

Result: The ROCAUCs and PRAUCs of the DVH, dosiomic, radiomic and dosiomic + radiomic models on esophageal cancer dataset were 0.67 ± 0.11 and 0.75 ± 0.10, 0.71 ± 0.10 and 0.77 ± 0.09, 0.71 ± 0.11 and 0.79 ± 0.09, and 0.75 ± 0.10 and 0.81 ± 0.09, respectively. The predictive performance of the dosiomic- and radiomic-based models was significantly higher than that of the DVH-based model with respect to esophageal cancer. The ROCAUCs and PRAUCs of the DVH, dosiomic, radiomic and dosiomic + radiomic models on the lung cancer dataset were 0.64 ± 0.18 and 0.37 ± 0.20, 0.67 ± 0.17 and 0.37 ± 0.20, 0.67 ± 0.16 and 0.45 ± 0.23, and 0.68 ± 0.16 and 0.44 ± 0.22, respectively. On the lung cancer dataset, the predictive performance of the radiomic and

dosiomic + radiomic models was significantly higher than that of the DVH-based model. However, the PRAUC of the dosiomic-based model showed no significant difference relative to the corresponding RP prediction performance on the lung cancer dataset.

Conclusion: The results suggested that dosiomic and CT radiomic features could improve RP prediction in thoracic radiotherapy. Dosiomic and radiomic feature knowledge might be transferrable from esophageal cancer to lung cancer.

Keywords: radiotherapy, dosiomic, radiomic, machine learning, DVH, radiation pneumonitis, esophageal cancer

INTRODUCTION

In thoracic radiation therapy, organs at risk, such as the lungs, are the limiting factors of radiation treatment due to radiation toxicity. Radiation pneumonitis (RP) is one type of lung toxicity. Many studies have tried to develop RP prediction models based on dose volume histograms (DVHs) and/or the clinical profiles of patients (1–3). However, DVHs and clinical factors are only some of the many pieces of information that can be extracted from patients.

Recently, quantitative image features such as the dosiomic (quantitative features of dose distribution) and/or radiomic features of computed tomography (CT) images have been reported to improve the performance of prediction models for radiation toxicity (4–8). Dosiomic features contain more dose distribution information than DVH features and have been shown to be able to improve toxicity prediction in radiation therapy. Information that can be used for the prediction of RP can also be found in CT images. For example, interstitial lung disease was found to be a risk factor for RP (9–11). RP prediction models for lung cancer have also been shown to benefit from the use of radiomic features obtained from CT images (6–8). The quantitative imaging features of fluorine 18 fluorodeoxyglucose (FDG) positron emission tomography (PET)/CT were previously studied in esophageal cancer patients (12). While the radiomic features from CT were not found to be significant, the radiomic features from FDG-PET SUV were significantly associated with grade 2 RP. However, only a subset of radiomics features in CT images was explored.

Studies of dosiomic and radiomic features can result in feature selection bias, as demonstrated by a systematic review by Chalkidou et al., who generated 100 random features and found that 10% of the features were significant predictors (13). Furthermore, some random variables achieved higher performance metric scores than other significant features, as reported in other studies. To reduce the false-positive rates in radiomic studies, external validation was recommended (14–18).

This study aimed to investigate the benefit of using radiomic and dosiomic features in an RP prediction model for esophageal cancer patients. We compared four predictive models with DVH features, dosiomic features, radiomic features and combined dosiomic and radiomic features. Furthermore, to investigate the generalizability of dosiomic and radiomic features, we incorporated an external dataset with lung cancer patients and investigated a predictive model using features selected from esophageal cancer data.

MATERIAL AND METHODS

Data

The CT images, ROIs, and 3D dose distributions of 333 esophageal cancer patients and 110 lung cancer patients >15 years of age who were treated with radiation therapy from 2011 to 2019 were extracted from the Varian Eclipse v16.1 treatment planning system (TPS) (Varian Medical Systems, Palo Alto, CA) at the Ramathibodi Hospital at Mahidol University. The study was approved by the ethical committee of the Ramathibodi Hospital at Mahidol University (IRB MURA2021/283). Patients with previous histories of thoracic radiation therapy, diagnoses of interstitial lung disease, follow-up times under one year, no treatment data or diagnoses of lung metastasis within one year were excluded from the study. After exclusion, 101 patients and 93 patients had esophageal cancer and lung cancer, respectively. The clinical and treatment characteristics are shown in **Table 1**. All dose distribution were calculated by Anisotropic Analytical Algorithm (AAA) from Varian Eclipse TPS. The script for the extraction of the treatment plan from the Varian Eclipse TPS based on the Eclipse Scripting Application Programming Interface (ESAPI) is available at GitHub at https://github.com/44REAM/ExportFractionDose.git.

Radiation pneumonitis grading was performed by radiation oncologists based on the National Cancer Institute Common Terminology Criteria for Adverse Events version 5.0 (CTCAE v5.0). In practice, grade 0 RP was defined as negative for RP, grade 1 RP was defined as patients with symptoms or radiographic features without the need for steroids. Grade 2 RP was defined as patients requiring steroids or with symptoms that interfered with daily activities. Grade 3 RP was defined as patients requiring oxygen and steroids. Grade 4 RP was defined as patients requiring intubation. The aim of this study was to evaluate the performance of dosiomic and radiomic features for prediction of presence of any RP. However, due to unavailability of grade 1 RP data in lung cancer dataset, the positive class for esophageal cancer was defined as grade 1 or above, while for lung cancer, positive class was defined as grade 2 or above.

Equivalent Dose in 2 Gy Fractions

Dose distributions were extracted as fractions. The dose distributions of fractions and voxels were referred to as "doses per fraction per voxel". The equivalent dose in the 2 Gy fraction a voxel with EQD2 fractions was calculated as follows (19):

TABLE 1 | Clinical and treatment characteristics of esophageal and lung cancer patients.

Clinical and Treatment Characteristics	Esophageal Cancer	Lung Cancer
	Median (Range)/n (%)	
Age	61 (26–93)	67 (32–87)
Sex		
Male	89 (0.88%)	63 (68%)
Female	12 (0.12%)	30 (32%)
Stage		
1	4 (4%)	31 (34%)
2	3 (3%)	58 (62%)
3	71 (70%)	3 (3%)
4	23 (23%)	1 (1%)
Prescription dose	50.4 (30.0–60.0)	59.4 (43.2–66)
Prescription fraction	1.8 (1.8–3.0)	2.0 (1.8–2.0)
Treatment setting		
CCRT	95 (94%)	91 (98%)
RT	6 (6%)	2 (2%)
RT modality		
3D conformal RT	78 (77%)	38 (41%)
IMRT/VMAT	9 (9%)	27 (29%)
Combine	14 (14%)	28 (30%)
RT aim		
Preoperative	47 (47%)	0 (0%)
Postoperative (adjuvant)	1 (1%)	0 (0%)
Definitive	49 (48%)	93 (100%)
Palliative	4 (4%)	0 (0%)
RP grade		
0	38 (38%)	77 (83%)
1	58 (57%)	
2	5 (5%)	14 (15%)
3	0 (0%)	2 (2%)
4	0 (0%)	0 (0%)

$$D_{EQD2} = \sum_{i}^{N} \frac{d_{i,k} + d_{i,k}^2/(\alpha/\beta)}{1 + 2/(\alpha/\beta)}.$$

The value of the α/β ratio in the equation was assumed to be 3 (20–26). The variable $d_{i,j}$ is the dose per fraction per voxel, i is the number of fractions and k is the number of voxels. The equation above was suitable for our dataset because of its compatibility with different doses per fraction per voxel. Although we used a similar prescription fraction size (1.8–3 Gy per fraction), the actual doses the patient received in different locations and with different fractions might have been different. For example, the first fraction may have been delivered by an antero-posterior beam, and the second fraction may have been delivered by 2 lateral beams, resulting in different doses per fraction for different voxels.

Features

Resampling to 1.5 × 1.5 × 1.5 mm^3 by b-spline algorithm was performed for all dose distributions and CT images. ROIs was resampled by nearest neighbor algorithm to match CT image. All CT images were free-breathing CT scans. The mean lung doses (MLDs), the volumes of the lungs that received doses greater than x Gy, Vx (ranging from V5 to V70 over 5 Gy steps), were used as DVH features. The Pyradiomics library in Python (27), which contains the most common feature definitions based on the Imaging Biomarker Standardization Initiative (IBSI) (28), was used to extract dosiomic and radiomic features. Dosiomic features were extracted from the resampled dose distribution. Both texture features and first-order features were then extracted from the CT images (radiomic) and dose distributions (dosiomic). The dosiomic features were extracted from lung ROIs, and the radiomics features were extracted from the lung ROIs of patients who received doses greater than x Gy for x = 10 and 20. The lung ROIs for esophageal cancer were defined as all the bilateral areas of the lungs, and that for lung cancer was defined as all the bilateral areas of the lungs minus the gross tumor volume (GTV). All ROIs were segmented by different physicians.

The dosiomic and radiomic features included in this study were based on the Pyradiomics library. However, we excluded one feature among the first-order statistics of the dosiomic features, "mean dose", because this feature was redundant with the DVH features. All features in this study were based on 51 (17 × 3) first-order statistics features and 225 (61 × 3) texture features. The dose distributions and CT images were further processed before the calculation of dosiomic and radiomic features. The dose distribution gray-level intensity was binned to the 100 Gy level with a fixed bin size of 1 Gy. The CT image Hounsfield units (HUs) above 100 HU and below −1,000 HU were set to zero, resulting in an HU range of [−1,000 100]. Each HU value was then converted to a positive number in the range [0 1,100] and binned with a fixed bin size of 50. The texture features were based on the gray level cooccurrence matrix (GLCM) with 72 (24 × 3) features, gray level run length matrix (GLRLM) with 48 (16 × 3) features, gray level size zone matrix (GLSZM) with 48 (16 × 3) features and neighborhood gray tone difference matrix (NGTDM) with 15 (5 × 3)

features. Both the DVH and dosiomic features were extracted from lung the ROIs from dose distributions with or without corrections to EQD2. All features were standardized to zero mean and unit variance. In summary, 15 DVH features, 78 dosiomic features and 156 radiomic features were extracted from each patient. The complete list of features is provided in **Supplementary Table 2**.

Model Building

The predictive models for radiation pneumonitis were built separately for esophageal cancer patients and lung cancer patients. An overview of the process is shown in **Figure 1**.

a. First, we performed feature selection *via* univariate analysis. A univariate logistic regression model was developed for all features using the entire esophageal dataset. Features that had p-values ≥0.1 were eliminated. We further trained the logistic regression model without regularization by repeat 5-fold cross-validation 50 times for the esophageal patients on the entire esophageal dataset. The top 10 features corresponding to the average area under the receiver operating characteristic curve (ROCAUC) from each feature group (DVH, dosiomic and radiomic) were selected for multivariate analysis.
b. The esophageal data (500 instances) were randomly separated into a training set (80%) and test set (20%). We trained the following models: DVH (10 features), dosiomic (10 features), radiomic (10 features) and dosiomic + radiomic (20 features) models. Multivariate logistic regression with L2 norm regularization was used. The L2 norm was utilized to prevent model overfitting. The hyperparameter (regularization strength) was determined by inner cross-validation (CV) by 250 Monte Carlo CV (inner training set 80%; validation 20%) runs to maximize the average ROCAUC on the validation set using grid search. The minority class in the inner training set was oversampled by the synthetic minority oversampling technique (SMOTE) to equalize the two classes.
c. Next, we trained multivariate logistic regression models using the DVH, dosiomic and radiomic features selected previously from the esophageal dataset and the lung cancer dataset. We also selected DVH features from the lung dataset (DVH_{lung}) and trained the logistic regression model as described above to introduce some bias to favor the DVH features.

The model performance was evaluated by the mean ROCAUC. We also calculated the AUCs of the precision-recall curves (PRAUC) because the ROCAUC could be biased when used with imbalanced datasets (29). The mean, standard deviation (SD) and 10th–90th percentiles of the ROCAUCs and PRAUCs for the test set results of 500 models in each group were calculated. A Z-test was used to test the statistical significance of the mean AUC between each pair models. Statistical analyses were performed using the Python and SciPy packages (30). A p-value <0.05 was considered significant. For more details of the model building, we refer to **Supplementary S1**.

RESULTS

The selected features for the DVH, dosiomic and radiomic groups are shown in **Supplementary Table 1**. For the DVH features selected from the esophageal cancer dataset, only V45 had a p-value less than 0.1 in the univariate analysis of the lung cancer dataset. For DVH_{lung}, only 3 features with p-values less than 0.1 were selected (V45, V50, and V55). The univariate analysis of lung cancer showed that 5 of 10 and 8 of 10 features had p-values less than 0.1 in the dosiomic and radiomic analyses, respectively (**Supplementary Table 1**).

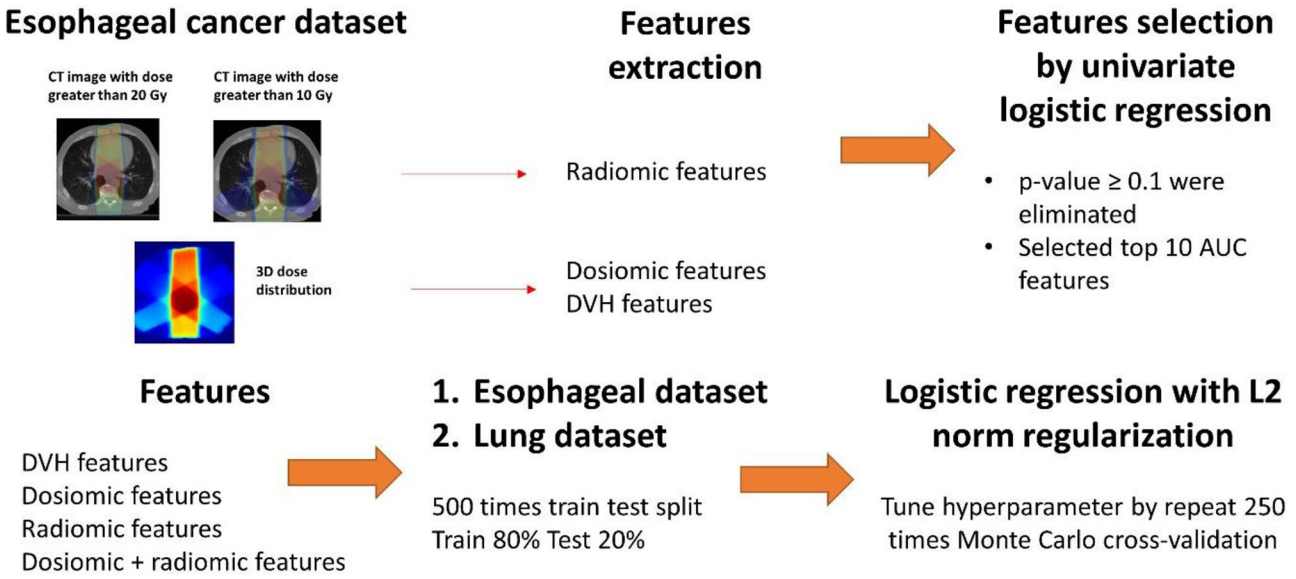


FIGURE 1 | Overview of the process from inputting data to model training.

The ROCAUC and PRAUC curves are shown in **Figure 2**. For the esophageal dataset, the model based on the DVH features resulted in an ROCAUC of 0.67 ± 0.11 and a PRAUC of 0.75 ± 0.10. The model based on dosiomic features resulted in an ROCAUC of 0.71 ± 0.10 and a PRAUC of 0.77 ± 0.09. The model based on radiomic features resulted in an ROCAUC of 0.71 ± 0.11 and a PRAUC of 0.79 ± 0.09. The model based on dosiomic + radiomic features resulted in an ROCAUC of 0.75 ± 0.10 and a PRAUC of 0.81 ± 0.09. The results of esophageal cancer dataset are included in **Table 2**. The ROCAUC and PRAUC of the model using dosiomic + radiomic features were significantly higher than those of the models with DVH, dosiomic and radiomic features (p-value <0.05). The AUCs of both the dosiomic model and radiomic model were also significantly higher than that of the DVH model (p-value <0.05). However, the ROCAUCs of the dosiomic and radiomic models were not significantly different (p-value = 0.62), although the PRAUC of the radiomic model was significantly higher than that of the dosiomic model (p-value <0.05).

For the lung dataset, the model based on DVH features resulted in an ROCAUC of 0.64 ± 0.18 and a PRAUC of 0.37 ± 0.20. The model based on dosiomic features resulted in an ROCAUC of 0.67 ± 0.17 and a PRAUC of 0.37 ± 0.20. The model based on radiomic features resulted in an ROCAUC of 0.67 ± 0.16 and a PRAUC of 0.45 ± 0.23. The model based on dosiomic + radiomic features resulted in an ROCAUC of 0.68 ± 0.16 and a PRAUC of 0.44 ± 0.22. The results of lung cancer dataset were included in **Table 3**. The ROCAUCs of the dosiomic, radiomic and dosiomic + radiomic models was significantly higher than that of the DVH model. However, only the PRAUCs of the radiomic and dosiomic + radiomic models were significantly higher than that of the DVH model, and the PRAUCs of the dosiomic and DVH models were not significantly different (p-value = 0.61).

We also provided the results when selecting the features within CV loop in **Supplementary Tables 3–5**. The model building method was included in **Supplementary S2**. For esophageal dataset, the model based on the DVH features resulted in an ROCAUC of 0.67 ± 0.10 and a PRAUC of 0.74 ± 0.09. The model based on dosiomic features resulted in an ROCAUC of 0.70 ± 0.10 and a PRAUC of 0.77 ± 0.09. The model based on radiomic features resulted in an ROCAUC of 0.63 ± 0.11 and a PRAUC of 0.72 ± 0.09. The model based on dosiomic + radiomic features resulted in an ROCAUC of 0.70 ± 0.11 and a PRAUC of 0.78 ± 0.10. For lung cancer dataset, the model based on the DVH features resulted in an ROCAUC of 0.61 ± 0.17 and a PRAUC of 0.26 ± 0.16. The model based on dosiomic features resulted in an ROCAUC of 0.67 ± 0.18 and a PRAUC of 0.38 ± 0.21. The model based on radiomic features resulted in an ROCAUC of 0.66 ± 0.17 and a PRAUC of 0.46 ± 0.24. The model based on dosiomic + radiomic features resulted in an ROCAUC of 0.70 ± 0.17 and a PRAUC of 0.43 ± 0.22.

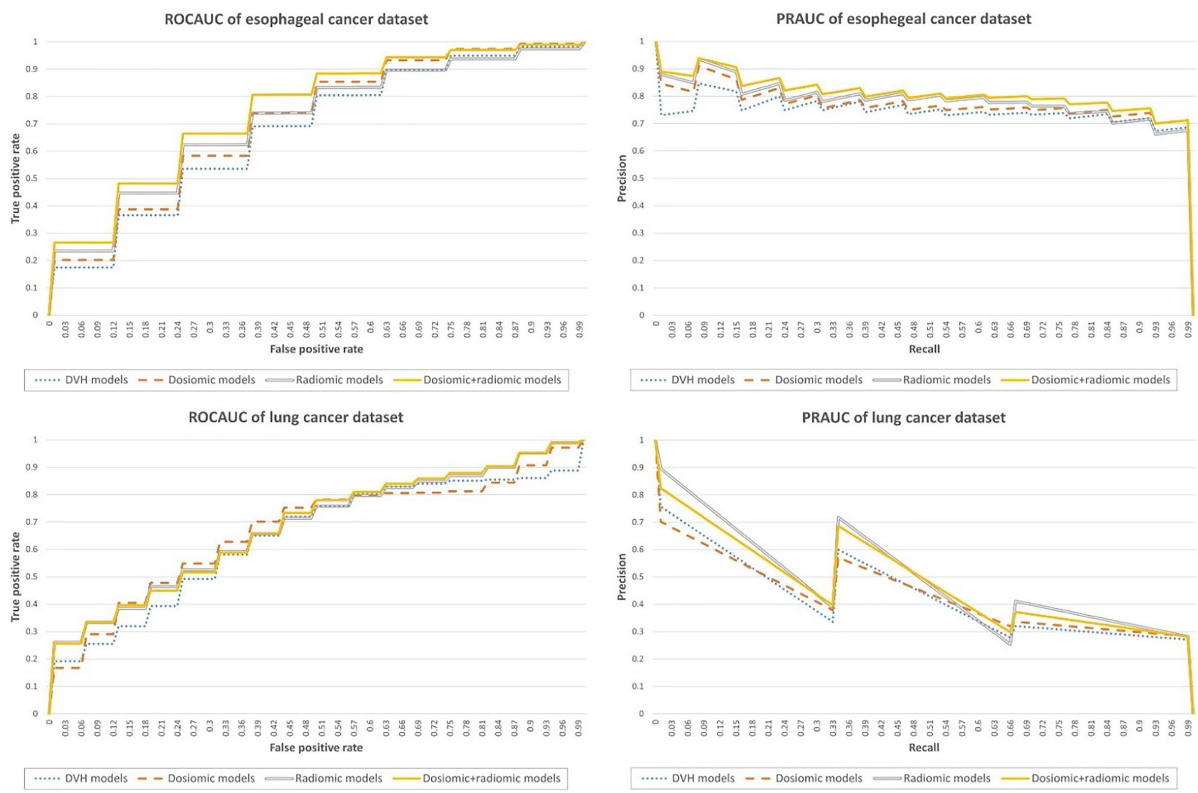


FIGURE 2 | Performance metrics of the predictive models obtained on the esophageal cancer and lung cancer datasets.

TABLE 2 | ROCAUC and PRAUC scores for the esophageal cancer dataset.

	ROCAUC	10th, 90th ROCAUC	PRAUC	10th, 90th PRAUC
DVH	0.67 ± 0.11	0.53, 0.82	0.75 ± 0.10	0.61 0.87
Dosiomic	0.71 ± 0.10	0.58, 0.84	0.77 ± 0.09	0.65 0.89
Radiomic	0.71 ± 0.11	0.57, 0.85	0.79 ± 0.09	0.66 0.90
Dosiomic + Radiomic	0.75 ± 0.10	0.63, 0.88	0.81 ± 0.09	0.68, 0.92

DISCUSSION

Our results showed that the dosiomic and radiomic models achieved higher AUCs than the DVH-based models on the esophageal cancer dataset. The results from our studies, obtained on an esophageal cancer dataset, were different from those of previous studies on the use of quantitative CT image features for esophageal cancer to predict RP grade ≥2, which found that SUV95 was a predictive feature but that CT images were bad predictors (12). However, we investigated more CT radiomic features than a previous study by Castillo et al. with respect to esophageal cancer. Furthermore, in our population, PET might not have been available for all patients due to restrictions in Thailand. Another study on esophageal cancer data also found that CT image-based delta-radiomics improved discriminative ability of patient developing grade ≥ 3RP within 3 months (31). Delta-radiomics was the technique that analyzes the radiomics features at different time. It was show that delta-radiomics features were robust than simple radiomics features (32). Delta-radiomics also have the advantage of more data over radiomics features. An advantage of using only pre-treatment data is that it might allow clinician to revise the treatment plan before initiating treatment, therefore preventing radiation pneumonitis.

In the lung dataset, only the radiomic and dosiomic + radiomic models achieved better performance than the DVH-based model. Although the ROCAUC of the dosiomic model was better than that of the DVH-based model, the PRAUC of the dosiomic-based model was not different from that of the DVH-based model. This demonstrated that dosiomics and radiomics could improve the performance of predictive models for RP, as observed in previous studies (6, 33, 34). Furthermore, knowledge of radiomic and dosiomic features might be transferable from one dataset to another dataset with performance that is equivalent to or better than that of standard DVH features.

The DVH features selected using the lung cancer dataset (DVH_{lung}) were expected to differ from the DVH features selected from the esophageal cancer dataset, and the model using DVH_{lung} was expected to have better performance than the DVH features selected from the esophageal cancer dataset. From the results, DVH_{lung} was different from DVH, as expected, but the performance of DVH was not different from that of DVH_{lung}. Nonetheless, radiomics and dosiomics still performed slightly better regarding the prediction of RP than DVH_{lung}.

Previous studies on the use of radiomics and dosiomics for radiation pneumonitis prediction reported a variety of "most important" features. Among CT radiomic studies with respect to RP, Hirose et al. were the first to investigate a predictive model using only pretreatment CT radiomics for predicting RP grade ≥2 in lung cancer after stereotactic body radiotherapy (SBRT) (6). One of the most selected features was "correlation" from the GLCM. Nevertheless, a radiomic study by Krafft et al. (7) did not find any features that were common with those of Hirose et al. It was difficult to compare the two studies since the treatment modalities, extracted features and methods employed to build the models were not the same. The results of dosiomic studies relative to RP were also difficult to compare due to differences in the extracted features. For example, Liang et al. found "contrast" from the GLCM and "low grey level run emphasis" from the GLRLM as the most predictive features of RP ≥2 in lung cancer patients treated with volumetric modulated arc therapy (VMAT) (34), while the study of dosiomics in lung cancer patients treated with VMAT by Bourbonne et al. investigated acute and late lung toxicity separately, which was different approach from that of Liang et al. (35). Adachi et al. made a study of dosiomics that utilized different modalities (SBRTs) and different techniques for feature extraction (33).

A systemic review of PET/CT texture features also found that many texture features have been reported even though the datasets used were similar in terms of cancer types and modalities (13). The study in which the dataset and features were the most similar to those in our study was the work of Liang et al., although the patients were different, and there was no work done regarding CT radiomics (34). One drawback of their study was that the result was not validated on a test set. However, we separated a test dataset for the evaluation of our model. From the obtained results, our selected dosiomic features were different from those of their study. The differences in the selected features might be derived from the difference between the training sets,

TABLE 3 | ROCAUC and PRAUC scores for the lung cancer dataset.

	ROCAUC	10th, 90th ROCAUC	PRAUC	10th, 90th PRAUC
DVH_{lung}	0.64 ± 0.18	0.42, 0.88	0.38 ± 0.20	0.13, 0.66
DVH	0.64 ± 0.18	0.42, 0.88	0.37 ± 0.20	0.13, 0.66
Dosiomic	0.67 ± 0.17	0.44, 0.90	0.37 ± 0.20	0.15, 0.68
Radiomic	0.67 ± 0.16	0.46, 0.88	0.45 ± 0.23	0.14, 0.75
Dosiomic + Radiomic	0.68 ± 0.16	0.46, 0.90	0.44 ± 0.22	0.15, 0.75

since we trained the model on esophageal cancer patients, as opposed to lung cancer patients.

Some studies have reported that the results from radiomic features can be biased due to false positives, and an external dataset is required to confirm the predictability of models (13, 36). Our study did not use an external dataset to validate the performance of the predictive model but to validate the radiomic and dosiomic features instead. To avoid biases in the radiomic and dosiomic features, we incorporated lung cancer patients receiving RT as an external dataset with an end point of predicted RP grade ≥2. The aim of incorporating an external dataset was to ensure the superiority of the predictive abilities of dosiomic and radiomic features over that of DVH features in the same organ. The results obtained on an external dataset indicated that dosiomic and radiomic feature performed equally or superior to DVH-based features in the same organ, even with different primary cancers.

There were several limitations in our study. First, our study was based on retrospective data, which might have resulted in false positives (36). Nevertheless, we tried to overcome this limitation by introducing lung cancer patients as an external dataset. Another limitation was that, from a biological standpoint, grade 1 RP and grade 2 RP are different. This is due to unavailability of grade 1 RP data in lung cancer. Grade 1 RP is viewed as local damage from the criterion of CT image changes. This might be a drawback of the study regarding grade 1 RP, which could cause the results to be inapplicable to grade 2 RP. However, grade 1 RP can also be viewed as whole organ damage if it is classified as grade 1 by symptom criteria, which would be biologically similar to grade 2 RP. The input spaces in the lung cancer and esophageal cancer datasets were also different. The input space in the lung cancer dataset had smaller ROIs than that in the esophageal cancer dataset, and the doses administered to the lungs of lung cancer patients were also higher than the doses administered to the lungs of esophageal cancer patients. Another difference was that the risk factors, such as the locations of the primary tumors, were not the same. Despite all the differences described above, we found that the dosiomic and radiomic feature models could achieve performance that was equal or superior to that of the DVH model.

The method of features selection in esophageal cancer dataset might cause overfitting problem for esophageal cancer results since the features selection process was carry out of the CV loop but not be for lung cancer dataset. Thus, we also provided the results of models using feature selection within the CV loop which eliminated the problem of overfitting in **Supplementary S2**. In summary, only the performance of the model that included radiomic decrease less than DVH model in esophageal cancer dataset (**Supplementary Table 3**), while in lung cancer dataset, the performance of radiomic and dosiomic model still greater than DVH model (**Supplementary Table 4**). The featured select in both methods were similar (**Supplementary Table 5**). The purpose was to test the transferability of dosiomic and radiomic features to lung cancer dataset which the process does not cause overfitting in the lung cancer results.

CONCLUSION

In conclusion, studies on the dosiomic and radiomic feature of RP are in the early stage. Our study found that dosiomic and radiomic models could enhance the performance of RP prediction models for esophageal and lung cancer patients treated with RT. Further prospective studies are required to validate the effectiveness of dosiomic and radiomic features.

DATA AVAILABILITY STATEMENT

The data analyzed in this study is subject to the following licenses/restrictions: No. Requests to access these datasets should be directed to yodchanan.won@mahidol.ac.th.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Human research ethics committee, Faculty of Medicine Ramathibodi Hospital, Mahidol University (IRB MURA2021/283). Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

CP, NS, NT and CJ contributed to the acquisition of the data. NS and PP contributed to clinical criteria. CP contributed to processing the data, features extraction, machine learning, statistical analysis and drafted the manuscript. SK, PP, CJ and YW were senior author supervising the project.

FUNDING

This study is supported in part by the National Higher Education Science Research and Innovation Policy Council, PMU B (B05F640079).

ACKNOWLEDGMENTS

We sincerely appreciate Pubordee Aussavavirojekul (a research fellow from the Urology Division, Surgery Department, Siriraj Hospital, Mahidol University) for the technical support regarding this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.768152/full#supplementary-material

REFERENCES

1. Hope AJ, Lindsay PE, El Naqa I, Alaly JR, Vicic M, Bradley JD, et al. Modeling Radiation Pneumonitis Risk With Clinical, Dosimetric, and Spatial Parameters. *Int J Radiat Oncol Biol Phys* (2006) 65(1):112–24. doi: 10.1016/j.ijrobp.2005.11.046
2. Valdes G, Solberg TD, Heskel M, Ungar L, Simone CBII. Using Machine Learning to Predict Radiation Pneumonitis in Patients With Stage I Non-Small Cell Lung Cancer Treated With Stereotactic Body Radiation Therapy. *Physiol Behav* (2018) 176(1):139–48. doi: 10.1016/j.physbeh.2017.03.040
3. Yakar M, Etiz D, Metintas M, Ak G, Celik O. Prediction of Radiation Pneumonitis With Machine Learning in Stage III Lung Cancer: A Pilot Study. *Technol Cancer Res Treat* (2021) 20:1–10. doi: 10.1177/15330338211016373
4. Avanzo M, Pirrone G, Vinante L, Caroli A, Stancanello J, Drigo A, et al. Electron Density and Biologically Effective Dose (BED) Radiomics-Based Machine Learning Models to Predict Late Radiation-Induced Subcutaneous Fibrosis. *Front Oncol* (2020) 10:490(April). doi: 10.3389/fonc.2020.00490
5. Gabryś HS, Buettner F, Sterzing F, Hauswald H, Bangert M. Design and Selection of Machine Learning Methods Using Radiomics and Dosiomics for Normal Tissue Complication Probability Modeling of Xerostomia. *Front Oncol* (2018) 8:35. doi: 10.3389/fonc.2018.00035
6. rose T, Arimura H, Ninomiya K, Yoshitake T, Fukunaga J, Shioyama Y. Radiomic Prediction of Radiation Pneumonitis on Pretreatment Planning Computed Tomography Images Prior to Lung Cancer Stereotactic Body Radiation Therapy. *Sci Rep* (2020) 10(1):1–9. doi: 10.1038/s41598-020-77552-7
7. Krafft SP, Rao A, Stingo F, Briere TM, Court LE, Liao Z, et al. The Utility of Quantitative CT Radiomics Features for Improved Prediction of Radiation Pneumonitis. *Med Phys* (2018) 45(11):5317–24. doi: 10.1002/mp.13150
8. Moran A, Daly ME, Yip SSF, Yamamoto T. Radiomics-Based Assessment of Radiation-Induced Lung Injury After Stereotactic Body Radiotherapy. *Clin Lung Cancer* (2017) 18(6):e425–31. doi: 10.1016/j.cllc.2017.05.014
9. Kong FM, Wang S. Nondosimetric Risk Factors for Radiation-Induced Lung Toxicity. *Semin Radiat Oncol* (2015) 25(2):100–9. doi: 10.1016/j.semradonc.2014.12.003
10. Sanuki N, Ono A, Komatsu E, Kamei N, Akamine S, Yamazaki T, et al. Association of Computed Tomography-Detected Pulmonary Interstitial Changes With Severe Radiation Pneumonitis for Patientstreated With Thoracic Radiotherapy. *J Radiat Res* (2012) 53(1):110–6. doi: 10.1269/jrr.110142
11. Ueki N, Matsuo Y, Togashi Y, Kubo T, Shibuya K, Iizuka Y, et al. Impact of Pretreatment Interstitial Lung Disease on Radiation Pneumonitis and Survival After Stereotactic Body Radiation Therapy for Lung Cancer. *J Thorac Oncol* (2015) 10(1):116–25. doi: 10.1097/JTO.0000000000000359
12. Castillo R, Pham N, Castillo E, Aso-Gonzalez S, Ansari S, Hobbs B, et al. Pre-Radiation Therapy Fluorine 18 Fluorodeoxyglucose Pet Helps Identify Patients With Esophageal Cancer at High Risk for Radiation Pneumonitis. *Radiology* (2015) 275(3):822–31. doi: 10.1148/radiol.14140457
13. Chalkidou A, O'Doherty MJ, Marsden PK. False Discovery Rates in PET and CT Studies With Texture Features: A Systematic Review. *PloS One* (2015) 10 (5):1–18. doi: 10.1371/journal.pone.0124165
14. Desideri I, Loi M, Francolini G, Becherini C, Livi L, Bonomo P. Application of Radiomics for the Prediction of Radiation-Induced Toxicity in the IMRT Era: Current State-Of-the-Art. *Front Oncol* (2020) 10:1708. doi: 10.3389/fonc.2020.01708
15. Fornacon-Wood I, Faivre-Finn C, O'Connor JPB, Price GJ. Radiomics as a Personalized Medicine Tool in Lung Cancer: Separating the Hope From the Hype. *Lung Cancer* (2020) 146:197–208. doi: 10.1016/j.lungcan.2020.05.028
16. Leijenaar RTH, Carvalho S, Hoebers FJP, Aerts HJWL, Van Elmpt WJC, Huang SH, et al. External Validation of a Prognostic CT-Based Radiomic Signature in Oropharyngeal Squamous Cell Carcinoma. *Acta Oncol* (2015) 54 (9):1423–9. doi: 10.3109/0284186X.2015.1061214
17. 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
18. Taylor JMG, Ankerst DP, Andridge RR. Validation of Biomarker-Based Risk Prediction Models. *Clin Cancer Res* (2008) 14(19):5977–83. doi: 10.1158/1078-0432.CCR-07-4534
19. Niebuhr NI, Splinter M, Bostel T, Seco J, Hentschke CM, Floca RO, et al. Biologically Consistent Dose Accumulation Using Daily Patient Imaging. *Radiat Oncol* (2021) 16(1):1–16. doi: 10.1186/s13014-021-01789-3
20. Borst GR, Ishikawa M, Nijkamp J, Hauptmann M, Shirato H, Bengua G, et al. Radiation Pneumonitis After Hypofractionated Radiotherapy: Evaluation of the LQ(L) Model and Different Dose Parameters. *Int J Radiat Oncol Biol Phys* (2010) 77(5):1596–603. doi: 10.1016/j.ijrobp.2009.10.015
21. Dhami G, Zeng J, Vesselle HJ, Kinahan PE, Miyaoka RS, Patel SA, et al. Framework for Radiation Pneumonitis Risk Stratification Based on Anatomic and Perfused Lung Dosimetry. *Physiol Behav* (2018) 63(5):1–18. doi: 10.1007/s00066-017-1114-0.Framework
22. Palma G, Monti S, Xu T, Scifoni E, Yang P, Hahn SM, et al. Spatial Dose Patterns Associated With Radiation Pneumonitis in a Randomized Trial Comparing Intensity-Modulated Photon Therapy With Passive Scattering Proton Therapy for Locally Advanced Non-Small Cell Lung Cancer. *Int J Radiat Oncol Biol Phys* (2019) 104(5):1124–32. doi: 10.1016/j.ijrobp.2019.02.039
23. Ren C, Ji T, Liu T, Dang J, Li G. The Risk and Predictors for Severe Radiation Pneumonitis in Lung Cancer Patients Treated With Thoracic Reirradiation. *Radiat Oncol* (2018) 13(1):1–7. doi: 10.1186/s13014-018-1016-z
24. Selvaray J, Lebesque J, Hope A, Guckenberger M, Werner-Wasik M, Peulen H, et al. Modeling Radiation Pneumonitis of Pulmonary Stereotactic Body Radiotherapy: The Impact of a Local Dose–Effect Relationship for Lung Perfusion Loss. *Radiother Oncol* (2019) 132:142–7. doi: 10.1016/j.radonc.2018.12.015
25. Stefan LS, Lebesque JV, Theuws JCM, Marks LB, Munley MT, Bentel G, et al. Radiation Pneumonitis as a Function of Mean Lung Dose: An Analysis of Pooled Data of 540 Patients. *N Engl J Med* (1970) 282(2):104. doi: 10.1056/NEJM197001082820220
26. Thor M, Deasy J, Iyer A, Bendau E, Fontanella A, Apte A, et al. Toward Personalized Dose-Prescription in Locally Advanced Non-Small Cell Lung Cancer: Validation of Published Normal Tissue Complication Probability Models. *Physiol Behav* (2016) 176(12):139–48. doi: 10.1016/j.radonc.2019.05.011.Toward
27. Van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res* (2017) 77(21):e104–7. doi: 10.1158/0008-5472.CAN-17-0339
28. Zwanenburg A, Leger S, Vallières M, Löck S. *Image Biomarker Standardisation Initiative*. (2016). doi: 10.1148/radiol.2020191145.
29. 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):1–21. doi: 10.1371/journal.pone.0118432
30. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat Methods* (2020) 17(3):261–72. doi: 10.1038/s41592-019-0686-2
31. Wang L, Gao Z, Li C, Sun L, Li J, Yu J, et al. Computed Tomography-Based Delta-Radiomics Analysis for Discriminating Radiation Pneumonitis in Patients With Esophageal Cancer After Radiation Therapy. *Int J Radiat Oncol Biol Phys* (2021) 111(2):443–55. doi: 10.1016/j.ijrobp.2021.04.047
32. Nardone V, Reginelli A, Guida C, Belfiore MP, Biondi M, Mormile M, et al. Delta-Radiomics Increases Multicentre Reproducibility: A Phantom Study. *Med Oncol* (2020) 37(5):1–7. doi: 10.1007/s12032-020-01359-9
33. Adachi T, Nakamura M, Shintani T, Mitsuyoshi T, Kakino R, Ogata T, et al. Multi-Institutional Dose-Segmented Dosiomic Analysis for Predicting Radiation Pneumonitis After Lung Stereotactic Body Radiation Therapy. *Med Phys* (2021) 48(4):1781–91. doi: 10.1002/mp.14769
34. 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
35. Bourbonne V, Da-ano R, Jaouen V, Lucia F, Dissaux G, Bert J, et al. Radiomics Analysis of 3D Dose Distributions to Predict Toxicity of Radiotherapy for Lung Cancer. *Radiother Oncol* (2021) 155:144–50. doi: 10.1016/j.radonc.2020.10.040
36. Yip SSF, Aerts HJWL. Applications and Limitations of Radiomics. *Phys Med Biol* (2016) 61(13):R150–66. doi: 10.1088/0031-9155/61/13/R150

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

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: 18 March 2022
doi: 10.3389/fonc.2022.827991

Multiscale Local Enhancement Deep Convolutional Networks for the Automated 3D Segmentation of Gross Tumor Volumes in Nasopharyngeal Carcinoma: A Multi-Institutional Dataset Study

Geng Yang[1,2,3†], Zhenhui Dai[3†], Yiwen Zhang[1,2], Lin Zhu[3], Junwen Tan[4], Zefeiyun Chen[1,2], Bailin Zhang[3], Chunya Cai[3], Qiang He[3], Fei Li[3], Xuetao Wang[3] and Wei Yang[1,2*]*

[1] School of Biomedical Engineering, Southern Medical University, Guangzhou, China, [2] Guangdong Provincial Key Laboratory of Medical Image Processing, Southern Medical University, Guangzhou, China, [3] Department of Radiation Therapy, The Second Affiliated Hospital of Guangzhou University of Chinese Medicine, Guangzhou, China, [4] Department of Oncology, The Fourth Affiliated Hospital of Guangxi Medical University, Liuzhou, China

Purpose: Accurate segmentation of gross target volume (GTV) from computed tomography (CT) images is a prerequisite in radiotherapy for nasopharyngeal carcinoma (NPC). However, this task is very challenging due to the low contrast at the boundary of the tumor and the great variety of sizes and morphologies of tumors between different stages. Meanwhile, the data source also seriously affect the results of segmentation. In this paper, we propose a novel three-dimensional (3D) automatic segmentation algorithm that adopts cascaded multiscale local enhancement of convolutional neural networks (CNNs) and conduct experiments on multi-institutional datasets to address the above problems.

Materials and Methods: In this study, we retrospectively collected CT images of 257 NPC patients to test the performance of the proposed automatic segmentation model, and conducted experiments on two additional multi-institutional datasets. Our novel segmentation framework consists of three parts. First, the segmentation framework is based on a 3D Res-UNet backbone model that has excellent segmentation performance. Then, we adopt a multiscale dilated convolution block to enhance the receptive field and focus on the target area and boundary for segmentation improvement. Finally, a central localization cascade model for local enhancement is designed to concentrate on the GTV region for fine segmentation to improve the robustness. The Dice similarity coefficient (DSC), positive predictive value (PPV), sensitivity (SEN), average symmetric surface distance (ASSD) and 95% Hausdorff distance (HD95) are utilized as qualitative evaluation criteria to estimate the performance of our automated segmentation algorithm.

OPEN ACCESS

Edited by:
Jose Eduardo Villarreal Barajas, Royal Devon and Exeter Hospital, United Kingdom

Reviewed by:
Frankie Mo, The Chinese University of Hong Kong, Hong Kong SAR, China
Haihua Yang, Wenzhou Medical University, China

***Correspondence:**
Wei Yang
weiyanggm@gmail.com
Xuetao Wang
wangxuetao0625@126.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: *02 December 2021*
Accepted: *24 February 2022*
Published: *18 March 2022*

Citation:
Yang G, Dai Z, Zhang Y, Zhu L, Tan J, Chen Z, Zhang B, Cai C, He Q, Li F, Wang X and Yang W (2022) Multiscale Local Enhancement Deep Convolutional Networks for the Automated 3D Segmentation of Gross Tumor Volumes in Nasopharyngeal Carcinoma: A Multi-Institutional Dataset Study. Front. Oncol. 12:827991. doi: 10.3389/fonc.2022.827991

Results: The experimental results show that compared with other state-of-the-art methods, our modified version 3D Res-UNet backbone has excellent performance and achieves the best results in terms of the quantitative metrics DSC, PPR, ASSD and HD95, which reached 74.49 ± 7.81%, 79.97 ± 13.90%, 1.49 ± 0.65 mm and 5.06 ± 3.30 mm, respectively. It should be noted that the receptive field enhancement mechanism and cascade architecture can have a great impact on the stable output of automatic segmentation results with high accuracy, which is critical for an algorithm. The final DSC, SEN, ASSD and HD95 values can be increased to 76.23 ± 6.45%, 79.14 ± 12.48%, 1.39 ± 5.44mm, 4.72 ± 3.04mm. In addition, the outcomes of multi-institution experiments demonstrate that our model is robust and generalizable and can achieve good performance through transfer learning.

Conclusions: The proposed algorithm could accurately segment NPC in CT images from multi-institutional datasets and thereby may improve and facilitate clinical applications.

Keywords: nasopharyngeal carcinoma, segmentation, deep learning, radiotherapy, CT images

1 INTRODUCTION

Originating in the nasopharynx epithelium, NPC is a malignant tumor with the highest incidence among otolaryngological cancers in Southwest Asia, Southern China and Northern Africa (1, 2). Radiation therapy is the preferred treatment strategy for NPC because the poorly differentiated squamous cell carcinoma discovered upon pathological examination of patients with NPC is commonly radiosensitive (3). Delineating the tumor contour is the essential step in radiotherapy planning, which is the mainstay of NPC treatment. Based on the gross tumor volume (GTV) and organs at risk (OARs), the dose distribution of irradiation can be calculated by a radiation physicist. Therefore, accurate segmentation contributes to delivering the prescribed dose to the tumor volume while improving the sparing of OARs (4). However, in clinical practice, tumor segmentation is carried out manually by slices using multimodal or multiparametric imaging datasets, which is time-consuming (5, 6). Interobserver variability, especially in accuracy, is based on the expertise and experience of the radiation oncologist (7, 8) because on imaging, NPC often has a more complex tissue structure that has a similar intensity to its neighboring organs and a high variations in shape and size among cases (9). Accordingly, it is clinically desirable to develop a robust, accurate and automatic algorithm for target segmentation, which is helpful for reducing the labor intensity and interobserver variability.

Among many of the proposed autosegmentation approaches, atlas-based segmentation (10) has been widely used for the delineation of targets and/or OARs in head-and-neck radiotherapy (11–14), and it can obtain acceptable results without supervision. Currently, with the enormous success of deep learning in object detection (15), image classification (16), and segmentation (17), the applications of deep learning in medical imaging have received great attention. As the most popular algorithm for deep learning, convolutional neural networks (CNNs) have made significant progress in semantic segmentation with the advantage of an end-to-end framework for feature learning and model training. After the full convolutional network (FCN) was proposed by Long et al. (18), segmentation was achieved more efficiently in inference and learning for images with arbitrary sizes. UNet is the most successful FCNs utilized in medical image segmentation investigations and has been cited more than 29000 times since it was proposed in 2015 (17). Its success is largely attributed to the U-shaped architecture and skip connection in which the fusion of multiscale features and the recovery of fine-grained details can be realized effectively.

Deep learning with the CNN technique in tumor segmentation has recently made progress in the brain (19), rectum (20) and breast (21). For the segmentation of the GTV of NPC, a majority of CNN-based approaches have been applied in magnetic resonance imaging (MRI), which demonstrates superb resolution and soft tissue contrast and obtains satisfying results (22–24). However, radiotherapy plans are designed based on CT images, and MRI-based radiation therapy techniques have not been widely applied in clinical practice. Benefitting from the complementary information from both CT and MRI images, Ma et al. (25) proposed a multimodality segmentation framework based on CNN. While the performance of this kind of approach largely depends on the image registration accuracy, its difficulty will be further increased due to different body positions, scan times and imaging mechanisms. Therefore, the CT-based NPC segmentation technique is the core element that can actually solve the above clinical problems.

Although several methods have been explored for CT-based GTV segmentation (4, 26, 27), the results are barely satisfactory, and it remains the most challenging task primarily because of 1) CT images with lower contrast result in a lack of clear tumor boundaries; 2) tumors present a great variety of sizes and morphologies between different stages, especially for stages T3 and T4, which always have lymph node metastasis and other distant metastases, thus increasing the difficult of distinguishing

the primary tumor. Therefore, the presented algorithms demonstrated good performance for early nasopharyngeal carcinoma, although the accuracy decreased sharply when advanced-stage tumors were included. Meanwhile, most of these methods were based on 2D segmentation models with a lack of complementary information between CT image layers. Notably, current research is based only on one specific institution dataset and does not consider differences between different data sources, which limits the universality of clinical applications.

Hence, we explored automatic delineation of GTV based on CT images and assessed its applicability for stages T1-T4. In our previous study (28), a modified version 3D U-Net model based on Res-block and SE-block to delineate the GTV for NPC was developed; however, the accuracy still needs to be improved. With the aim of promoting network performance, we propose a cascaded multiscale local enhancement CNN structure, which can realize NPC segmentation from global to local scales by concentrating on the GTV region, and multiscale features in CT images can be captured simultaneously. Comprehensive experiments on diverse multi-institutional planning CT datasets were performed to demonstrate the effectiveness of our algorithm. Both the qualitative and quantitative evaluation results show that our approach can achieve good segmentation performance and outperforms other state-of-the-art segmentation methods.

2 MATERIALS AND METHODS

2.A. Datasets

We collected retrospective data on 257 patients who were diagnosed with NPC with stage T1-T4 and underwent radiotherapy in our institution from 2016 to 2020. Data were derived from radiotherapy treatment planning, including plain CT (pCT) and contrast-enhanced CT (CE-CT) images, with the scanned region covering the overall head, neck and partial chest obtained by simulation CT. All images were axially reconstructed with a matrix size of 512 × 512 pixels, a resolution of 0.748~0.976 mm and a slice thickness of 3.0 mm. The radiotherapy contours were jointly delineated slice by slice by two radiation oncologists on the pCT image by fusion of CE-CT and MR images according to the consensus as the ground truth. The director with 20 years' experience of the radiation oncology department was consulted in cases of disagreement.

Meanwhile, multi-institutional datasets were also employed in the experiments to evaluate the performance of the proposed algorithm. These datasets from institution B are composed of 40 NPC patient pCT images, and the MICCAI 2019 StructSeg challenge (GTV segmentation task) is composed of 50 NPC patient pCT images. **Table 1** shows the details of the multi-institutional datasets.

TABLE 1 | Details of the multi-institutional datasets.

Dataset	A	B	C
Source	Our department	Institution B	MICCAI 2019
Images	pCT + CE-CT	pCT	pCT
Number	257	40	50
Stage	T1-T4 (30:93:87:47)	NA	NA
Axially size	512 × 512	512 × 512	512 × 512
Slice thickness	3 mm	3 mm	3 mm

2.B. Overview of the Method

In this paper, inspired by cascaded method (29), we propose a two-stage cascaded multiscale local enhancement network structure to achieve the goal of building a precise GTV segmentation method. In the training phase, two networks were trained simultaneously: one was for globally coarse segmentation predictions and named MDR-UNet$_1$, and the other was for locally fine segmentation predictions concentrated on the GTV and named MDR-Uet$_2$. The testing phase could be divided into three steps: 1) Obtain the globally coarse segmentation result by MDR-UNet$_1$; 2) Identify the central location of the ROI from coarse results and acquire precise segmentation for the cropped target region by MDR-UNet$_2$; and 3) Assemble the two-stage results to output as the final prediction. The overall framework is presented in **Figure 1**.

The above two-stage independent cascade network structure with central localization was designed to achieve 'coarse-to-fine' segmentation for the GTV region. Moreover, there are two other key points that should be noted in this method: 1) Crop the ROI using a body mask, which is obtained by morphology and geometry; and 2) Use multiscale dilated convolution blocks in the skip connections between the encoder and decoder to enhance the receptive field to improve the segmentation performance. These points will be described specifically in the following.

2.C. Preprocessing

In this study, a 3D CNN was introduced in our automated delineation GTV network structure. However, calculation efficiency was an issue because of the massive occupation of graphic processing unit (GPU) memory caused by 3D data. Reducing neural network channels, layers, and batch size are always selected as the solution; however, such changes will have a great negative impact on the training results. Hence, we preprocessed the data to minimize the data cropped from the original images and simultaneously guarantee that the target was completely included. For the preprocessing procedures, we first removed beds and obtained the body mask by utilizing Hough transform line detection, threshold and morphological methods, which are binary images with targets shown in pixels (value=1) and ground shown in pixels (value=0). Second, the body mask was integrated along the vertical direction to find the minimum point, and the corresponding abscissa was the neck position. Then, the volume above the neck was clipped as the ROI. Finally, for normalization, 1) the target voxel spacing was normalized to 0.952×0.952×3 mm for all of the data with third-order spline interpolation; and 2) the intensities were normalized by subtraction of the global mean and division by the global standard deviation, which are obtained by computing the foreground voxels based on body mask cropping in the dataset. After preprocessing, the image overall mean size with standard

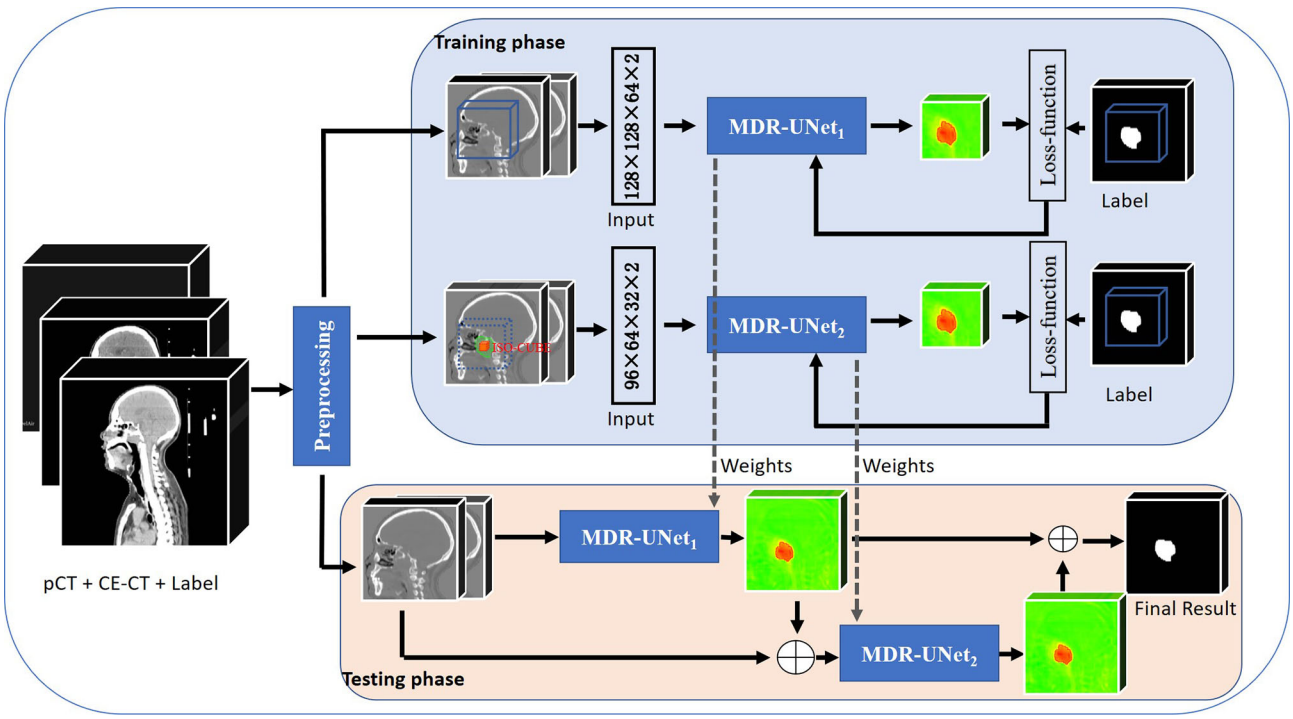


FIGURE 1 | Overview of the workflow of the proposed method.

deviation decreased from 512×512× (130 ± 11.7) to (190 ± 32.8) × (222 ± 17.4) × (74 ± 4.8).

2.D. Proposed Two-Stage Cascade Architecture

2.D.1. Stage 1: Initial Coarse Segmentation and Central Localization

Although the image size was greatly reduced to half of the original after preprocessing, it was still too large to put all the images into the network for 3D CNN training. Therefore, the preprocessed images were randomly cropped by a sliding window before being put into the network during the training stage to satisfy the calculation requirement. However, although the cropped image might contain massive background information, a few ROIs result from random cutting. As a result, only the globally coarse segmentation region of the GTV has been extracted at this stage. Based on the preliminary segmentation, the centroid of the GTV can be determined and used as preparation for the next locally precise segmentation. As a trade-off between model performance and memory consumption, the size of the window was set to 128×128×64.

2.D.2. Stage 2: Central ROI Fine-Segmentation

The aim of this stage was to further narrow the target region and perform precise segmentation around the center of the tumor. To achieve this goal, the first step was to realize the central localization of the tumor. For the training data, the center of the GTV can be identified from the ground truth, while for the test data, the location can only be determined from the segmentation predicted in stage 1. Considering the error between them, the position differences for the training data were compared, and the results demonstrated that the centroid of the GTV between the ground truth and stage 1 segmentation results varied (1.46 ± 1.34, 1.84 ± 1.69, 1.14 ± 1.09) pixels. Hence, by considering the error, we allowed the central location to vary with the range in (± 3, ± 3, ± 3) pixels of the center determined from the ground truth or stage 1 segmentation result. In addition to the central localization, the cropped cube size of the GTV had to be determined. The cube size of the GTV is known to correlate with the stage of the tumor; therefore, we calculated the size of the GTV for all of the data and found that the overall mean with standard deviation was 64 ± 8.5, 43 ± 12.1, and 15 ± 5.7 pixels. Thus, the cropped cube size of the ROI was set to 96×64×32. The process not only made it possible to further reduce the size of the input image, which contains much more ROI area and less background information, but also optimized the network model and facilitated its convergence in training.

2.D.3. Network Architecture

Our network architecture is shown in **Figure 2**. We used UNet (17) and its 3D counterpart VNet (30)-like architecture as the backbone due to their excellent performance in medical image segmentation. UNet consists of three components: 1) downsampling for feature extraction, 2) upsampling for resolution restoration and 3) skip-connection for feature fusion, which can achieve the multiscale feature extraction of large medical images. Res-block (31) was introduced in the

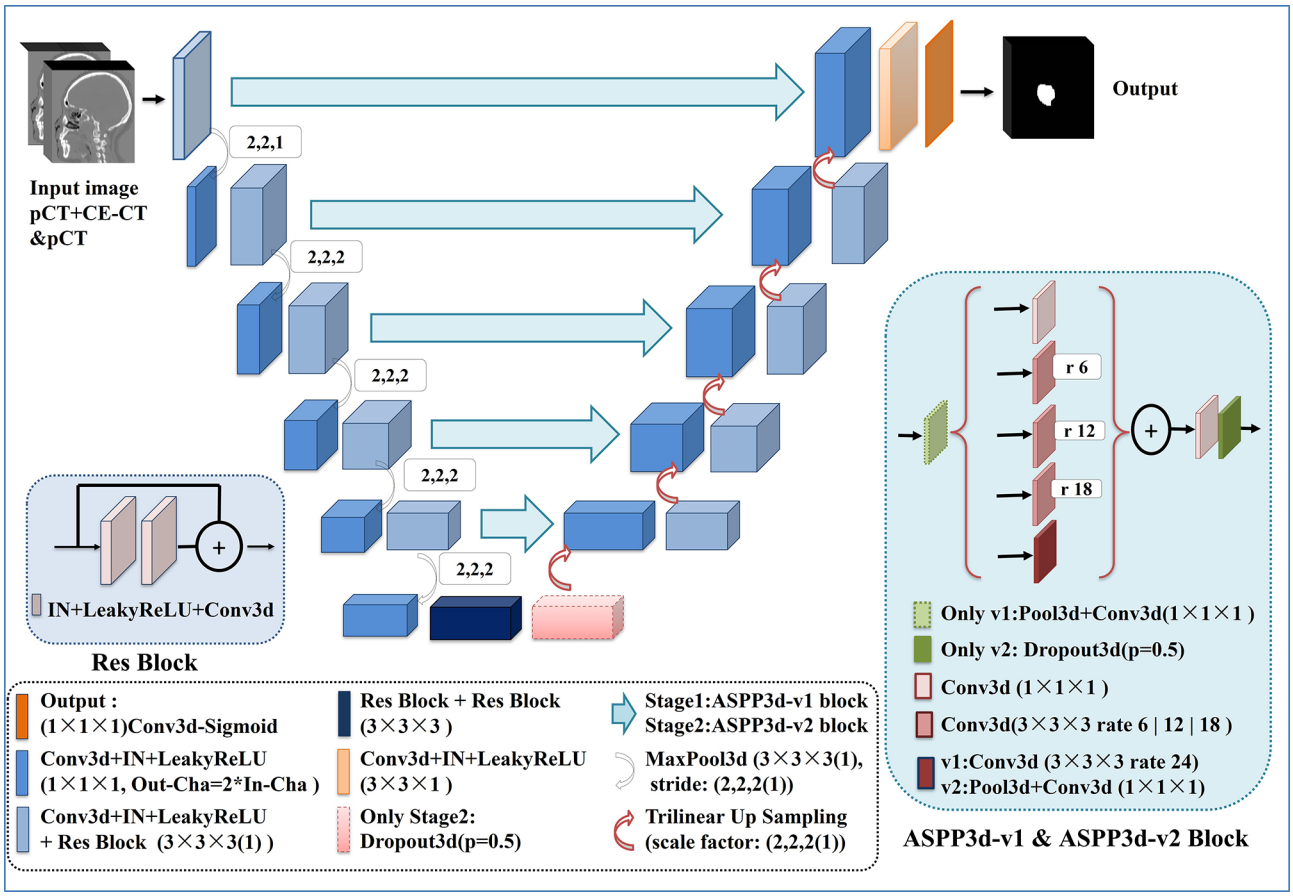


FIGURE 2 | Network architecture of the proposed CNN model.

procedure of downsampling and upsampling because it has excellent performance in deep networks for solving gradient dispersion and precision reduction. In the skip connection, we employed the modified 3D ASPP (32, 33) block because it can use multiscale dilated convolution to enlarge the field of view. The proposed network contains 11 Res-blocks and 5 ASPP-blocks in total. The network adopts the common configuration of blocks per resolution step in both the encoder and decoder, with each block consisting of a Conv-block and Res-block. The proposed network can only be trained with a small batch size for a large patch size. In the case of a small batch size, we utilized instance normalization (34) to speed up or stabilize the training because of the poor performance shown by batch normalization (35). Furthermore, the leaky ReLU (36) was used as the activation function in the hidden layers. Following the final segmentation map, sigmoid activation function outputs are obtained. As a compromise between model performance and memory consumption, the initial number of feature maps was set to 32 and doubled or halved in each downsampling or upsampling procedure. To limit the parameter size of the final model, the number of feature maps was additionally capped at 512 and 1024 for stage 1 and stage 2, respectively.

For the best performance, the two-stage network architectures were slightly different to allow for adaptations to different data characteristics. In stage 1, due to the large input image, we introduced the ASPP block (32) with a larger dilation rate in the skip connection to obtain multiscale feature information of different fields of view. For stage 2, further reducing the size and range of the input image would contribute to the appearance of overfitting during the network training. Hence, to obtain a better training effect, the ASPP_v3+ block (33), which has a small dilation rate and includes a dropout layer, was employed to replace the ASPP block in the skip connection for this stage. In addition, the dropout layer was also added at the end of the encoder to prevent overfitting.

2.E. Postprocessing

In the test phase, to obtain the initial coarse segmentation result, the test samples were input into the stage 1 network by grid sliding with an input size of 128×128×64 and a stride of 64 × 64 × 32. The stage 2 input ROI could be extracted from the original image by taking the centroid of the GTV determined from the stage 1 segmentation result as the center of a cube. For both stage networks, the segmentation results were acquired after activation with a sigmoid function. Then, the segmentation results obtained

in stage 1 and stage 2 were assembled by $S = \omega S_1 + (1 - \omega)S_2$, ω was set to 0.5, and the final segmentation score map was output, where the threshold value was set to 0.6 and values lower than the threshold were regarded as background. Finally, the predicted ROI was reconstructed and returned to the corresponding position of the original CT images.

3 EXPERIMENTS AND RESULTS

3.A. Experimental Scheme

3.A.1 Implementation Details

Fivefold cross-validation was adopted to evaluate the network performance among our 257-patient dataset, which meant that the dataset was randomly divided into 5 subsamples. Specifically, in each round, one of the subsamples (20%) was used as the test set and the remaining four folds (80%) were used as training set. We repeated this procedure five times until all the five folds have been used as the test set. Varieties of data augmentation techniques were introduced into the training dataset, and they consisted of random scaling (scaling factors: 0.7~1.3), random elastic transformation (scaling factor: 34; elasticity coefficient: 10), random rotation (angle: -10~10), random noise (Gaussian noise: 0~0.1 or uniform noise: -0.1~0.1), and random flipping.

In our 257 dataset, for the two networks, the 'kaiming normal' (37) strategy was employed for weight parameter initialization, the stochastic gradient descent (SGD) optimizer was used for training with momentum of 0.99, the learning rate (LR) abided by the 'poly' policy (32) decaying with $LR = lr_{ini} * (1 - epoch/epoch_{max})^{0.9}$, initial learning rate (lr_{ini}) was set to 0.01, and the loss function was the sum of cross-entropy and dice loss. Based on previous experiments and experience, as a compromise between runtime and reward, each fold of both networks was set to 120 epochs. The batch size was set to 1, and the network parameters were updated every 2 batch sizes with the gradient accumulation method.

The experiments were implemented on a workstation powered by a NVIDIA GeForce RTX 2080Ti with 11 GB GPU memory. The code was implemented with Pytorch 1.4.0 in Ubuntu 18.04.3 LTS.

3.A.2 Evaluation Criteria

Three volumetric overlap metrics, the Dice similarity coefficient (DSC), positive predictive value (PPV), and sensitivity (SEN), and two distance metrics, the average symmetric surface distance (ASSD) and 95% Hausdorff distance (HD95), were utilized as qualitative evaluation criteria to estimate the performance of our automated segmentation algorithm. These metrics are defined as follows.

a) Volumetric overlap metrics:

$$DSC = \frac{2|G \cap A|}{|G| + |A|} \tag{1}$$

$$PPR = \frac{|G \cap A|}{|A|} \tag{2}$$

$$SEN = \frac{|G \cap A|}{|G|} \tag{3}$$

b) Distance metrics:

$$ASSD = \frac{1}{2}\left\{ \underset{g \in G_s}{\text{meanmin}}\ \underset{a \in A_s}{d(a,g)} + \underset{a \in A_s}{\text{meanmin}}\ \underset{g \in G_s}{d(a,g)} \right\} \tag{4}$$

$$HD_{95}(G_s, A_s) = \max\left(d_{95}(G_s, A_s), d_{95}(A_s, G_s)\right) \tag{5}$$

Where

$$d_{95}(G_s, A_s) = K^{95}_{g \in G_s}(\min_{a \in A_s}||a - g||) \tag{6}$$

For the volumetric overlap metrics, G and A represent the voxel sets of the ground truth and automatic delineation, respectively. DSC, PPV and SEN, which are a series of ratios, are used to describe the corresponding spatial overlap between the ground truth and the automated delineation, and a higher value indicates better performance.

For distance metrics, where G_s and A_s are the corresponding surface voxel sets of G and A, $d(a,g)$ and $||a - g||$ are the Euclidean distance of the voxel between a and g, $d_{G_s A_s}$ describes the point $x \in X_s$ that is farthest from any point of Y_s and calculates the distance from x to its nearest neighbor in Y_s. ASSD and HD95 describe the mean surface distance between the ground truth and automated delineation, and a lower value shows a higher delineation accuracy.

In addition, IBM SPSS (version 23; New York, NY) was used for statistical analysis. The mean DSC, PPR, SEN, ASSD and HD95 values for the GTV segmentation were evaluated the dispersion with standard deviation (SD) and analyzed with paired t-tests between different models. All values are presented as mean ± SD. Two-tailed p-values <0.05 were considered statistically significant.

3.A.3 Comparison of Model Performance

Automated segmentation models of the GTV based on the proposed cascaded multiscale local enhancement CNN structure were achieved in this study. First, several successful network architectures and our proposed 3D Res-UNet without ASPP block (33) and cascade architecture were applied to data for 257 NPC patient cases from our institution for GTV segmentation to compare the performance. Two modalities (pCT and CE-CT) were used to train the network models, because the advantage of CE-CT in terms of contrast visibility between tumor and normal tissues can help improve the performance of network segmentation (28). The other successful network architectures were as follows: 3D CNN (22), which is based on the 3D CNN architecture of VoxResNet and used to segment the complicated NPC GTV based on MRI showing outstanding performance, Attention-UNet (38), which adds AGs to filter the features propagated through the skip connections of UNet; and UNet++ (39), which improves UNet performance by alleviating the unknown network depth, redesigning the skip connections, and devising

a pruning scheme to the architecture. For comparison, the Attention-UNet and UNet++ were changed from 2D to 3D network architectures. Second, based on the 3D Res-UNet backbone model, the important roles of the multiscale dilated convolution blocks and cascade architecture, namely, the five ASPP blocks and the two-stage cascade method shown in **Figure 2**, were further explored experimentally in the proposed algorithm. Moreover, the feature maps and the boundary probability maps were output further to confirm the ASPP block mechanism on segmentation performance.

Finally, additional datasets of 40 NPC patients from institution B and 50 NPC patients from the MICCAI 2019 StructSeg challenge were also adopted to validate the effectiveness of the model further and evaluate the generalizability of the model application. The model was applied to two additional datasets for further algorithm verification experiments in MDR-UNet without the cascade architecture to reduce time consumption. Because of the lack of CE-CT data in the institution B and C datasets, only pCT data were used for training. It is important to note that transfer learning was used to fine-tune the network based on the model trained to reduce estimation errors between different datasets. Specifically, the model was first trained on the A dataset, and its performance was evaluated on the B and C datasets. Then, based on the model pre-trained on the A dataset, transfer learning was employed for training datasets B, C and B+C. Moreover, the original and target domains were switched in the experiment to verify further the generalization ability of the network in transfer learning. Datasets B, C, and B+C were utilized for training the network to verify the network performance and tested on the A dataset. The network was then fine-tuned with dataset A using the transfer learning method. It should be noted that during transfer learning stage, datasets A, B and C were randomly split into 70% and 30% respectively for fine-tuning and testing. The network was trained for 60 epochs, and the initial learning rate lr_{ini} was adjusted to 0.001 during the fine-tuning phase.

3.B. Results and Analysis

3.B.1. Comparison Between Different Network Architectures

The quantitative comparison results of the fivefold cross-validation of the different network models trained by pCT and CE-CT are summarized in **Table 2**. As shown in the table, Res-UNet performed better in mean DSC, PPR, ASSD and HD95 with standard deviations of 74.49 ± 7.81%, 79.97 ± 13.9%, 1.49 ± 0.65mm, 5.06 ± 3.30mm, respectively, compared to other networks, benefiting from the dominance of the RES-block in solving gradient dispersion and precision reduction, which verifies the effectiveness of the proposed backbone. However, the mean SEN index of Res-UNet was only 73.9%, which is a poor result due to the lack of relevant attention mechanism, indicating that the network is under-segmented. **Figure 3** shows the visual segmentation results for the comparison of automatic ground truth delineation by the public networks, the method used for this research and human experts. It is obvious that the estimation of the proposed Res-UNet produces fewer false positive predictions and presents more robust segmentation results, particularly in the coronal and sagittal views, compared to the 3D CNN, 3D Attention-UNet and 3D UNet++. To summarize, the experimental results demonstrate that although the Res-UNet backbone network might not have the best performance and statistical significance in all indicators. Nevertheless, the key indicator DSC was significantly improved ($P<0.05$) and had better performance than other successful architectures when dealing with anisotropic 3D resolution.

3.B.2. Evaluation of the Network Mechanism

When the 3D Res-UNet backbone was combined with the ASPP blocks and cascade architecture, the experimental results trained by pCT and CE-CT illustrated significant improvements in most of the evaluation indicators under the effect of local enhancement mechanisms, as shown in **Table 3**. The mean values of DSC, SEN, ASSD and HD95 increased to 76.23 ± 6.45%, 79.14 ± 12.48%, 1.39 ± 5.44mm and 4.72 ± 3.04mm. It is found that the mean PPR decreased from 79.97 ± 13.90% to 77.34 ± 14.04% ($P<0.001$), while the mean SEN increased from 73.90 ± 14.58% to 79.14 ± 12.48% ($p<0.001$). Although the network model changed from under-segmentation to over-segmentation, the model was able to keep the segmentation results in a relatively optimal state. For DSC, which is the key indicator of segmentation performance, a boxplot is shown in **Figure 4**. The differences can be found in the following three statistical intervals, i.e., the 5%-95%, 25%-75% and 50%, and the results show that the model with the ASPP block and cascade architecture, named CMDR-UNet, has a higher average value along with a smaller variance ($P<0.001$).

TABLE 2 | Quantitative comparison of different backbone models for GTV segmentation performance, including mean DSC, PPR, SEN, ASSD and HD95 with standard deviation.

Method/P -value		DSC (%)	PPR (%)	SEN (%)	ASSD (mm)	HD95 (mm)
① 3D CNN (22)		73.67 ± 7.88	76.74 ± 14.71	75.27 ± 14.31	1.84 ± 3.91	6.32 ± 13.77
② 3D Attention-UNet (38)		73.54 ± 7.16	75.95 ± 14.59	**76.04 ± 14.48**	1.80 ± 1.64	6.74 ± 11.91
③ 3D UNet++ (39)		73.87 ± 7.07	77.73 ± 14.40	74.82 ± 14.26	1.53 ± 0.60	5.17 ± 3.04
④ Proposed 3D Res-UNet		**74.49 ± 7.81**	**79.97 ± 13.90**	73.90 ± 14.58	**1.49 ± 0.65**	**5.06 ± 3.30**
P-value	④ *vs.* ①	0.026*	<0.001*	0.041*	0.179	0.190
	④ *vs.* ②	0.007*	<0.001*	0.001*	0.005*	0.043*
	④ *vs.* ③	0.027*	<0.001*	0.107	0.213	0.520

Asterisks (*) indicate that the difference between the proposed 3D Res-UNet method and the competing method is statistically significant ($p < 0.05$) using a paired t-test. The best result is highlighted in bold.

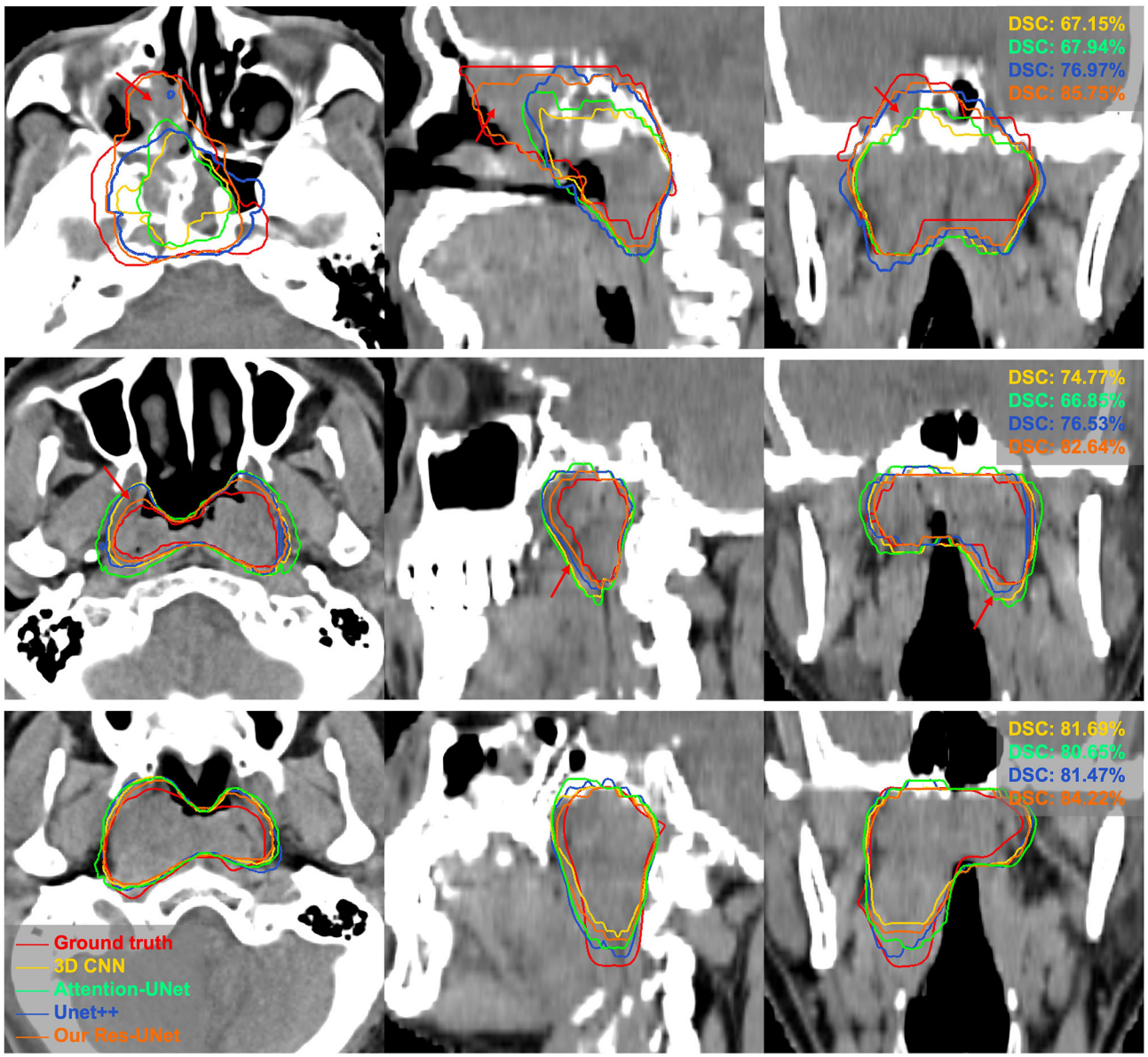


FIGURE 3 | Visual comparison of different networks for GTV segmentation. The red arrows denote false positives and poorly segmented areas. Note that these results derived from the models trained on pCT and CE-CT data are shown in pCT.

TABLE 3 | Comparison of the effects of adding the ASPP blocks and the proposed cascade architecture.

Method/P -value		DSC (%)	PPR (%)	SEN (%)	ASSD (mm)	HD95 (mm)
① Res-UNet		74.49 ± 7.81	**79.97 ± 13.90**	73.90 ± 14.58	1.49 ± 0.65	5.06 ± 3.30
② MDR-UNet		75.16 ± 6.76	78.73 ± 13.50	75.81 ± 13.65	1.68 ± 3.44	5.82 ± 13.57
③ CMDR-UNet		**76.23 ± 6.45**	77.34 ± 14.05	**79.14 ± 12.48**	**1.39 ± 5.44**	**4.72 ± 3.04**
P-value	② *vs.* ①	0.065	0.002*	<0.001*	0.409	0.418
	③ *vs.* ②	<0.001*	<0.001*	<0.001*	0.233	0.244
	③ *vs.* ①	<0.001*	<0.001*	<0.001*	<0.001*	0.002*

Two-tailed p-values <0.05 were considered statistically significant between the proposed different models using paired t-tests. The best result is highlighted in bold.
**p < 0.05 was considered significant. The values were represented as mean ± standard deviation. MDR-UNet: Adding the multiscale dilate CNN of the ASPP module. CMDR-UNet: Adding our proposed cascade architecture.*

To further validate our network mechanism, we extracted two representative examples for boundary segmentation feature maps as shown in **Figure 5**. Because of the effect of the ASPP module, the weight distribution of the feature map is relatively uniform and the tumor boundary is sharp, which are indicated by the red arrows, when compared with the 3D Res-UNet backbone. **Figure 6** shows the visual comparison results of the three different modules for four instances of NPC segmentation along with the DSC quantification results. We found that the ASPP block is more sensitive to the boundaries of the GTV than the Res-UNet backbone, as shown by the red arrows indicating positions, while the cascade architecture further improves the stability of the network. In summary, the multiscale dilated convolution block and proposed cascade architecture can make a large difference in stably outputting automatic segmentation results with high accuracy, which is critical for an algorithm.

3.B.3. Multi-Institution Experiments

The quantitative segmentation results of the models pretrained, fine-tuned and validated on the multi-institution datasets using mean DSC and ASSD with standard deviations were listed in **Table 4**. For the A(pCT) pretrained model, the mean results of the DSC and ASSD for the data with five-fold cross-validation were 74.63% and 1.58 mm, respectively. The DSC value was inferior to the 75.16% obtained for the DSC trained with pCT and CE-CT, as shown in **Table 3**. In these five models pretrained on dataset A, the model with the best performance was applied to the datasets B, C and B+C, the mean DSC and ASSD results were 68.21% and 2.09 mm, 61.64% and 2.13 mm and 64.26% and 2.12mm, respectively. Clearly, these results were not satisfactory.

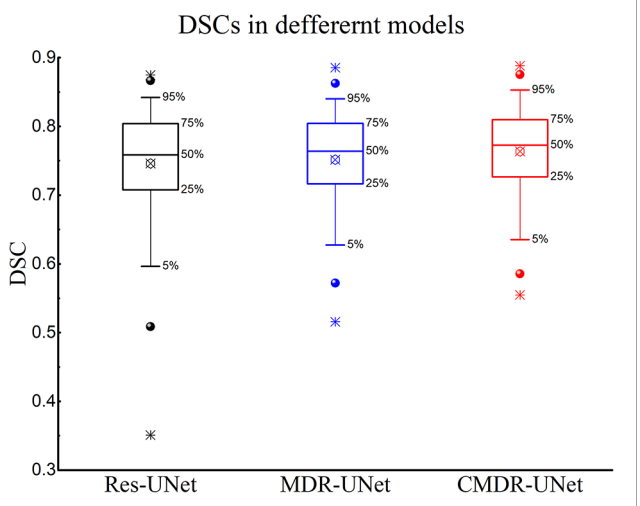


FIGURE 4 | Boxplot showing the DSCs of different models. The symbols ⊠ represents the means, ◕ represents the 1% and 99%, ✳ represents the min and max.

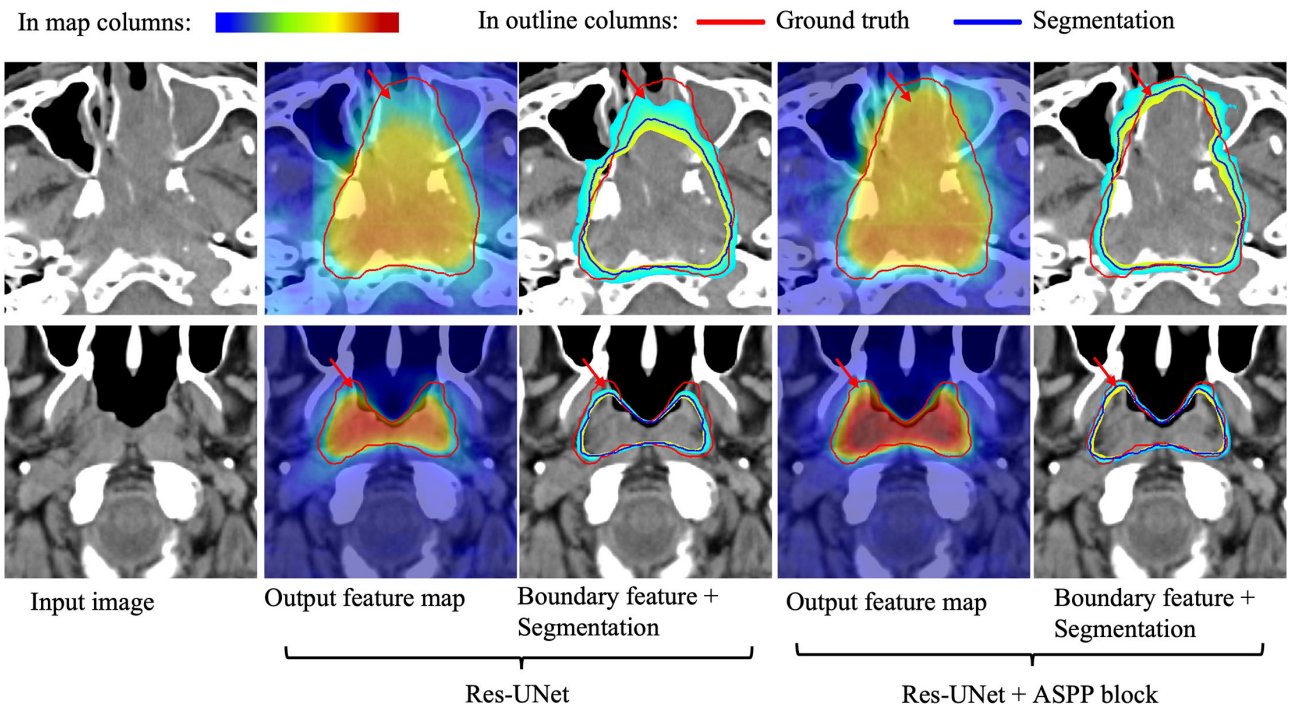


FIGURE 5 | Visual comparison of the boundary feature maps obtained by including the Res-UNet backbone and adding the ASPP block. Map columns: includes output feature maps and boundary feature maps, where warmer colors represent higher attention. Outline columns: red and blue denote the ground truth and automatic segmentation results, respectively. The red arrows denote boundary areas noticed by the ASPP block, which has better boundary segmentation performance. Note that the two rows are from different patients and shown in pCT.

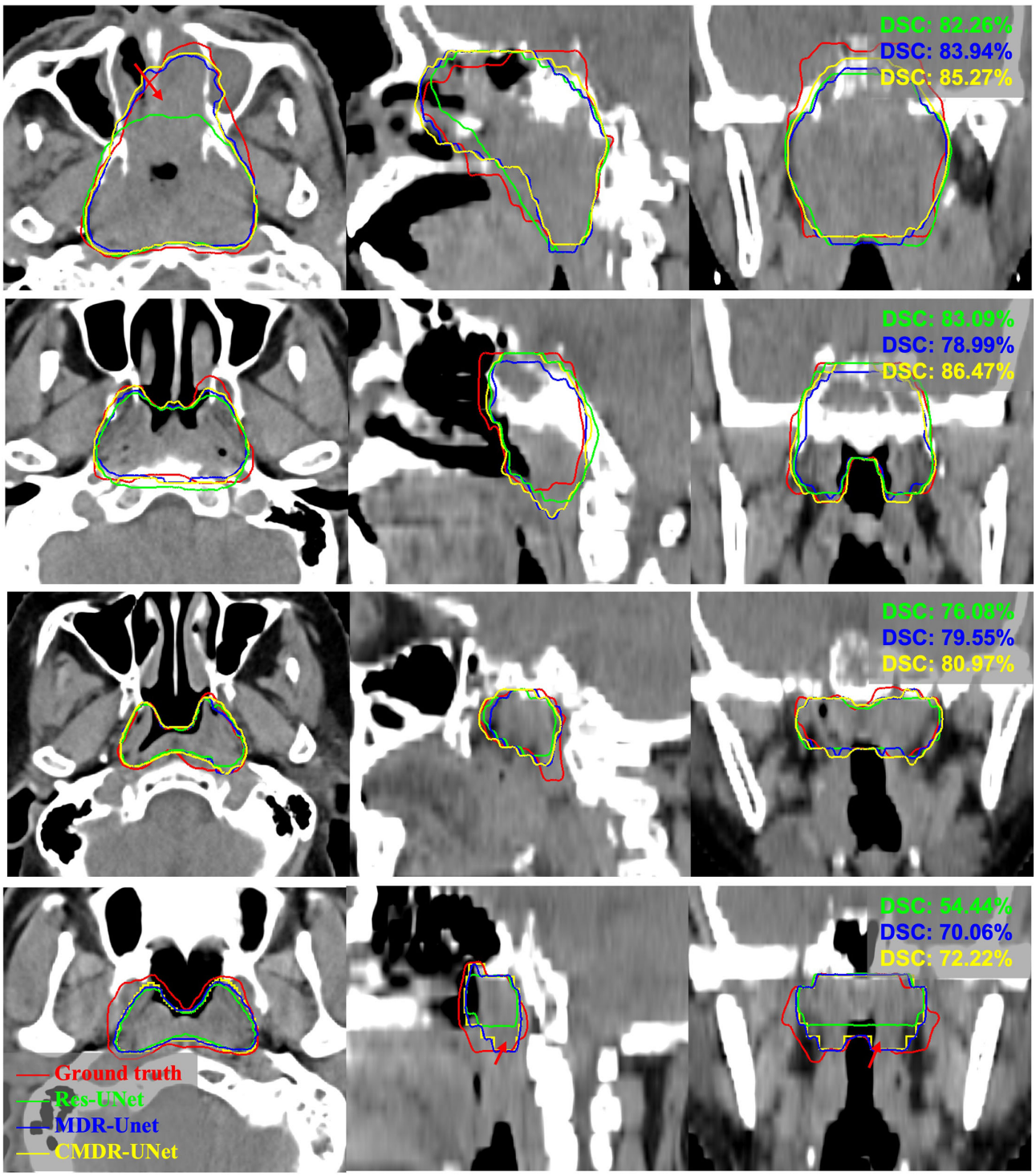


FIGURE 6 | Example pCT images show the level of consistency for GTVnx between the automatic delineation with our method and the ground truth. Red lines denote the human expert-delineated ground truth, and the other lines denote the contours of the automatic delineation.

Then this pretrained model was fine-tuned with 70% of datasets B, C and B+C and the remaining 30% dataset was utilized for test, respectively. The mean DSC and ASSD results reached to 74.49%, 1.87 mm and 73.95%, 1.89mm for institution B, and 69.64%, 1.42 mm and 67.43%, 1.60mm for institution C, respectively. Compared with the model fine-tuned on the single dataset B or C, the evaluation metrics of model fine-tuned on the mixed dataset B+C showed slightly lower

TABLE 4 | The segmentation results of the models pretrained, fine-tuned and validated on the dataset A, B, C and B+C.

Training datasets	Test datasets	DSC (%)	ASSD (mm)
A (pCT)	A	**74.63 ± 7.05**	**1.58 ± 0.94**
	B	68.21 ± 5.51	2.09 ± 0.74
	C	61.64 ± 13.55	2.13 ± 1.02
	B+C	64.26 ± 11.54	2.12 ± 0.93
A Pretrained + B (70%) fine-tuning	① B (30%)	**74.49 ± 6.86**	**1.87 ± 0.67**
A Pretrained + C (70%) fine-tuning	① C (30%)	**69.64 ± 10.02**	**1.42 ± 0.36**
A Pretrained + (B+C) (70%) fine-tuning	② B (30%)	73.95 ± 8.66	1.89 ± 0.78
	② C (30%)	67.43 ± 12.35	1.60 ± 0.49
P-value	② B *vs.* ① B	0.810	0.893
	② C *vs.* ① C	0.180	0.041*

Two-tailed p-values <0.05 were considered statistically significant. The best result is highlighted in bold.
*The values were represented as mean ± standard deviation. *p < 0.05 was considered significant. A: Our institution; B: institution B; and C: MICCAI2019.*

performance, which was probably caused by the pCT scanned from different devices and particularly the ground truth delineated by different doctors. Moreover, the problem of sample shortage leaded to the lack of stability in statistical differences. In summary, these results demonstrate that transfer learning can achieve good performance on external datasets.

To further test the robustness and transferability of the model application, the original and target domains were switched for the following experiments. The segmentation results of the models pretrained, fine-tuned and validated on the dataset B, C, B+C and A were listed in the **Supplementary Table 1**. The pretrained network model achieved good performance on the B and C datasets with the mean DSC and ASSD of 72.84%, 1.89 mm and 65.77%, 2.56mm, respectively. It shows that the model also has good robustness on other datasets. However, direct application of the trained models based on B and C datasets to the A dataset shows poor performance with mean DSC and ASSD as low as 61.23%, 2.80mm and 58.88%, 2.72mm, respectively. Meanwhile, dataset A (70%) was used to fine-tune the models pretrained on the B, C and B+C datasets, respectively. The mean DSC and ASSD results reached 74.81%, 1.36mm, 74.27%, 1.40mm and 74.46%, 1.61mm, respectively. It is easy to notice that there is no improvement in the above results compared to the dataset directly training results of 74.46%, 1.45mm (P>0.05). These results suggest that models trained on small datasets have poor generalization ability and transfer learning being introduced to large datasets may not improve performance.

4 DISCUSSION AND CONCLUSION

Blurred tumor boundaries and large shape variations have always been huge challenges for GTV segmentation of NPC in lower-contrast CT images. In this work, a 3D automatic segmentation algorithm has been proposed to solve these issues, in which the concept of multiscale local enhancement is employed with the foundation of the 3D Res-UNet backbone model. First, for small target areas, such as GTV, we introduced the multiscale ASPP-block in skip-connection to guide the network to focus more on the target area, especially for the boundary during learning, to promote the segmentation performance. Second, benefiting from the excellent ability to capture both global and local feature information simultaneously, a cascade architecture was adopted in training for outputting robust segmentation results. Moreover, we adopted an automatic preprocessing method to reduce the image background to solve the problem of heavy computation and memory consumption in loading the 3D network model data.

In multi-institution experiments, as shown in **Tables 3**, **4**, CE-CT is helpful for improving segmentation performance. Due to the differences in data sources and manual delineation for different institutions, as demonstrated in the literature (22), DSCs can vary from 71% to 80% compared to the manual contours of eight radiation oncologists with ground truth contours. In our experiments, it was found that direct application of the model trained with the internal dataset to the external dataset yielded suboptimal results. However, the performance of the models can be significantly improved by transfer learning, which only requires fine-tuning the network with a smaller dataset to output comparable results obtained from a larger dataset. Although transfer learning may not be used to improve the performance of applying a model trained on a small dataset to the results of a large dataset, it can output comparable results with higher training speed. In addition, the models were applied to the small datasets B and MICCAI, where the DSC could reach 72.84% and 65.77%, respectively. Another study based on MICCAI data showed that the DSC of the ensemble multi-scale model was only 65.66% (40), which is comparable to the result (65.77%) but lower than the proposed method with transfer learning (69.64%). It shows that the model is robust to small datasets. In general, although a trained model cannot not be valid for all dataset estimations, the proposed model can achieve good performance through transfer learning.

Indeed, limitations have been observed within the current model. In the multi-institution experiment, since the labels are delineated by different radiation oncologists, the dataset cannot be evaluated uniformly when multiple institutions are integrated into the network model for supervised training. In future work, the segmentation performance will be further improved in the following three aspects: 1) To overcome the differences in tumor delineations by various doctors, the accuracy and consistency of delineation will be further advanced. 2) Considering the

significant differences between numerous data sources, a semi-supervised method will be introduced in training to enhance the robustness of the network. 3) Multimodal data will be utilized to assist CT image segmentation.

Extensive experiments on our CT dataset show that our proposed CMDR-UNet method based on the modified 3D Res-UNet backbone outperforms other state-of-the-art methods and achieves the best results for the quantitative indicators DSC, PPR, ASSD and HD95. In multi-institution experiments, due to the differences in data sources and manual delineations for different institutions, the segmentation results of other datasets acquired from a single institution trained model is unsatisfactory, although this issue can be resolved by transfer learning. This finding is desirable because it partially reflects the universality of our proposed algorithm for clinical application.

DATA AVAILABILITY STATEMENT

The data analyzed in this study is subject to the following licenses/restrictions: The datasets for this article are not publicly available because the data were collected by multiple centers and publicly accessible data is not approved. Requests to access these datasets should be directed to Zhenhui Dai, dzh_dzh@126.com.

AUTHOR CONTRIBUTIONS

GY and ZD: design of methodology, development and implement of models, original drafting. YZ and ZC: data curation and preprocessing. LZ, BZ, CC, and QH: experimental results analysis, draft reviewing. JT and FL: data collection. WY and XW: design of methodology, review and editing. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Natural Science Foundation of China (No.82172020), the Guangzhou Science and Technology Plan (No.202102010264), and the Youth Committee of Medical Engineering Branch of Guangdong Medical Association Research Projects (No.2019-GDMAZD-01).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.827991/full#supplementary-material

REFERENCES

1. Wei WI, Sham J. Nasopharyngeal Carcinoma. *Lancet* (2005) 365(9476):2041–54. doi: 10.1016/S0140-6736(05)66698-6
2. Guo F, Shi C, Li X, Wu X, Zhou J, Lv J. Image Segmentation of Nasopharyngeal Carcinoma Using 3D CNN With Long-Range Skip Connection and Multi-Scale Feature Pyramid. *Soft Comput (Berlin Germany)* (2020) 24(16):12671–80. doi: 10.1007/s00500-020-04708-y
3. Zhao L, Lu Z, Jiang J, Zhou Y, Wu Y, Feng Q. Automatic Nasopharyngeal Carcinoma Segmentation Using Fully Convolutional Networks With Auxiliary Paths on Dual-Modality PET-CT Images. *J Digit Imaging* (2019) 32(3):462–70. doi: 10.1007/s10278-018-00173-0
4. Men K, Chen X, Zhang Y, Zhang T, Dai J, Yi J, et al. Deep Deconvolutional Neural Network for Target Segmentation of Nasopharyngeal Cancer in Planning Computed Tomography Images. *Front Oncol* (2017) 7:315. doi: 10.3389/fonc.2017.00315
5. Das IJP, Moskvin VP, Johnstone PAM. 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. Tomé SWA. Emphasizing Conformal Avoidance Versus Target Definition for IMRT Planning in Head-And-Neck Cancer. *Int J Radiat OncologyBiologyPhysics* (2010) 77(3):950–8. doi: 10.1016/j.ijrobp.2009.09.062
7. Breen SL, Publicover J, De Silva S, Pond G, Brock K, O Sullivan B, et al. Intraobserver and Interobserver Variability in GTV Delineation on FDG-PET-CT Images of Head and Neck Cancers. *Int J Radiat OncologyBiologyPhysics* (2007) 68(3):763–70. doi: 10.1016/j.ijrobp.2006.12.039
8. Feng MU, Demiroz C, Vineberg KA, Balter JM, Eisbruch A. Intra-Observer Variability of Organs at Risk for Head and Neck Cancer: Geometric and Dosimetric Consequences. *Int J Radiat OncologyBiologyPhysics* (2010) 78(3): S444–5. doi: 10.1016/j.ijrobp.2010.07.1044
9. Fountzilas G, Psyrri A, Giannoulatou E, Tikas I, Manousou K, Rontogianni D, et al. Prevalent Somatic BRCA1 Mutations Shape Clinically Relevant Genomic Patterns of Nasopharyngeal Carcinoma in Southeast Europe. *Int J Cancer* (2018) 142(1):66–80. doi: 10.1002/ijc.31023
10. Iglesias JE, Sabuncu MR. Multi-Atlas Segmentation of Biomedical Images: A Survey. *Med Image Anal* (2015) 24(1):205–19. doi: 10.1016/j.media.2015.06.012
11. Sims R, Isambert A, Grégoire V, Bidault F, Fresco L, Sage J, et al. A Pre-Clinical Assessment of an Atlas-Based Automatic Segmentation Tool for the Head and Neck. *Radiother Oncol* (2009) 93(3):474–8. doi: 10.1016/j.radonc.2009.08.013
12. Han X, Hoogeman MS, Levendag PC, Hibbard LS, Teguh DN, Voet P, et al. Atlas-Based Auto-Segmentation of Head and Neck CT Images. *Med Image Comput Comput Assist Interv* (2008) 11(Pt 2):434–41. doi: 10.1007/978-3-540-85990-1_52
13. Teguh DN, Levendag PC, Voet PWJ, Al-Mamgani A, Han X, Wolf TK, et al. Clinical Validation of Atlas-Based Auto-Segmentation of Multiple Target Volumes and Normal Tissue (Swallowing/Mastication) Structures in the Head and Neck. *Int J Radiat OncologyBiologyPhysics* (2011) 81(4):950–7. doi: 10.1016/j.ijrobp.2010.07.009
14. Tao C, Yi J, Chen N, Ren W, Cheng J, Tung S, et al. Multi-Subject Atlas-Based Auto-Segmentation Reduces Interobserver Variation and Improves Dosimetric Parameter Consistency for Organs at Risk in Nasopharyngeal Carcinoma: A Multi-Institution Clinical Study. *Radiother Oncol* (2015) 115 (3):407–11. doi: 10.1016/j.radonc.2015.05.012
15. Lin T, Goyal P, Girshick R, He K, Dollar P. Focal Loss for Dense Object Detection. *IEEE T Pattern Anal* (2020) 42(2):318–27. doi: 10.1109/TPAMI.2018.2858826
16. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. *Comput Vision Pattern Recogn: IEEE*. Las Vegas, NV, USA (2016). p:770–8. doi: 10.1109/CVPR.2016.90
17. Ronneberger O, Fischer P, Brox T. *U-Net: Convolutional Networks for Biomedical Image Segmentation*. Munich, Germany: Medical Image Computing and Computer Assisted Intervention (MICCAI) (2015). pp. 234–41.
18. Long J, Shelhamer E, Darrell T. *Fully Convolutional Networks for Semantic Segmentation. Computer Vision and Pattern Recognition*. Boston: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2015). pp. 3431–40.
19. Dong H, Yang G, Liu F, Mo Y, Guo Y. Automatic Brain Tumor Detection and Segmentation Using U-Net Based Fully Convolutional Networks. Edinburgh, UK: 21st Annual Conference, MIUA (2017). p:506–17. doi: 10.1007/978-3-319-60964-5_44

20. 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
21. Men K, Zhang T, Chen X, Chen B, Tang Y, Wang S, et al. Fully Automatic and Robust Segmentation of the Clinical Target Volume for Radiotherapy of Breast Cancer Using Big Data and Deep Learning. *Physica Med* (2018) 50:13–9. doi: 10.1016/j.ejmp.2018.05.006
22. Lin L, Dou Q, Jin Y, Zhou G, Tang Y, Chen W, 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
23. Chen H, Qi Y, Yin Y, Li T, Liu X, Li X, et al. MMFNet: A Multi-Modality MRI Fusion Network for Segmentation of Nasopharyngeal Carcinoma. *Neurocomputing* (2020) 394:27–40. doi: 10.1016/j.neucom.2020.02.002
24. Li Q, Xu Y, Chen Z, Liu D, Feng S, Law M, et al. Tumor Segmentation in Contrast-Enhanced Magnetic Resonance Imaging for Nasopharyngeal Carcinoma: Deep Learning With Convolutional Neural Network. *BioMed Res Int* (2018) 2018:1–7. doi: 10.1155/2018/9128527
25. Ma Z, Zhou S, Wu X, Zhang H, Yan W, Sun S, et al. Nasopharyngeal Carcinoma Segmentation Based on Enhanced Convolutional Neural Networks Using Multi-Modal Metric Learning. *Phys Med Biol* (2019) 64 (2):25005. doi: 10.1088/1361-6560/aaf5da
26. Daoud B, Morooka KI, Kurazume R, Leila F, Mnejja W, Daoud J. 3D Segmentation of Nasopharyngeal Carcinoma From CT Images Using Cascade Deep Learning. *Comput Med Imag Grap* (2019) 77:101644. doi: 10.1016/j.compmedimag.2019.101644
27. Li S, Xiao J, He L, Peng X, Yuan X. The Tumor Target Segmentation of Nasopharyngeal Cancer in CT Images Based on Deep Learning Methods. *Technol Cancer Res T* (2019) 18:1–8. doi: 10.1177/1533033819884561
28. Wang X, Yang G, Zhang Y, Zhu L, Xue X, Zhang B, et al. Automated Delineation of Nasopharynx Gross Tumor Volume for Nasopharyngeal Carcinoma by Plain CT Combining Contrast-Enhanced CT Using Deep Learning. *J Radiat Res Appl Sci* (2020) 13(1):568–77. doi: 10.1080/16878507.2020.1795565
29. Pereira S, Oliveira A, Alves V, Silva CA. On Hierarchical Brain Tumor Segmentation in MRI Using Fully Convolutional Neural Networks: A Preliminary Study. *IEEE Portuguese Meeting Bioeng* (2017), 1–4. doi: 10.1109/ENBENG.2017.7889452
30. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. *3d U-Net: Learning Dense Volumetric Segmentation From Sparse Annotation.* Athens: Medical Image Computing and Computer-Assisted Intervention (MICCAI). (2016). pp. 424–32.
31. He K, Zhang X, Ren S, Sun J. *Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR): IEEE.* (2016). pp. 770–8.
32. Chen L, Papandreou G, Kokkinos I, Murphy K, Yuille AL. DeepLab: Semantic Image Segmentation With Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE T Pattern Anal* (2018) 40(4):834–48. doi: 10.1109/TPAMI.2017.2699184
33. Chen L, Zhu Y, Papandreou G, Schroff F, Adam H. Encoder-Decoder With Atrous Separable Convolution for Semantic Image Segmentation. Munich: European Conference on Computer Vision (2018) 11211:833–51. doi: 10.1007/978-3-030-01234-2_49
34. Ulyanov D, Vedaldi A, Lempitsky V. Instance Normalization: The Missing Ingredient for Fast Stylization. *arXiv: Comput Vision Pattern Recogn* (2016). arXiv:1701.02096. doi: 10.48550/arXiv.1607.08022
35. Ioffe S, Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. Proceedings of the 32nd International Conference on International Conference on Machine Learning: Lille, France (2015), 448–56.
36. Maas AL, Hannun AY, Ng AY. *Rectifier Nonlinearities Improve Neural Network Acoustic Models. Proc. Icml*. (2013). p. 3.
37. He K, Zhang X, Ren S, Sun J. *Delving Deep Into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. International Conference on Computer Vision*. Santiago, Chile (2015). pp. 1026–34.
38. Oktay O, Schlemper J, Folgoc LL, Lee M, Heinrich M, Misawa K, et al. *Attention U-Net: Learning Where to Look for the Pancreas*. (2018).
39. Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J. UNet++: Redesigning Skip Connections to Exploit Multiscale Features in Image Segmentation. *IEEE T Med Imaging* (2020) 39(6):1856–67. doi: 10.1109/TMI.2019.2959609
40. Mei H, Lei W, Gu R, Ye S, Sun Z, Zhang S, et al. Automatic Segmentation of Gross Target Volume of Nasopharynx Cancer Using Ensemble of Multiscale Deep Neural Networks With Spatial Attention. *Neurocomputing* (2021) 438:211–22. doi: 10.1016/j.neucom.2020.06.146

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

frontiers | Frontiers in Oncology

ORIGINAL RESEARCH
published: 26 May 2022
doi: 10.3389/fonc.2022.844052

Application of Multi-Scale Fusion Attention U-Net to Segment the Thyroid Gland on Localized Computed Tomography Images for Radiotherapy

*Xiaobo Wen[1], Biao Zhao[1], Meifang Yuan[1], Jinzhi Li[1], Mengzhen Sun[1], Lishuang Ma[1], Chaoxi Sun[2] and Yi Yang[1]**

[1] Department of Radiotherapy, Yunnan Cancer Hospital, Kunming, China, [2] Department of Neurosurgery, Yunnan Cancer Hospital, Kunming, China

OPEN ACCESS

Edited by:
Jose Eduardo Villarreal Barajas, Royal Devon and Exeter Hospital, United Kingdom

Reviewed by:
Gregory Sharp, Massachusetts General Hospital and Harvard Medical School, United States
Weiwei Zong, Henry Ford Health System, United States

***Correspondence:**
Yi Yang
yiyangrt@126.com

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

Received: 27 December 2021
Accepted: 26 April 2022
Published: 26 May 2022

Citation:
Wen X, Zhao B, Yuan M, Li J, Sun M, Ma L, Sun C and Yang Y (2022) Application of Multi-Scale Fusion Attention U-Net to Segment the Thyroid Gland on Localized Computed Tomography Images for Radiotherapy. Front. Oncol. 12:844052. doi: 10.3389/fonc.2022.844052

Objective: To explore the performance of Multi-scale Fusion Attention U-Net (MSFA-U-Net) in thyroid gland segmentation on localized computed tomography (CT) images for radiotherapy.

Methods: We selected localized radiotherapeutic CT images from 80 patients with breast cancer or head and neck tumors; label images were manually delineated by experienced radiologists. The data set was randomly divided into the training set (n = 60), the validation set (n = 10), and the test set (n = 10). We expanded the data in the training set and evaluated the performance of the MSFA-U-Net model using the evaluation indices Dice similarity coefficient (DSC), Jaccard similarity coefficient (JSC), positive predictive value (PPV), sensitivity (SE), and Hausdorff distance (HD).

Results: For the MSFA-U-Net model, the DSC, JSC, PPV, SE, and HD values of the segmented thyroid gland in the test set were 0.90 ± 0.09, 0.82± 0.11, 0.91 ± 0.09, 0.90 ± 0.11, and 2.39 ± 0.54, respectively. Compared with U-Net, HRNet, and Attention U-Net, MSFA-U-Net increased DSC by 0.04, 0.06, and 0.04, respectively; increased JSC by 0.05, 0.08, and 0.04, respectively; increased SE by 0.04, 0.11, and 0.09, respectively; and reduced HD by 0.21, 0.20, and 0.06, respectively. The test set image results showed that the thyroid edges segmented by the MSFA-U-Net model were closer to the standard thyroid edges delineated by the experts than were those segmented by the other three models. Moreover, the edges were smoother, over–anti-noise interference was stronger, and oversegmentation and undersegmentation were reduced.

Conclusion: The MSFA-U-Net model could meet basic clinical requirements and improve the efficiency of physicians' clinical work.

Keywords: U-Net model, multi-scale fusions, medical-image segmentation, thyroid, radiotherapy

INTRODUCTION

Head and neck tumors and breast cancer are currently the tumors with relatively higher morbidity and mortality rates worldwide (1). In 2020, 19.29 million new cancer cases occurred worldwide, of which 4.57 million (23.7%) were in China. Radiotherapy is an effective and common method for treating head and neck cancer and breast cancer (2–4). Accurately delineating organs at risk (OARs) when designing radiotherapy plans can effectively avoid radiation side effects. At present, physicians are responsible for outlining OARs, making the process subjective, time consuming, and labor intensive.

The rapid development of artificial intelligence (AI) enabled Ronneberger et al. (5) to propose the U-Net neural network model in 2015. The delineation method based on deep learning (DL) has gradually been developed and applied in clinical work (6–10). Ye et al. (7) used an improved model, a dense-connectivity embedding U-Net, to train and segment the T1 and T2 magnetic resonance imaging (MRI) images of 44 patients with nasopharyngeal carcinoma; the authors obtained a Dice similarity coefficient (DSC) of 0.87 after tenfold cross-validation. Automatized delineation of the thyroid gland on localized CT images for radiotherapy has been critical in radiotherapy planning (11). Zhai et al. (12) found that the patients who received thyroid mean radio dose of ≥45 Gy had a 4.9 times increased risk of hypothyroidism than those with lower mean radio dose. Akın et al. (13) conducted a retrospective study on 122 patients who received three-dimensional conformal radiation therapy (3D-CRT) for breast cancer. They found that ①functional abnormalities occurred in the thyroid gland which was exposed to total radiation doses of 26 to 30 Gy; ②44% of the patients were exposed to a radiation dose of >26 Gy. Other studies showed that 2 years after patients with head and neck tumors received radiotherapy, their incidence of hypothyroidism was 36%; moreover, this incidence increased along with follow-up time (14, 15). Therefore, in radiotherapeutic planning, radiation must be limited to the thyroid gland. Narayanan D. et al. (16) used multi-atlas label fusion (MALF) and random forest (RF) to automatically segment the thyroid gland on CT and found that MALF with RF presented better segmentation with the DSC being 0.76 ± 0.11, which was significantly better than the individual MALF and RF methods. Chang et al. (17) used a progressive learning vector quantization neural network to segment the thyroid on CT and their experimental results showed that the proposed method could effectively segment thyroid glands with its average SE being 88.43%. He et al. (18) used deep convolutional neural network to segment the thyroid gland on noncontrast-enhanced head and neck CTs and found that their proposed method had significantly improved performance. Considering that CT localization for radiotherapy involves a simulated-positioning, large-aperture CT (SOMATOM Sensation Open, 24 rows, Φ85 cm; Siemens Healthcare, Forchheim, Germany), which is limited by small size and poor image resolution, automatic segmentation of the thyroid gland based on a DL model is difficult. The performance of such a model on localized CT images for radiotherapy requires further exploration. In the deep learning study, the combination of HRNet and SE is common (19). In HRNet, multiple parallel networks with different resolutions are used to extract features and multi-scale fusions are repeatedly performed during feature extraction to ensure that the model can fully obtain information of different scales (20). The cSE module enables the model to pay more attention to major channel features and suppresses those minor channel features (21). Therefore, in this study, we proposed a model that combined a Spatial Squeeze and Channel Excitation Block (cSE) attention mechanism with HRNet on the basis of U-Net and used it to segment the thyroid gland on localized CT images to help delineate the gland as an OAR in radiotherapy.

Abbreviations: CT, computed tomography; MSFA-U-Net, Multi-scale Fusion Attention U-Net; DSC, Dice similarity coefficient; JSC, Jaccard similarity coefficient; PPV, positive predictive value; SE, sensitivity; HD, Hausdorff distance; HRNet, High-Resolution net; OARs, organs at risk; AI, artificial intelligence; DL, deep learning; MRI, magnetic resonance imaging; MALF, multi-atlas label fusion; RF, random forest 3D, three-dimensional; 3D-CRT, three-dimensional conformal radiotherapy; ROI, region of interest; cSE, Spatial Squeeze and Channel Excitation Block; BN, batch normalization; GAP, global average pooling; DLC, deep learning contouring.

MATERIALS AND METHODS

Data Set Acquisition

We obtained the experimental data set in this study from 80 patients with nasopharyngeal carcinoma or breast cancer who were admitted to the Department of Radiotherapy of Yunnan Cancer Hospital (Kunming, China) from June 2014 to April 2019. Localization for each patient was simulated using a SOMATOM Sensation Open 24 CT scanner. CT images were obtained in Digital Imaging and Communications in Medicine (DICOM) format with slices being 5 or 3 mm thick and pixels being 512 × 512. Senior radiotherapists drew the label images on the CT images in DICOM, using 3D Slicer software version 4.11. The label images were converted from DICOM to PNG format (**Figure 1**).

We divided the data set (6:1:1) into training, validation, and test sets. Due to the small number of medical data sets and the high cost of drawing, collecting a sufficiently large number of data sets was difficult; however, a training data set that was too small would have created a risk of overfitting the model. To avoid this risk, in this study we expanded the training sample data set size by means of rotation, flipping, zooming, and shearing.

Data Set Preprocessing

To better highlight the region of interest (ROI), we first converted CT image pixels into Hounsfield unit (HU) values and then adjusted the window width and level of the converted data to highlight the thyroid gland. Finally, we used adaptive histogram equalization to further enhance the contrast and normalize the images.

Model Framework

We improved our model based on the U-Net and HRNet model architectures, called MSFA-U-Net. Main improvements were (a) replacing two feature extraction convolutions of different resolutions in the U-Net downsampling process with multiple convolution blocks in HRNet and feature fusion between different

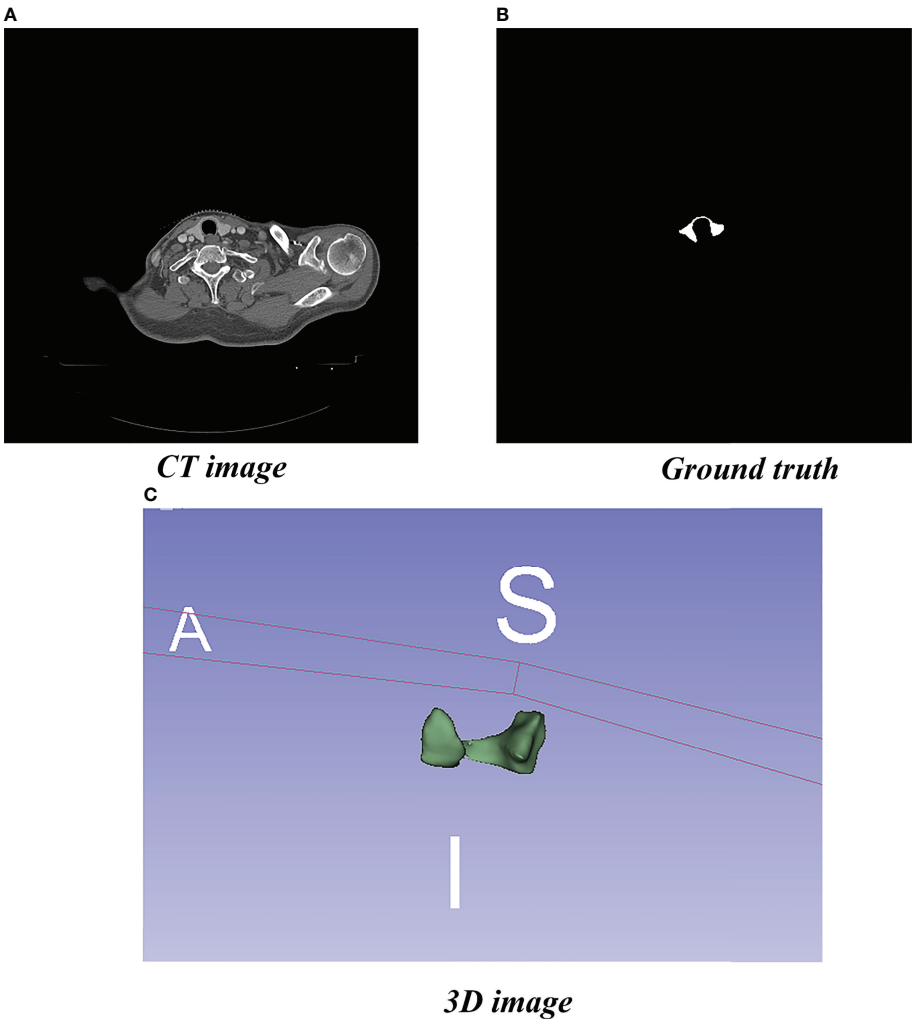


FIGURE 1 | Localized CT image, Ground truth, and 3D image. **(A)** Standard image of the imported model (CT image). **(B)** Corresponding label image (Ground truth). **(C)** Thyroid gland drawn in 3D.

scales: and (b) introducing the cSE attention mechanism into each convolution block (**Figure 2**). In the downsampling process of the model, we connected a cSE module after extracting two 3 × 3 convolutional features and fused the input features with the post–scale operation features by means of a residual connection that consisted of a 1 × 1 2D convolution and a normalization layer (22) [batch normalization (BN)]. In the cSE module, we used a global average pooling (GAP) layer to convert a feature map from channel × height × width to channel × 1 × 1 and then used Dense to reduce the feature channel by half, which we achieved by activating the function Relu. Next, we restored the feature channel to normal size using Dense and activated it using the function Sigmoid. Finally, we obtained a calibrated feature map *via* channelwise multiplication. The schematic diagram of the residual connection and cSE module structure is shown in **Figure 3**. Residual connection can prevent gradient vanishing and gradient explosion during training (23). Moreover, the cSE module could effectively reflect relationships between different channels and assign different weights, enabling the model to focus on important features for accurate segmentation of the thyroid gland during the training process. The whole module is called an Attention Resblock (**Figures 2**, **3**). The traditional U-Net model uses the maximum pooling layer to perform downsampling and reduce the number of parameters; this method can lead to loss of information during feature extraction. Therefore, in this study we used stepped convolution for downsampling. Stride convolution can remove redundant information, thereby reducing the size of the feature map. Our model used multiple branches of different resolutions to extract features in parallel during the training process, and it performed feature fusion among different scales after each attention residual block to achieve strong semantic information and precise location during the training process. One or more transposed convolutions (3 × 3) were used in the conversion from low to high resolution, while one or more stride convolutions (3 × 3) were used in the conversion from high to low resolution (**Figure 3**). In the upsampling, the attention residual block replaced the two

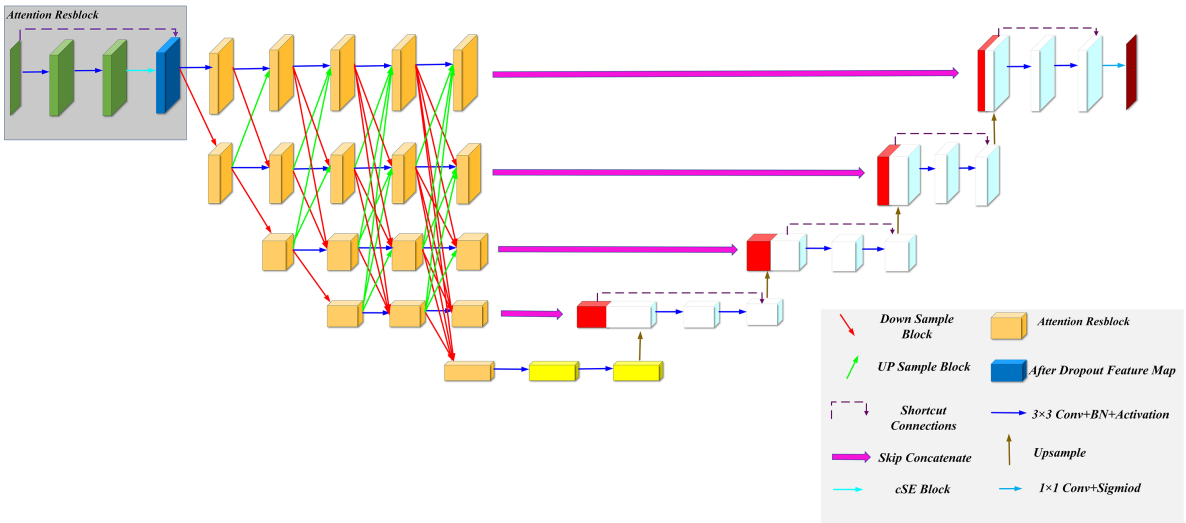


FIGURE 2 | MSFA-U-Net structure.

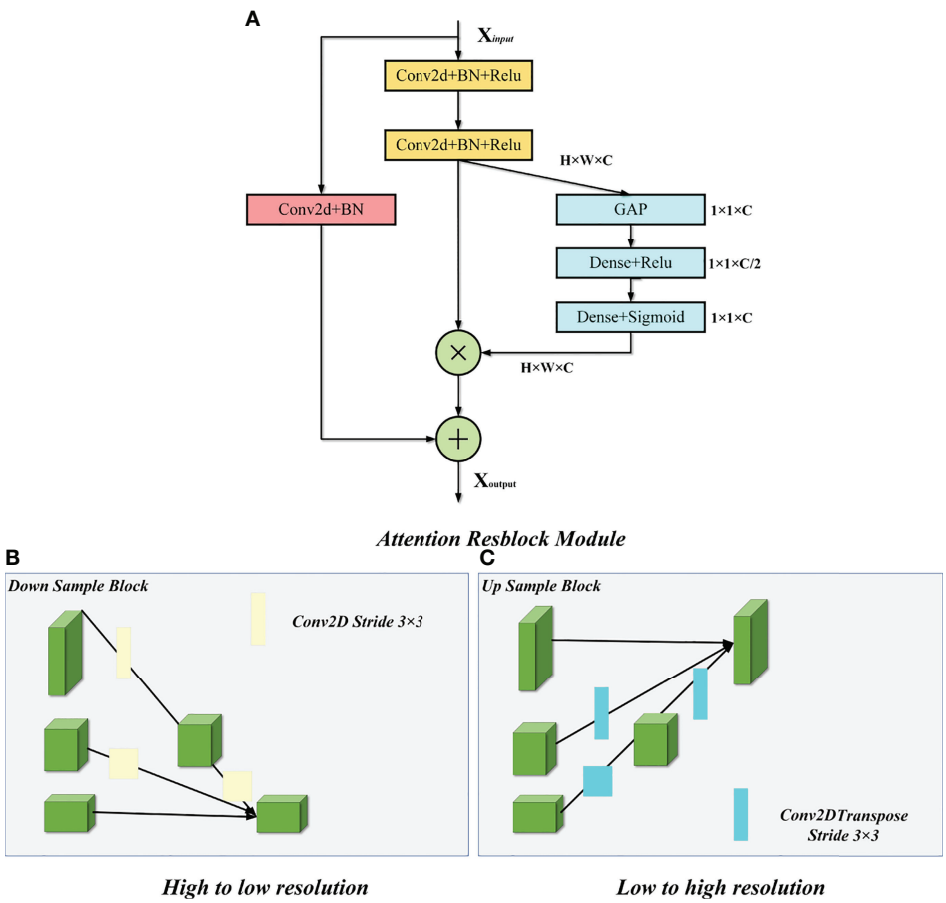


FIGURE 3 | Attention Resblock Module and feature fusion of different scales. **(A)** Attention Resblock Module; blue cuboid = cSE module, red cuboid = Resblock module. **(B)** One or more stride convolutions (3 × 3) were used in the conversion from high to low resolution. **(C)** One or more transposed convolutions (3 × 3) were used in the conversion from low to high resolution.

convolution operations in U-Net to avoid excessive parameters. Meanwhile, we added a dropout layer after each shortcut connection (parameter set to 0.2) to avoid a decrease in generalization caused by overfitting resulting from multiple feature fusions between different scales during the training process.

Model Operating Environment and Parameters

We used TensorFlow software version 2.4.0 (Google Brain Team, 2015; Mountain View, CA, USA) and Keras software version 2.4.3 (Chollet, 2015) to build the model, and Python 3 (Van Rossum and Drake, 2009) to program it. In addition, we used a Windows 10 64-bit operating system (Microsoft Corp., Redmond, WA, USA) with the following hardware: central processing unit (CPU), Intel Core i9-10900 KF @ 3.70 GHz (Intel Corp., Santa Clara, CA, USA); graphics card, NVIDIA GTX3090 24 G (NVIDIA Corp., Santa Clara, CA, USA); and 128 GB memory. Model hyperparameters were selected from the best results according to the experimental conditions (**Table 1**). Batch Size represents the number of input images per iteration, Epoch represents the batch to be trained, Image Size represents the input size of the image, Learning Rate represents the initial learning rate using exponential decay, Decay steps indicate how many steps have been experienced for a learning rate decay, and Decay Rate indicates the learning rate decay coefficient.

Loss Function

Due to its small size, the thyroid gland occupies minimal space on a CT image. Therefore, use of the traditional cross-entropy loss function would leave the model more inclined to predict the background and thus unable to accurately identify the thyroid gland. Milletari et al. (24) proposed a loss function for sample imbalance in medical-image segmentation while researching V-Net-Dice loss function, which is based on DSC. It directly compares the overlap between the model prediction and real segmentation, thereby effectively solving the problem of serious thyroid imbalance. The Dice loss function is calculated according to formula (1.1) below:

$$DL = 1 - 2\star\frac{|X \cap Y| + \varepsilon}{|X| + |Y| + \varepsilon}, \tag{1.1}$$

where X represents the label matrix of the real thyroid gland, Y is the prediction matrix of the model predicting the thyroid gland, and ε represents a constant included to avoid division by zero.

Evaluation Indices

We used the common indices of DSC, JSC, PPV, SE, and HD to further evaluate the generalization ability and segmentation accuracy of the model.

DSC (25) and JSC (26) were calculated according to formulas (1.2) and (1.3), respectively:

$$DSC = 2\star\frac{|X \cap Y|}{|X| + |Y|}, \tag{1.2}$$

$$JSC = \frac{|X \cap Y|}{|X \cup Y|}, \tag{1.3}$$

where X represents the standard segmentation map drawn by a radiologist, Y is the prediction image segmented by the neural-Network model, and $|X \cap Y|$ represents the overlap between the standard map drawn by the radiologist and the model-predicted image. The value range of DSC and JSC is 0–1; values closer to 1 indicate better predictive ability.

PPV (27) and SE (28) were calculated according to formulas (1.4) and (1.5), respectively:

$$PPV = \frac{TP}{TP + FP}, \tag{1.4}$$

$$SE = \frac{TP}{TP + FN}, \tag{1.5}$$

where TP represents the correctly predicted foreground target value, FP represents the incorrectly predicted foreground target value, and FN represents the incorrectly predicted background target value.

HD (29) was calculated according to formula (1.6):

$$H(X, Y) = \max(h(X, Y), h(Y, X)), \tag{1.6}$$

where:

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

$$= \max_{y \in Y} \min_{x \in X} \| y - x \|$$

Smaller values of HD indicate better predictive ability.

Comparison Model Design

To prove the validity of the proposed MSFA-U-Net model, we selected three model architectures related to MSFA-U-Net and conducted comparative experiments:

1. U-Net (5): A U-shaped symmetrical structure composed of upsampling, downsampling, and skip connection. The skip connection effectively combines feature information among different resolutions and compensates for the loss of high-resolution features in the downsampling process. It could also output the feature map more accurately. U-Net is a widely used model in medicine.

TABLE 1 | Network training parameters.

Model	Batch Size	Epoch	Image Size	Learning Rate	Decay Steps	Decay_Rate
U-Net	2	120	512 × 512	1e-5		
HRNet	2	120	512 × 512	8e-5	300	0.96
Attention U-Net	2	120	512 × 512	8e-4	300	0.96
MSFA-U-Net	2	120	512 × 512	2e-4	300	0.96

2. HRNet (20): This model maintains high-resolution output during feature extraction. It has multiple parallel subnets with different resolutions to compress and extract features and to fuse features on multiple scales in order to obtain richer high-resolution features. In their original study, the authors used bilinear-interpolation upsampling. To better extract features for fusion, in this study we used transposed convolution to convert from low to high resolution.
3. Attention U-Net (30): This model introduces an attention-gating mechanism, which gives the information in the jump connection of the U-Net model different weights and enables it to pay more attention to the ROI.

EXPERIMENTAL RESULTS

Qualitative Analysis of Results

Figure 4 shows the results of the thyroid gland segmentation of the four models in the test set. **Figure 5** shows the coverage map of the four models on the CT image of the radiotherapy location. Given the presence of many blood vessels and soft tissues with similar gray levels around the thyroid gland, oversegmentation and undersegmentation are expected in edge segmentation. As shown in **Figure 4**, some of the surrounding blood vessels and soft tissues were mistakenly segmented as part of the left lobe of the thyroid gland when the U-Net model segmented that lobe. Although HRNet and Attention U-Net decreased oversegmentation of the surrounding soft tissues and blood vessels, some noise points and uneven edges remained. However, the MSFA-U-Net architecture used in this study achieved smooth edges and decreased noise. Moreover, although we adjusted the window width and level and adopted adaptive contrast enhancement, some lesions at some levels of the thyroid might have been less obviously enhanced. At these levels, MSFA-U-Net exhibited more-robust thyroid segmentation than the other three models. In summary, compared with those other three models, MSFA-U-Net improved the performance of thyroid gland segmentation on localized CT images for radiotherapy.

Quantitative Analysis of Results

Table 2 compares the results for the four models in the test set of thyroid gland segmentation indices on localized CT images for radiotherapy. MSFA-U-Net had the best values for four evaluation indices: DSC, 0.90; JSC, 0.82; SE, 0.90; and HD, 2.39. Compared with the other three mainstream medical-image segmentation models, MSFA-U-Net greatly improved DSC (improvement range, 0.04–0.06), JSC (improvement range, 0.04–0.08), SE (improvement range, 0.04–0.11), and HD (improvement range, -0.21 to -0.06). On the PPV index, MSFA-U-Net was better than U-Net and worse than both HRNet and Attention U-Net; however, the HRNet and Attention U-Net models performed worse than the MSFA-U-Net model on the other evaluation indices.

Box Plot of Results

To further evaluate the differences among the four models, we made box plots of the evaluation indices (**Figure 6**). The results showed that MSFA-U-Net had a smaller distance between the upper and lower quartiles than the other three models; it also had fewer outliers, and the outliers it did have were closer to the median. These findings indicated that MSFA-U-Net was better at segmenting the thyroid gland on localized CT images for radiotherapy than the other three models were; moreover, it ensured more-consistent segmentation results.

Summary and Analysis of Results

Attention U-Net (which introduces gated attention) and HRNet did not show obvious advantages in thyroid gland segmentation on localized CT images for radiotherapy compared with U-Net; rather, they performed even worse on some of the evaluation indices. The reason might be that although the gated-attention mechanism can

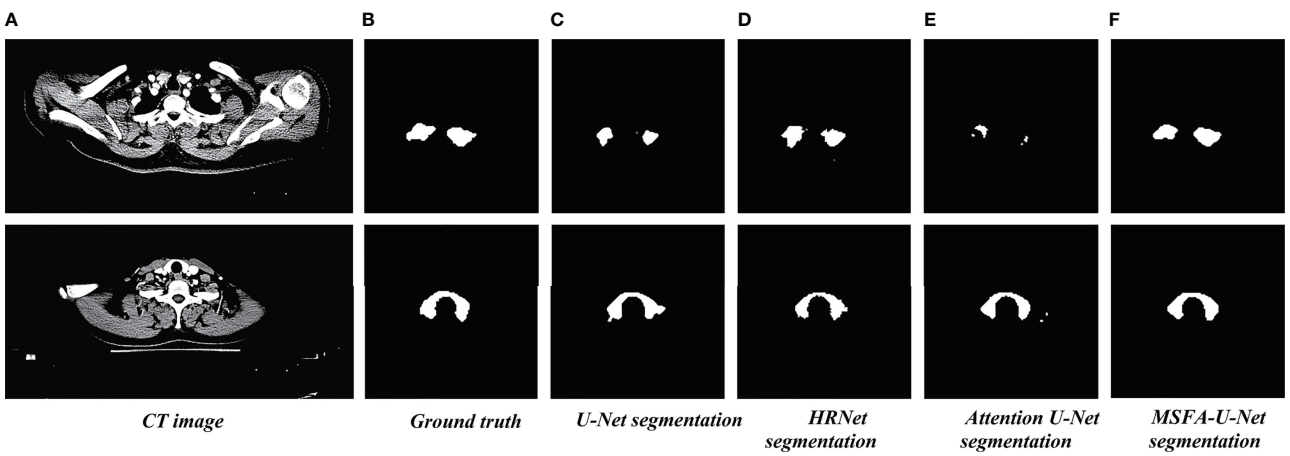


FIGURE 4 | Thyroid gland segmentation of the four models on localized CT images for radiotherapy. **(A)** Standard image of the imported model (CT image). **(B)** Corresponding label image (Ground truth). **(C)** Thyroid segmented by U-Net. **(D)** Thyroid segmented by HRNet. **(E)** Thyroid segmented by Attention U-Net. **(F)** Thyroid segmented by MSFA-U-Net.

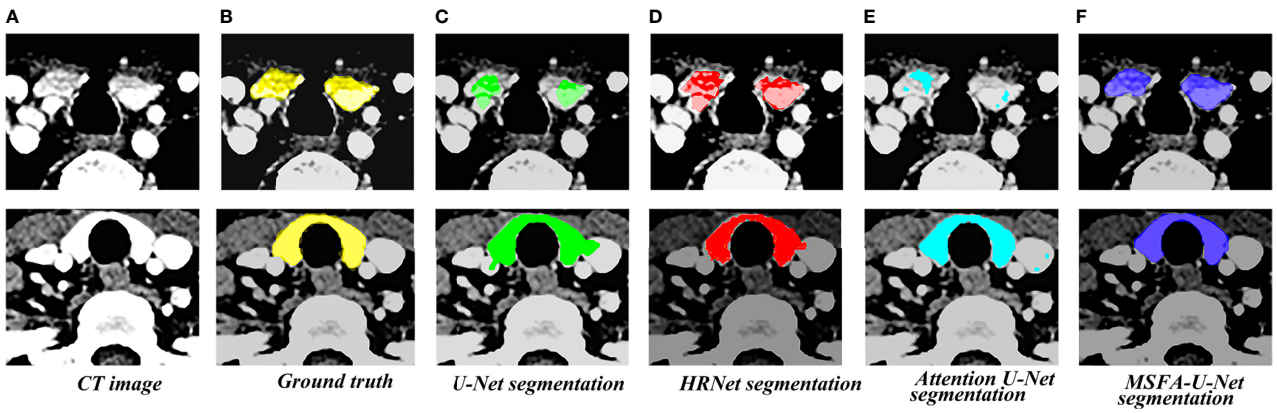


FIGURE 5 | Thyroid coverage map of the four models on localized CT images for radiotherapy. **(A)** CT image. **(B)** Coverage map of thyroid of corresponding label image(Ground truth) on CT image. **(C)** Coverage map of thyroid segmented by U-Net on CT image **(D)** Coverage map of thyroid segmented by HRUet on CT image. **(E)** Coverage map of thyroid segmented by Attention U-Net on CT image. **(F)** Coverage map of thyroid segmented by MSFA-U-Net on CT image.

effectively segment the target category and location, it can also lead to an increase in false-positive model predictions for small-volume segmentation with variability in shape. HRNet performs multiple simple feature fusions; although it can effectively fuse features and obtain rich high-resolution features, it is also more likely to cause overfitting if training data is relatively scarce. Therefore, the phenomenon of integration leads to a decline in HRNet's generalization ability. Clearly, the increases in model parameters and resource consumption might not necessarily improve results.

DISCUSSION

Radiotherapy is critical to comprehensive treatment of head and neck tumors and breast cancer. When radiotherapy plans are designed and implemented, accurate regulation of the radiation dose within the target area and limitation thereof to the surrounding OARs are important factors in the treatment plan's evaluation. OARs must be precisely delineated to effectively limit the dose outside the target area and avoid side effects of radiation (31). The thyroid gland, as an OAR during treatment of head and neck tumors and breast cancer, must be protected during radiotherapy. Atlas-based Auto- contouring (ABAS) and deep learning contouring (DLC) are most widely used in present automatic delineation of OARs. Choi Ms et al. (32) compared the performance between ABAS and DLC in delineating breast cancer OARs and clinical target volume, which showed that DLC performed better than ABAS in the properties of most structures. Besides, L. V. van Dijk et al. (33). compared the performance between ABAS and DLC in automatically delineating head and neck OARs, which revealed that DLC had better performance than ABAS and DLC presented significantly better thyroid automatic delineation with DSC increasing 0.23 (0.83 VS 0.60). However, there still exists improvement space in automatic delineation of thyroid due to its small volume and complex shape, which necessitates developing a new DLC model to improve the accuracy of automatic delineation of the thyroid.

TABLE 2 | Assessment indices of the test set ($\bar{x} \pm s$).

	U-Net	HRNet	Attention U-Net	MSFA-U-Net
DSC	0.86 ± 0.10	0.84 ± 0.09	0.86 ± 0.15	**0.90 ± 0.09**
JSC	0.77 ± 0.13	0.74 ± 0.13	0.78 ± 0.16	**0.82 ± 0.11**
PPV	0.88 ± 0.12	0.93 ± 0.08	**0.95 ± 0.07**	0.91 ± 0.09
SE	0.86 ± 0.12	0.79 ± 0.13	0.81 ± 0.17	**0.90 ± 0.11**
HD	2.60 ± 0.57	2.59 ± 0.54	2.45 ± 0.69	**2.39 ± 0.54**

Bold, optimal value.

This study proposed a multi-scale fusion attention U-Net model to address the problem of thyroid gland segmentation on localized CT images for radiotherapy. The innovation of this algorithm lay in the addition of multiple parallel channels on the basis of the traditional U-Net model. It fully integrated feature information between different resolutions, thereby avoiding single-resolution information in the U-Net downsampling process. In addition, our study also introduced the cSE attention mechanism, which inclined the model to the ROI during the training process. The experimental results showed that, compared with similar representative segmentation algorithms, the proposed model improved both qualitative and quantitative results to a certain extent and had better robustness and generalization. The image segmentation graphs revealed that MSFA-U-Net effectively reduced oversegmentation and undersegmentation and achieved smoother edges. It is believed in the relevant articles that DSC > 0.70 indicates acceptable agreement (34, 35). All models used in this study reached this threshold with the DSC value of MSFA-U-Net reaching 0.90, indicating that this model could effectively segment the thyroid gland on localized CT images for radiotherapy. The box plot diagrams demonstrated that MSFA-U-Net yielded good quantitative results: the upper- and lower-quartile gaps and outliers of most evaluation indicators were reduced, indicating that the model achieved consistent segmentation of the different layers of the thyroid gland and could segment the gland effectively.

Furthermore, some automatic delineation performance of the thyroid conducted by other researchers were compared with ours.

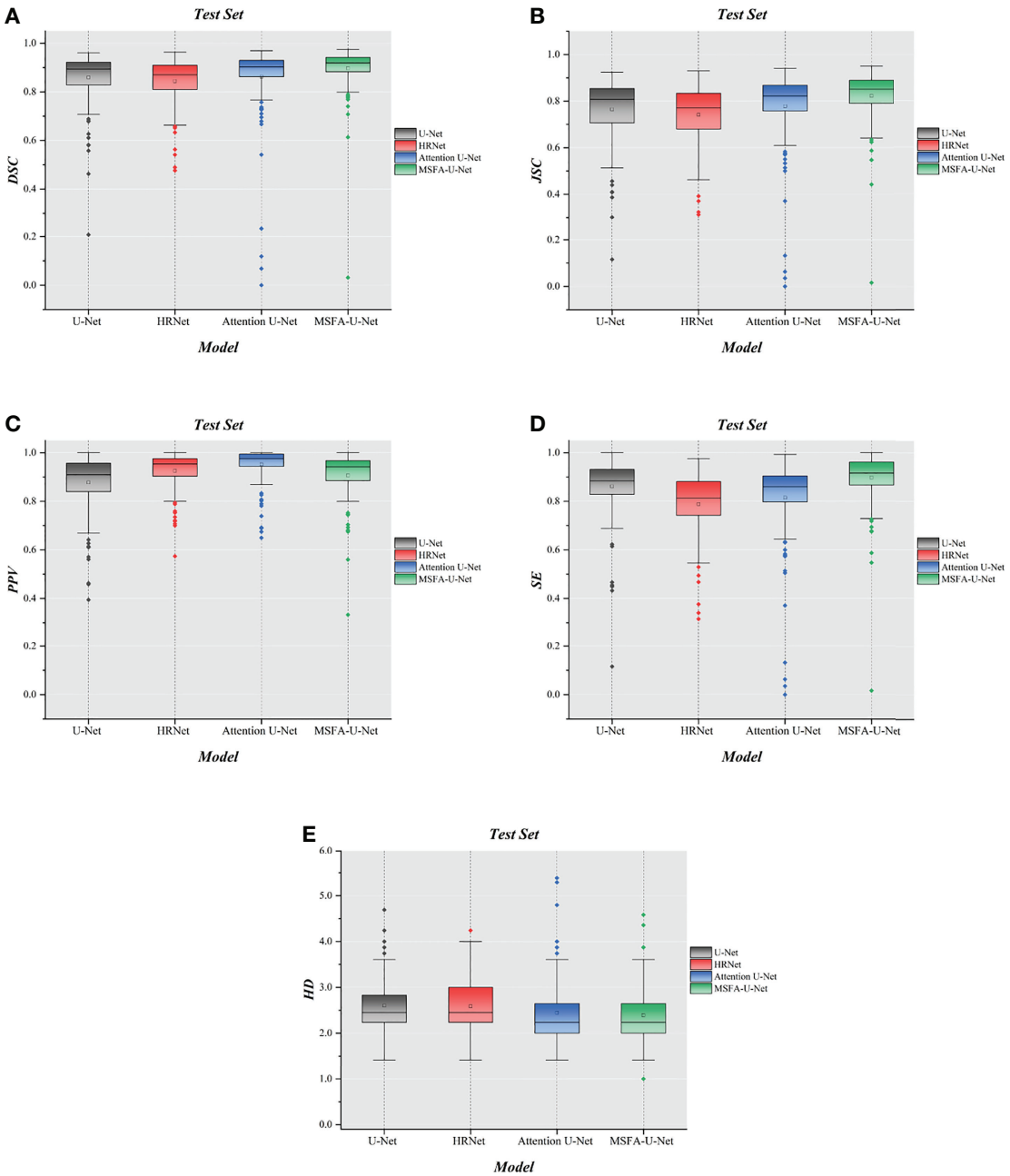


FIGURE 6 | Box plot diagrams in the test set. **(A)** Box plot diagram of DSC in the test set. **(B)** Box plot diagram of JSC in the test set. **(C)** Box plot diagram of PPV in the test set. **(D)** Box plot diagram of SE in the test set. **(E)** Box plot diagram of HD in the test set.

L. V. van Dijk et al. (33) adopted deep learning contouring to improve automatic delineation for head and neck OARs. In their study, automatic delineation of 693 patients were performed with DSC and HD being 0.83 ± 0.08 and 3.6 ± 3.0 mm for DLC. In our study, the DSC and HD for the proposed model, MSFA-U-Net, were 0.90 ± 0.09, and 2.39 ± 0.54 respectively. Yang et al. (36) used a self-adaptive Unet network to segment OARs on the CT images of 149 nasopharyngeal carcinoma patients and obtained thyroid segmentation with DSC being 0.83 ± 0.03 and HD being 4.5 ± 1.3. Compared with their model, our proposed model, MSFA-U-Net, increased DSC by 0.07 and reduced HD by 2.11, indicating that MSFA-U-Net had certain superiority in segmenting the thyroid gland on localized CT images for radiotherapy. Zhong et al. (37) proposed Boosting-based Cascaded Convolutional Neural Network to segment the head and neck OARs. In their model, DSC was above 92.29% and HD was 2.64 for the thyroid. Our proposed model produced a lower DSC, but a better HD evaluation. Cascaded and boosting were adopted in their model, which promoted its

performance than single-models but meanwhile increased its complexity. When we take model complexity and segmentation performance into consideration, we find that our model still has its advantages.

However, the algorithm proposed in this study had some limitations. First, certain MSFA-U-Net evaluation indices have not yet reached optimal results. Second, the introduction of numerous feature fusions among different scales increased the number of model parameters. Third, although Dice loss function could effectively solve the problem of class imbalance, its gradient characteristics could cause the model to oscillate during the training process. In future research, we will explore how to reduce the parameter count of the model while further optimizing the model by using different loss functions.

CONCLUSION

In summary, the MSFA-U-Net model enabled radiotherapy physicians to automatically delineate the thyroid gland on localized CT images for radiotherapy. Our results showed that the model could be applied in clinical work: compared with the three commonly used models in medicine, MSFA-U-Net could delineate the thyroid gland more accurately.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

All authors were involved in the conception of this study, design, and implementation. XW was a major contributor in designing the models, drawing the figures and writing the manuscript and was also involved in delineating some label images. YY performed the checking and proofing of the manuscript and the data apart from guidance to the writing of the manuscript. BZ, MY, JL, MS, LM, and CS were major contributors in data collection, delineation of most of the label images and output of data. All authors contributed to the article and approved the submitted version.

FUNDING

This study has received funding by Climbing Foundation of National Cancer Center (NCC201925B03); Funding unit: National Cancer Center; Funder: YY: Director, Department of Radiotherapy, Yunnan Cancer Hospital.

ACKNOWLEDGMENTS

We thank LetPub (www.letpub.com) for its linguistic assistance during the preparation of this manuscript. The manuscript was preprinted in Research Square (DOI: 10.21203/rs.3.rs-949323/v1; https://www.researchsquare.com/article/rs-949323/v1).

REFERENCES

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global Cancer Statistics 2018: Globocan Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* (2018) 68 (6):394–424. doi: 10.3322/caac.21492
2. Lin Z, Yang Z, He B, Wang D, Gao X, Tam SY, et al. Pattern of Radiation-Induced Thyroid Gland Changes in Nasopharyngeal Carcinoma Patients in 48 Months After Radiotherapy. *PloS One* (2018) 13(7):e0200310. doi: 10.1371/journal.pone.0200310
3. Buchholz TA. Radiation After Breast Conservative Surgery or After Mastectomy Reduces Local Recurrence and Improves Overall Survival in Breast Cancer. *Women's. Oncol Rev* (2007) 6(1-2):67–9. doi: 10.3109/14733400600733387
4. Forastiere AA, Zhang Q, Weber RS, Maor MH, Goepfert H, Pajak TF, et al. Long-Term Results of Rtog 91-11: A Comparison of Three Nonsurgical Treatment Strategies to Preserve the Larynx in Patients With Locally Advanced Larynx Cancer. *J Clin Oncol* (2013) 31(7):845–52. doi: 10.1200/jco.2012.43.6097
5. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: *Med. Image Comput Comput Interv* Munich, Germany: MICCAI (2015). p. 234–41. doi: 10.1007/978-3-319-24574-4_28
6. Hesamian MH, Jia W, He X, Kennedy P. Deep Learning Techniques for Medical Image Segmentation: Achievements and Challenges. *J Dig Imaging* (2019) 32(4):582–96. doi: 10.1007/s10278-019-00227-x
7. Ye Y, Cai Z, Huang B, He Y, Zeng P, Zou G, et al. Fully-Automated Segmentation of Nasopharyngeal Carcinoma on Dual-Sequence Mri Using Convolutional Neural Networks. *Front Oncol* (2020) 10:166. doi: 10.3389/fonc.2020.00166
8. Zhu W, Huang Y, Zeng L, Chen X, Liu Y, Qian Z, et al. Anatomynet: Deep Learning for Fast and Fully Automated Whole-Volume Segmentation of Head and Neck Anatomy. *Med Phys* (2019) 46(2):576–89. doi: 10.1002/mp.13300
9. Alom MZ, Yakopcic C, Hasan M, Taha TM, Asari VK. Recurrent Residual U-Net for Medical Image Segmentation. *J Med Imaging (Bellingham)* (2019) 6 (1):14006. doi: 10.1117/1.Jmi.6.1.014006
10. Zhang Z, Wu C, Coleman S, Kerr D. Dense-Inception U-Net for Medical Image Segmentation. *Comput Methods Prog BioMed* (2020) 192:105395. doi: 10.1016/j.cmpb.2020.105395
11. Chen A, Niermann KJ, Deeley MA, Dawant BM. Evaluation of Multiple-Atlas-Based Strategies for Segmentation of the Thyroid Gland in Head and Neck Ct Images for Imrt. *Phys Med Biol* (2012) 57(1):93–111. doi: 10.1088/0031-9155/57/1/93
12. Zhai RP, Kong FF, Du CR, Hu CS, Ying HM. Radiation-Induced Hypothyroidism After Imrt for Nasopharyngeal Carcinoma: Clinical and Dosimetric Predictors in a Prospective Cohort Study. *Oral Oncol* (2017) 68:44–9. doi: 10.1016/j.oraloncology.2017.03.005
13. Akın M, Ergen A, Unal A, Bese N. Irradiation Doses on Thyroid Gland During the Postoperative Irradiation for Breast Cancer. *J Cancer Res Ther* (2014) 10(4):942–4. doi: 10.4103/0973-1482.137991
14. Boomsma MJ, Bijl HP, Christianen ME, Beetz I, Chouvalova O, Steenbakkers RJ, et al. A Prospective Cohort Study on Radiation-Induced Hypothyroidism: Development of an Ntcp Model. *Int J Radiat Oncol Biol Phys* (2012) 84(3): e351–6. doi: 10.1016/j.ijrobp.2012.05.020
15. Luo R, Li M, Yang Z, Zhan Y, Huang B, Lu J, et al. Nomogram for Radiation-Induced Hypothyroidism Prediction in Nasopharyngeal

Carcinoma After Treatment. *Br J Radiol* (2017) 90(1070):20160686. doi: 10.1259/bjr.20160686
16. Narayanan D, Liu J, Kim L, Chang KW, Lu L, Yao J, et al. Automated Segmentation of the Thyroid Gland on Thoracic Ct Scans by Multiatlas Label Fusion and Random Forest Classification. *J Med Imaging (Bellingham Wash)* (2015) 2(4):44006. doi: 10.1117/1.jmi.2.4.044006
17. Chang CY, Chung PC, Hong YC, Tseng CH. A Neural Network for Thyroid Segmentation and Volume Estimation in Ct Images. *IEEE Comput Intell Mag* (2011) 6(4):43–55. doi: 10.1109/MCI.2011.942756
18. He X, Guo B, Lei Y, Tian S, Wang T, Curran WJ, et al. Thyroid Gland Delineation in Noncontrast-Enhanced Ct Using Deep Convolutional Neural Networks. *Phys Med Biol* (2020) 66(5):055007. doi: 10.1088/1361-6560/abc5a6
19. Liu T, Wang Y, Yang Y, Sun M, Fan W, Bunger C, et al. A Multi-Scale Keypoint Estimation Network With Self-Supervision for Spinal Curvature Assessment of Idiopathic Scoliosis From the Imperfect Dataset. *Artif Intell Med* (2022) 125:102235. doi: 10.1016/j.artmed.2021.102235
20. Wang J, Sun K, Cheng T, Jiang B, Deng C, Zhao Y, et al. Deep High-Resolution Representation Learning for Visual Recognition. *IEEE Trans Pattern Anal Mach Intell* (2021) 43(10):3349–64. doi: 10.1109/tpami.2020.2983686
21. Hu J, Shen L, Albanie S, Sun G, Wu E. Squeeze-and-Excitation Networks. *IEEE Trans Pattern Anal Mach Intell* (2020) 42(8):2011–23. doi: 10.1109/tpami.2019.2913372
22. Ioffe S, Szegedy C. *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*. Lille: JMLR.org (2015). p. 1502.03167.
23. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Ieee*. Las Vegas: IEEE (2016). p. 770–8.
24. Milletari F, Navab N, Ahmadi SA. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. In: *2016 Fourth International Conference on 3D Vision (3dv)*. USA: IEEE (2016). p. 565–71.
25. Hermsen M, de Bel T, den Boer M, Steenbergen EJ, Kers J, Florquin S, et al. Deep Learning-Based Histopathologic Assessment of Kidney Tissue. *J Am Soc Nephrol* (2019) 30(10):1968–79. doi: 10.1681/asn.2019020144
26. Eelbode T, Bertels J, Berman M, Vandermeulen D, Maes F, Bisschops R, et al. Optimization for Medical Image Segmentation: Theory and Practice When Evaluating With Dice Score or Jaccard Index. *IEEE Trans Med Imaging* (2020) 39(11):3679–90. doi: 10.1109/tmi.2020.3002417
27. Talluri SK. Positive Predictive Value. *BMJ* (2009) 339:b3835. doi: 10.1136/bmj.b3835
28. Skaik YA. Understanding and Using Sensitivity, Specificity and Predictive Values. *Indian J Ophthalmol* (2008) 56(4):341. doi: 10.4103/0301-4738.41424
29. van Kreveld M, Miltzow T, Ophelders T, Sonke W, Vermeulen JL. Between Shapes, Using the Hausdorff Distance. *Comput Geomet* (2022) 100:101817. doi: 10.1016/j.comgeo.2021.101817
30. Oktay O, Schlemper J, Folgoc LL, Lee M, Heinrich M, Misawa K, et al. *Attention U-Net: Learning Where to Look for the Pancreas*, Vol. . 1804. (2018). p. 03999.
31. Peng YL, Chen L, Shen GZ, Li YN, Yao JJ, Xiao WW, et al. Interobserver Variations in the Delineation of Target Volumes and Organs at Risk and Their Impact on Dose Distribution in Intensity-Modulated Radiation Therapy for Nasopharyngeal Carcinoma. *Oral Oncol* (2018) 82:1–7. doi: 10.1016/j.oraloncology.2018.04.025
32. 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. *Radioth Oncol* (2020) 153:139–45. doi: 10.1016/j.radonc.2020.09.045
33. van Dijk LV, Van den Bosch L, Aljabar P, Peressutti D, Both S, Steenbakkers RJHM, et al. Improving Automatic Delineation for Head and Neck Organs at Risk by Deep Learning Contouring. *Radioth Oncol* (2020) 142:115–23. doi: 10.1016/j.radonc.2019.09.022
34. 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
35. Vinod SK, Min M, Jameson MG, Holloway LC. A review of interventions to reduce inter-observer variability in volume delineation in radiation oncology. *J Med Imaging Radiat Oncol* (2016) 60:393–406. doi: 10.1111/1754-9485.12462
36. Yang X, Li X, Zhang X, Song F, Huang S, Xia Y. [Segmentation of Organs at Risk in Nasopharyngeal Cancer for Radiotherapy Using a Self-Adaptive Unet Network]. *Nan Fang Yi Ke Da Xue Xue Bao* (2020) 40(11):1579–86. doi: 10.12122/j.issn.1673-4254.2020.11.07
37. Zhong T, Huang X, Tang F, Liang S, Deng X, Zhang Y. Boosting-Based Cascaded Convolutional Neural Networks for the Segmentation of Ct Organs-at-Risk in Nasopharyngeal Carcinoma. *Med Phys* (2019) 46(12):5602–11. doi: 10.1002/mp.13825

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

frontiers | Frontiers in Oncology

ORIGINAL RESEARCH
published: 10 June 2022
doi: 10.3389/fonc.2022.726896

Impact of Interfractional Error on Dosiomic Features

Chanon Puttanawarut[1,2], *Nat Sirirutbunkajorn*[3], *Narisara Tawong*[3], *Suphalak Khachonkham*[3], *Poompis Pattaranutaporn*[3] *and Yodchanan Wongsawat*[2]*

[1] Chakri Naruebodindra Medical Institute, Ramathibodi Hospital, Mahidol University, Samutprakarn, Thailand, [2] Brain-Computer Interface Laboratory, Department of Biomedical Engineering, Faculty of Engineering, Mahidol University, Nakhorn Pathom, Thailand, [3] Department of Diagnostic and Therapeutic Radiology, Ramathibodi Hospital, Mahidol University, Bangkok, Thailand

Objectives: The purpose of this study was to investigate the stability of dosiomic features under random interfractional error. We investigated the differences in the values of features with different fractions and the error in the values of dosiomic features under interfractional error.

Material and Methods: The isocenters of the treatment plans of 15 lung cancer patients were translated by a maximum of ±3 mm in each axis with a mean of (0, 0, 0) and a standard deviation of (1.2, 1.2, 1.2) mm in the x, y, and z directions for each fraction. A total of 81 dose distributions for each patient were then calculated considering four fraction number groups (2, 10, 20, and 30). A total of 93 dosiomic features were extracted from each dose distribution in four different regions of interest (ROIs): gross tumor volume (GTV), planning target volume (PTV), heart, and both lungs. The stability of dosiomic features was analyzed for each fraction number group by the coefficient of variation (CV) and intraclass correlation coefficient (ICC). The agreements in the means of dosiomic features among the four fraction number groups were tested by ICC. The percent differences (PD) between the dosiomic features extracted from the original dose distribution and the dosiomic features extracted from the dose distribution with interfractional error were calculated.

Results: Eleven out of 93 dosiomic features demonstrated a large CV (CV ≥ 20%). Overall CV values were highest in GTV ROIs and lowest in lung ROIs. The stability of dosiomic features decreased as the total number of fractions decreased. The ICC results showed that five out of 93 dosiomic features had an ICC lower than 0.75, which indicates intermediate or poor stability under interfractional error. The mean dosiomic feature values were shown to be consistent with different numbers of fractions (ICC ≥ 0.9). Some of the dosiomic features had PD greater than 50% and showed different PD values with different numbers of fractions.

Conclusion: Some dosiomic features have low stability under interfractional error. The stability and values of the dosiomic features were affected by the total number of fractions. The effect of interfractional error on dosiomic features should be considered in further studies regarding dosiomics for reproducible results.

Keywords: dosiomics, radiomics, texture analysis, dose distribution, stability, generalizability, interfractional error

OPEN ACCESS

Edited by:
Jose Eduardo Villarreal Barajas, Royal Devon and Exeter Hospital, United Kingdom

Reviewed by:
Sunyoung Jang, Princeton Radiation Oncology Center, United States
Niclas Pettersson, University of Gothenburg, Sweden
Annarita Fanizzi, National Cancer Institute Foundation (IRCCS), Italy

***Correspondence:**
Yodchanan Wongsawat
yodchanan.won@mahidol.ac.th

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

Received: *17 June 2021*
Accepted: *18 May 2022*
Published: *10 June 2022*

Citation:
Puttanawarut C, Sirirutbunkajorn N, Tawong N, Khachonkham S, Pattaranutaporn P and Wongsawat Y (2022) Impact of Interfractional Error on Dosiomic Features. Front. Oncol. 12:726896. doi: 10.3389/fonc.2022.726896

INTRODUCTION

In radiation therapy, radiation dose information is analyzed to determine an appropriate radiation plan. Dosimetric values derived from organ-at-risk dose-volume histograms (DVHs) or dosimetric features, such as the mean dose or V20, are commonly used to estimate the normal tissue complication probability (NTCP). However, dosimetric features do not incorporate spatial information from the dose distribution. Therefore, texture analysis of dose distributions, called dosiomics, has been proposed (1–3). Studies have shown that dosiomics can be used to predict complications from radiation therapy more accurately than dosimetric features (1, 3), yet there exist some concerns regarding the stability and generalizability of texture analysis (4, 5). Some dosiomic features were found to be unstable over different grid resolutions (6), dose calculation algorithms (6), and cube pixel spacing (7). This shows that the reproducibility of dosiomic features depends on the process of producing images.

Geometric errors in radiotherapy can be from random and systematic errors. Systematic geometrical error results in a total shift of the dose distribution, while random geometrical error, defined as interfractional error in this study, leads to blurring of the dose distribution (8). During treatment delivery, many random errors, such as setup error, organ shift, and respiratory motion can result in dose deviation from the original plan. Therefore, the actual dose distribution the patient receives could differ from the original treatment plan. These errors can also result in variations in the dose distribution within the same patient and the same treatment plan. The impact of setup errors in dosimetric features has been reported (9–12). Furthermore, these errors can result in overestimation or underestimation of probability according to the NTCP model (13–15). In other words, these errors might decrease the reproducibility of dosiomics as well.

For many cancers, the radiation dose and the number of fractions can vary from patient to patient. Because of interfractional error, a different total number of fractions may induce different error behavior. Random errors in the interfractional dose distribution cause the dose distribution to appear more blurred for a larger total number of fractions than for a smaller total number of fractions. These errors may also further affect the reproducibility of dosiomic features. To study the effect of such errors on dosiomic features, we investigated the following:

- The stability of dosiomic features extracted from dose distributions with interfractional error.
- The differences in dosiomic features extracted from original dose distributions and dose distributions with interfractional error.

MATERIALS AND METHODS

Datasets

Original CT image data, original dose distributions, treatment plan data, and regions of interest (ROIs) for 15 lung cancer patients from the Varian Eclipse Treatment Planning System (TPS; version 16.1, Varian Medical Systems, Palo Alto, CA, USA) at Ramathibodi Hospital were used in this study. Four different ROIs, including the GTV, PTV, heart, and both lungs, were labeled by radiation oncologists. All of the treatment plans corresponded to IMRT/VMAT for lung cancer. The information for clinical factors, radiation setting, and some dose profiles of all patients are included in **Supplementary Tables S1-S3**.

Error Simulation

We simulated the interfractional error by introducing Gaussian error into each fraction. Gaussian error was selected because the sum of arbitrary errors can be approximated as Gaussian error by the central limit theorem (16). The interfractional errors were simulated in Varian Eclipse Treatment Planning System by shifting the isocenter of the original treatment plans n times, where n is the number of fractions. The mean error was (0, 0, 0), and the standard deviation was (1.2, 1.2, 1.2) mm for the Gaussian distribution in the x, y, and z directions (with a maximum shift of ± 3.0 mm in the x, y, and z directions). The range of error -3.0 mm to 3.0 mm is used because in our institution, error < 3 mm is mostly acceptable and still within the PTV margin for IMRT/VMAT treatment. Therefore, for errors > 3 mm, a correction was applied, and the effect of error in that scenario did not exceed 3 mm.

The shifted treatment plans were used to calculate the dose distribution and the accumulative n dose to yield the final dose distribution (D_{err}). The dose distribution of the plan without errors (D_{ori}) was also recalculated to ensure that the parameters of the dose calculation were the same. To generate more dose distributions, we randomly sampled error from a Gaussian distribution to simulate 20 samples of D_{err} for each patient (with the same interfractional error for the same total number of fractions for different patients) and varied the total number of fractions among 2, 10, 20, and 30 while keeping the same total dose.

An example of a simulated dose distribution for patient 9 is shown in **Figure 1**. The overall process from error simulation to feature extraction is shown in **Figure 2**. The dose calculations for D_{ori} and D_{err} were performed using the analytical anisotropic algorithm (AAA) on Varian Eclipse™ TPS. The dose calculation grid resolution was set to 2.5 cm^3. A total of 81 dose distributions (20 images for each fraction group and one nonshifted dose distribution) were calculated for each patient.

Features

Dosiomic features were extracted from the dose distributions by using the pyradiomics library, an open-source Python package for extracting radiomic features from medical images (17). The features considered in this study were in compliance with the feature definitions described by the Imaging Biomarker Standardization Initiative (IBSI) (18). Before calculating a feature, dose distribution data were binned into 70 discrete levels from 0-70 Gy. Dosiomic feature extraction was

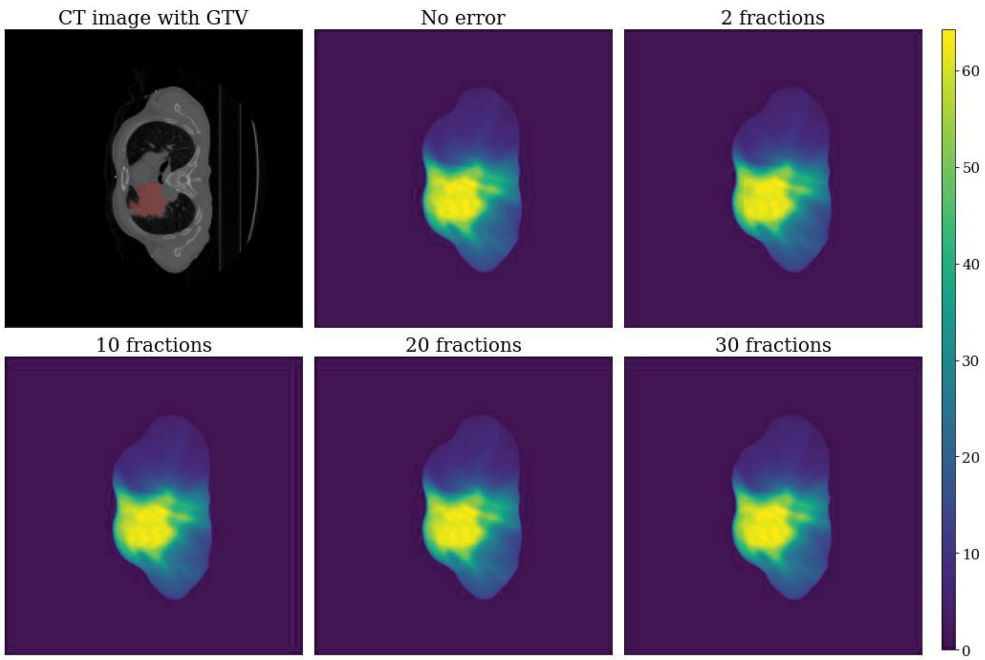


FIGURE 1 | Example dose distributions for patient 9. Top left: CT image with GTV label. Top center: Dose distribution without error (D_{ori}). Top right: Dose distribution with error for a total of two fractions. Bottom left: Dose distribution with error for a total of 10 fractions. Bottom center: Dose distribution with error for a total of 20 fractions. Bottom right: Dose distribution with error for a total of 30 fractions.

performed within the ROIs of the GTV, PTV, heart, and both lungs (bilateral lungs subtracted by GTV), as labeled by radiation oncologists. Features were extracted from D_{err} to obtain sets of dosiomic features for the total number of fractions corresponding to 2, 10, 20, and 30, denoted as $F2_{err}$, $F10_{err}$, $F20_{err,}$ and $F30_{err}$, respectively, as shown in **Figure 3**. We also extracted features from D_{ori} to obtain F_{ori}. We denoted FX_{err} as dosiomic features calculated from DX_{err} for any X total number of fractions. A total of 93 dosiomic features were calculated in this study: first-order features (18 features), gray-level cooccurrence matrix (GLCM) (24 features), gray-level run-length matrix (GLRLM) (16 features), gray-level size zone matrix (GLSZM) (16 features), neighboring gray-tone difference matrix (NGTDM) (five features), and gray-level dependence matrix (GLDM) (14 features). For a list of all features included in the study, we refer to **Supplementary Table S4**.

Data Analysis

Stability analysis of dosiomic features under interfractional errors was performed by the coefficient of variation (CV) and intraclass correlation coefficient (ICC) for each feature $f_{err} \in F_{err}$

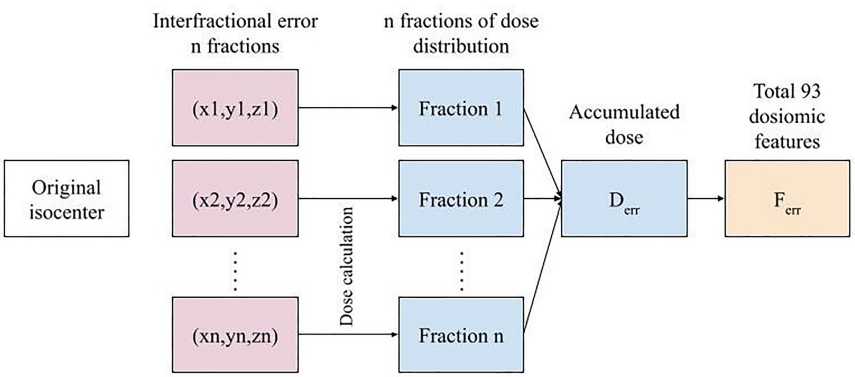


FIGURE 2 | The overall error simulation procedure for n fractions to feature extraction for a single treatment plan.

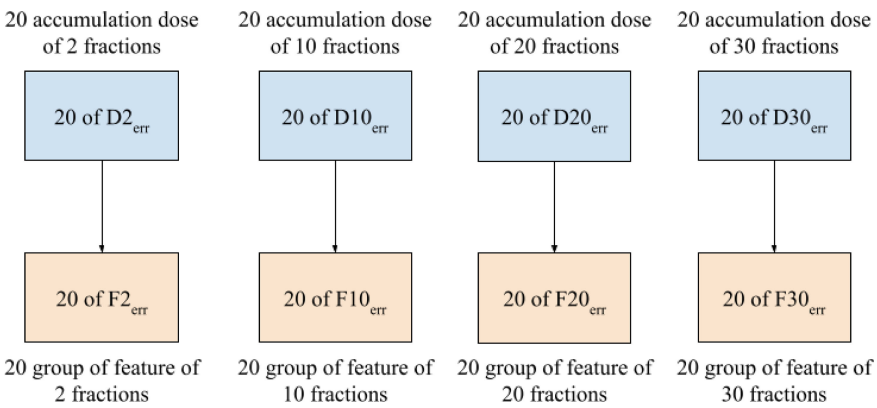


FIGURE 3 | Total of 80 groups of features extracted from D_{err} for each patient.

over the 20 accumulated dose distributions for the different total numbers of fractions, as follows:

$$CV = \left| \frac{\sigma_f}{f_{avg}} \right| \times 100\,\%$$

for f_{avg} is the mean, and σ_f is the standard deviation of the dosiomic features f_{err}. f_{avg} and σ_f were calculated by $f_{avg} = \frac{\sum_{i=1}^{20} f_{err,i}}{20}$ and $\sigma_f = \frac{\sqrt{\sum_{i=1}^{20}(f_{err,i} - f_{avg})^2}}{19}$, respectively. CV has been used in previous studies for radiomic feature stability analysis (19–22). The CV was calculated for all features and all total numbers of fraction groups. The CVs of all features were categorized into four groups as follows: very small CV (CV < 5%), small CV (5% ≤ CV < 10%), intermediate CV (10% ≤ CV < 20%), and large CV (CV ≥ 20%). The CV results are reported as the average CV among all patients.

Another stability test under interfractional errors was performed by ICC (6, 23, 24) of a single-measurement, absolute-agreement, two-way random effect model. ICC is a statistical measure of agreement between different raters; in this case, different interfractional errors give different dosiomic values to the subjects (25). A threshold of ICC ≥ 0.9 indicated good stability under interfractional error (26, 27).

The differences among the values of features extracted from the original dose distribution (F_{ori}) and the features extracted from the dose distribution with interfractional error (F_{err}) were calculated according to the percent difference (PD) over all patients. PD was calculated by

$$PD = \sum_{i=1}^{20} \frac{f_{err,i} - f_{ori}}{f_{ori}} \times \frac{1}{20} \times 100\,\%$$

where $f_{err} \in F_{er}$ and $f_{ori} \in F_{ori}$. PDs were calculated for all features and all total numbers of fraction groups. The PD results were reported as the average PD among all patients.

The effect of the total number of fractions on the mean dosiomic features was analyzed for all patients by ICC by comparing the features of each patient that were extracted from four groups corresponding to different total numbers of fractions.

All statistical analyses were performed by using in-house software implemented in the Python programming language. A p-value less than 0.01 indicated that the test was significant.

RESULTS

Stability under interfractional error was analyzed for each group and ROI. A summary of CV is shown in **Figure 4**. The results show that the overall stability of dosiomic features decreased with decreasing total number of fractions. When comparing four different ROIs, GTV and PTV had more features with large CVs than the other groups. Among the five groups, the GLSZM features were less stable under interfractional error than other feature classes (**Supplementary Figure S1**). Most of the features in the "lung" region had CV less than 10%, except "small area low gray-level emphasis" from GLSZM in Groups $F2_{err}$, $F10_{err}$, and $F20_{err}$ (**Supplementary Figure S1**). Dosiomic features with large CVs (CV ≥ 20%) were "Skewness" (2 times) and "minimum" (1 time) from first-order statistics, "ClusterShade" (4 times) and "ClusterProminence" (1 time) from GLCM, "SizeZoneNonUniformity" (3 times), "SizeZoneNonUniformity Normalized" (1 time), "SmallAreaEmphasis" (5 times), "Small AreaHighGrayLevelEmphasis" (7 times), and "SmallAreaLow GrayLevelEmphasis" (7 times) from GLSZM, and "Complexity" (2 times) and "Strength" (2 times) from NGTDM (**Supplementary Figure S1**). There were no features that had a CV greater than 50%. Additional details of CV regarding each of the dosiomic features in all groups are provided in **Supplementary Figure S1**.

From the ICC results, some dosiomic features were found to have an ICC < 0.9 (p-value < 0.01) (**Table 1**). From across four different ROIs, GTV and PTV had the highest number of features, with ICC < 0.9, and lung was the lowest. The result from ICC was also similar to CV, which showed that there were more features with

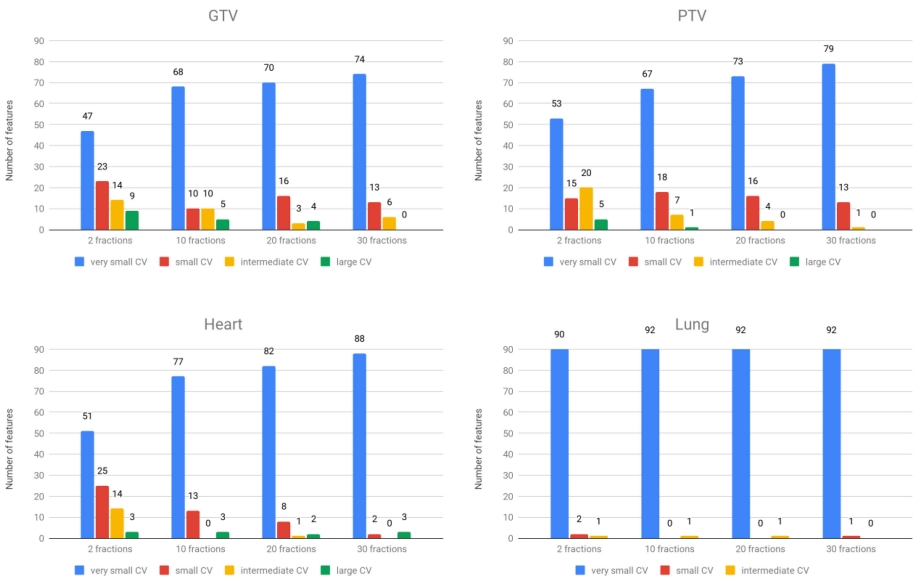


FIGURE 4 | The number of features was categorized into four groups based on CVs. Top left: GTV region. Top right: PTV region. Bottom left: heart region. Bottom right: lung region.

ICC < 0.9 when decreasing the number of fractions. The overall ICC results showed that the GLSZM dosiomic features had a lower ICC than the other groups. The common features in all groups that had ICC < 0.9 were "SmallAreaLowGrayLevelEmphasis" and "SmallAreaHighGrayLevelEmphasis" from GLSZM. The features that had ICC < 0.9 are shown in **Table 1**. The value of 1 - ICC for all features is shown in **Supplementary Figure S3**.

ICCs were tested for consistency of dosiomic features across different numbers of fractions for all ROIs. The expectations of all dosiomic feature values were found to be consistent (ICC ≥ 0.9) with respect to different numbers of fractions (p-value < 0.01). Note that some dosiomic features that had ICC < 0.95 were GTV: "SmallAreaEmphasis" and "SizeZoneNonUniformityNormalized" from GLSZM, PTV: "ClusterProminence" from GLCM, Heart: SmallAreaLowGrayLevelEmphasis from GLSZM and Lung: SmallAreaLowGrayLevelEmphasis from GLSZM.

The results of percent differences (PDs) are shown in **Supplementary Figure S2**. We excluded patient number 14 from calculating the PD of the "Minimum" of the heart and lungs ROI due to the minimum dose being 0. Some of the dosiomic features had F_{ori} that differed from F_{err} by more than 50%. The dosiomic features that had a PD greater than 50% were "ClusterShade" (4 times) from GLCM, "SmallAreaEmphasis" (3 times), "SmallAreaHighGrayLevelEmphasis" (3 times), and "SmallAreaLowGrayLevelEmphasis" (3 times) from GLSZM.

DISCUSSION

Many studies have been performed on the stability of radiomic features, whereas only a few studies have reported on the stability of dosiomic features (6, 7, 22). This is the first report to introduce

TABLE 1 | Features that had ICC < 0.9 are shown for different ROIs and different numbers of fractions.

Number of fractions (n)	(0.75 ≤ ICC < 0.9)	(0.5 ≤ ICC < 0.75)	(ICC < 0.5)
GTV 2 fractions (8)	**GLSZM**	**GLSZM**	
	Gray Level Non Uniformity Normalized	Size Zone Non Uniformity Normalized	
	NGTDM	Small Area Emphasis	
	Contrast	Small Area High Gray Level Emphasis	
	GLDM	Small Area Low Gray Level Emphasis	
	Dependence Non Uniformity Normalized		
	Dependence Variance		
GTV 10 fractions (4)	**GLSZM**	**GLSZM**	
	Small Area High Gray Level Emphasis	Size Zone Non Uniformity Normalized	
		Small Area Emphasis	
		Small Area Low Gray Level Emphasis	
GTV 20 fractions (4)	**GLSZM**	**GLSZM**	
	Small Area Emphasis	Size Zone Non Uniformity Normalized	

(Continued)

TABLE 1 | Continued

Number of fractions (n)	(0.75 ≤ ICC < 0.9)	(0.5 ≤ ICC < 0.75)	(ICC < 0.5)
	Small Area High Gray Level Emphasis Small Area Low Gray Level Emphasis		
GTV 30 fractions (4)	**GLSZM** Size Zone Non Uniformity Normalized Small Area Emphasis Small Area High Gray Level Emphasis Small Area Low Gray Level Emphasis		
PTV 2 fractions (4)	**GLCM** Cluster Shade **GLSZM** Small Area Emphasis	**GLCM** Cluster Prominence **GLSZM** Small Area High Gray Level Emphasis	
PTV 10 fractions (1)	**GLCM** Small Area High Gray Level Emphasis		
PTV 20 fractions (1)	**GLSZM** Small Area High Gray Level Emphasis		
PTV 30 fractions (1)	**GLSZM** Small Area High Gray Level Emphasis		
Heart 2 fractions (2)	**GLSZM** Size Zone Non Uniformity Normalized		**GLSZM** Small Area Low Gray Level Emphasis
Heart 10 fractions (1)			**GLSZM** Small Area Low Gray Level Emphasis
Heart 20 fractions (1)			**GLSZM** Small Area Low Gray Level Emphasis
Heart 30 fractions (1)			**GLSZM** Small Area Low Gray Level Emphasis
Lung 2 fractions (2)	**GLSZM** Low Gray Level Zone Emphasis	**GLSZM** Small Area Low Gray Level Emphasis	
Lung 10 fractions (1)		**GLSZM** Small Area Low Gray Level Emphasis	
Lung 20 fractions (1)	**GLSZM** Small Area Low Gray Level Emphasis		
Lung 30 fractions (1)	**GLSZM** Small Area Low Gray Level Emphasis		

a concern about the stability of dosiomic features under interfractional error in radiation therapy.

Many ML and DL applications with texture analysis in the field of radiotherapy have been recently proposed (1, 3, 28–30). Many studies have extracted features from treatment plan dose distributions and used these features as input to machine learning models; however, such dose distributions might not represent the true dose distribution the patient received due to interfractional error. The effect of interfractional error is known to impact the equivalent uniform dose (EUD), the tumor control probability (TCP), and the normal tissue complication probability (NTCP) (13). Changes in EUD, TCP, and NTCP can affect the decision of the physician during treatment planning. In the same way, errors in dosiomic features can result in prediction errors in ML models. Therefore, this study aimed to investigate the impact of interfractional error on dosiomic features.

We calculated CVs from each dosiomic feature and compared them to assess stability. Most of the features in all total numbers of fractions ($F2_{err}$, $F10_{err}$, $F20_{err}$, and $F30_{err}$) groups and all ROIs had high stability (very small CV and small CV). Comparing four ROIs, the lungs were the regions with high overall stability, while GTVs had low overall stability (**Figure 4**, **Supplementary Figure S1**). The most common dosiomic features with high variation compared to other features were similar to previous studies on dosiomic feature stability. For stability under different dose cube pixel spacings with a CV threshold of 0.3 (CV > 0.3), the stability of "Skewness" from the first-order statistic, "ClusterShade" and "ClusterProminence" from GLCM, and "SmallAreaLowGrayLevelEmphasis" from GLSZM were similar to our results (7). For the stability under dose grid resolution and dose calculation algorithms with a CV threshold of 0.5 (CV > 0.5), the stability of "Skewness" from the first-order statistic, "ClusterShade" from GLCM, and "SmallAreaLowGray LevelEmphasis" from GLSZM were similar to our results (6). It should be noted that we only investigated the stability of dosiomic in primary lung cancer patients under interfractional error, while the previous studies have explored stability under other factors such as dose cube pixel spacing (7), dose grid resolution, and dose calculation algorithms (6) across several primary cancer such as breast and brain. Further study on investigating and comparing the results obtained with respect to other primary cancers should be listed as our future work.

CV also revealed that the stability of dosiomic features decreased as the total number of fractions decreased (**Figure 4**). The high variation in the low total number of fractions group could be explained by the law of large numbers, causing a lower total number of fractions to have higher variance than a higher total number of fractions.

The ICC showed that some dosiomic features (**Table 1**) did not demonstrate excellent reproducibility defined by an ICC

threshold of ≥ 0.9 (26, 27). A guideline of ICC by Terry K. Koo et al. (25) suggested that ICC values less than 0.5 could be defined as poor reproducibility, which corresponded to the dosiomic feature "SmallAreaLowGrayLevelEmphasis" from GLSZM extracted from the heart ROIs. ICC values of 0.5 ≤ ICC < 0.75 could be defined as moderate reliability, which corresponded to dosiomic features: "SmallAreaLowGrayLevelEmphasis," "SmallAreaHighGrayLevelEmphasis," "SmallAreaEmphasis" "SizeZoneNonUniformityNormalized" from GLSZM and "ClusterProminence" from GLCM (**Table 1**, **Supplementary Figure S3**). ICC values of 0.75 ≤ ICC < 0.90 and ICC ≥ 0.9 indicated good and excellent reproducibility. Using dosiomic features with poor or moderate reproducibility might result in limited generalizability in some ROIs while lowering the number of fractions might lower the reproducibility.

We reviewed the predictive dosiomic features reported in other studies. The results presented here were similar to those (31), which showed that the predictive features of treatment response had high stability (32). "LongRunHighGrayLevelEmphasis" from GLRLM "Contrast" from GLCM and "LowGrayLevelEmphasis" from GLDM were selected as predictive factors in genitourinary and gastrointestinal complications in prostate cancer (1), radiation pneumonitis in lung cancer (3), and locoregional recurrences in head and neck cancer (30), respectively. These selected features had high stability (CV < 10%) and excellent reproducibility (ICC ≥ 0.9) according to the number of delivered fractions.

We investigated the PD between F_{ori} and F_{err}, comparing four ROIs. The lungs were the regions that had the lowest overall PD, and GTVs were the regions with the highest overall PD (**Supplementary Figure S2**). Note that a small or large PD between F_{ori} and F_{err} does not always lead to small or large errors in the predictive performance of a model, as there are many other factors, such as model parameters and techniques, in developing the model that may impact performance. For example, if the dosiomic features with large errors are normalized to small values and the weight of the features is small, then the change in the model result will also be small. The features are generally normalized before being input to the model, and models usually have regularization constraints.

A limitation of this study was that our data only included lung cancer patients. Different cancer types or different ROIs may give different feature values. However, from our results, the value of CVs and ICC in four tested ROIs showed a similar pattern (although the magnitudes were not the same). For example, in the same ROI, some features in GLSZM showed high CVs (**Supplementary Figure S1**) or high 1-ICC (**Supplementary Figure S3**) when compared with other dosiomic features. The difference in magnitude of stability may arise from the total dose in that region. Therefore, we also expected a similar pattern of stability in different cancer sites, with the magnitude of stability depending on the total dose. The dose to the heart ROIs was lower than the dose to the lung ROIs (**Supplementary Table S3**). However, lung ROIs had higher overall stability than heart ROIs. We believe that this came from the interfractional error, which caused the dose to the heart ROIs to be different in each dose distribution (**Supplementary Figure S4**), while the dose to the lungs was similar even with interfractional error.

In summary, this study investigated the stability of dosiomic features with IMRT and VMAT in a lung cancer dataset. Our results showed that some dosiomic features might not be reliable under interfractional error and with lower fraction numbers, even more susceptible to the effects of interfractional error resulting in unstable features. Our study also investigated plans with a higher dose per fraction and lower number of fractions than usual of IMRT/VMAT plans. Therefore, we expected that the treatment plan using the stereotactic body radiotherapy (SBRT) technique would yield similar results as the number of fractions 2 and 10. The stability of texture features should be further investigated using SBRT datasets.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available. Requests to access the datasets should be directed to yodchanan.won@mahidol.ac.th.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Human research ethics committee, Faculty of Medicine Ramathibodi Hospital, Mahidol University (IRB MURA2021/283).

AUTHOR CONTRIBUTIONS

CP, NS, and NT contributed to the acquisition of the data. CP, PP, NS, and SK contributed to the experimental design and statistical analysis. CP contributed to processing the data and feature extraction. CP and NS drafted the manuscript. SK, PP, and YW were senior authors supervising the project. All authors read and approved the final manuscript.

FUNDING

This project is supported in part by the National Higher Education Science Research and Innovation Policy Council, PMU B (B05F640079).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.726896/full#supplementary-material

REFERENCES

1. 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
2. Gabryś HS, Buettner F, Sterzing F, Hauswald H, Bangert M. Design and Selection of Machine Learning Methods Using Radiomics and Dosiomics for Normal Tissue Complication Probability Modeling of Xerostomia. *Front Oncol* (2018) 8(MAR):1–20. doi: 10.3389/fonc.2018.00035
3. 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(APR):1–7. doi: 10.3389/fonc.2019.00269
4. Yip SSF, Aerts HJWL. Applications and Limitations of Radiomics. *Phys Med Biol* (2016) 61(13):R150–66. doi: 10.1088/0031-9155/61/13/R150
5. Park JE, Park SY, Kim HJ, Kim HS. Reproducibility and Generalizability in Radiomics Modeling: Possible Strategies in Radiologic and Statistical Perspectives. *Kor J Radiol* (2019) 20(7):1124–37. doi: 10.3348/kjr.2018.0070
6. Placidi L, Lenkowicz J, Cusumano D, Boldrini L, Dinapoli N, Valentini V. Stability of Dosiomics Features Extraction on Grid Resolution and Algorithm for Radiotherapy Dose Calculation. *Phys Med* (2020) 77:30–5. doi: 10.1016/j.ejmp.2020.07.022
7. 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
8. van Herk M, Remeijer P, Rasch C, Lebesque JV. The Probability of Correct Target Dosage: Dose-Population Histograms for Deriving Treatment Margins in Radiotherapy. *Int J Radiat Oncol* (2000) 47(4):1121–35. doi: 10.1016/S0360-3016(00)00518-6
9. Algan O, Jamgade A, Ali I, Christie A, Thompson JS, Thompson D, et al. The Dosimetric Impact of Daily Setup Error on Target Volumes and Surrounding Normal Tissue in the Treatment of Prostate Cancer With Intensity-Modulated Radiation Therapy. *Med Dosim* (2012) 37(4):406–11. doi: 10.1016/j.meddos.2012.03.003
10. Goddu SM, Yaddanapudi S, Pechenaya OL, Chaudhari SR, Klein EE, Khullar D, et al. Dosimetric Consequences of Uncorrected Setup Errors in Helical Tomotherapy Treatments of Breast-Cancer Patients. *Radiother Oncol* (2009) 93(1):64–70. doi: 10.1016/j.radonc.2009.07.013
11. Lee M. A Study for the Dosimetric Evaluation of Rotational Setup Error for Lung Stereotactic Body Radiation Therapy. *J Radiother Pract* (2018) 17 (4):422–6. doi: 10.1017/S1460396918000250
12. Fu W, Yang Y, Li X, Heron DE, Huq MS, Yue NJ. Dosimetric Effects of Patient Rotational Setup Errors on Prostate IMRT Treatments. *Phys Med Biol* (2006) 51(20):5321–31. doi: 10.1088/0031-9155/51/20/016
13. Boughalia A, Marcie S, Fellah M, Chami S, Mekki F. Assessment and Quantification of Patient Set-Up Errors in Nasopharyngeal Cancer Patients and Their Biological and Dosimetric Impact in Terms of Generalized Equivalent Uniform Dose (gEUD), Tumour Control Probability (TCP) and Normal Tissue Complication Pro. *Br J Radiol* (1050) 2015:88. doi: 10.1259/bjr.20140839
14. Chamunyonga C. The Impact of Inter-Fraction Set-Up Errors on the Probability of Pulmonary and Cardiac Complication in Left-Sided Breast Cancer Patients. *J Radiother Pract* (2013) 13(4):393–402. doi: 10.1017/S1460396914000065
15. McCulloch MM, Muenz DG, Schipper MJ, Velec M, Dawson LA, Brock KK. A Simulation Study to Assess the Potential Impact of Developing Normal Tissue Complication Probability Models With Accumulated Dose. *Adv Radiat Oncol* (2018) 3(4):662–72. doi: 10.1016/j.adro.2018.05.003
16. van Herk M, Witte M, van der Geer J, Schneider C, Lebesque JV. Biologic and Physical Fractionation Effects of Random Geometric Errors. *Int J Radiat Oncol* (2003) 57(5):1460–71. doi: 10.1016/j.ijrobp.2003.08.026
17. Griethuysen JJMV, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res* (2017) 77(21):e104–7. doi: 10.1158/0008-5472.CAN-17-0339
18. Zwanenburg A, Leger S, Vallières M, Löck S. *Image Biomarker Standardisation Initiative* (2016). Available at: http://arxiv.org/abs/1612.07003%0.
19. Yan J, Chu-Shern JL, Loi HY, Khor LK, Sinha AK, Quek ST, et al. Impact of Image Reconstruction Settings on Texture Features in 18F-FDG PET. *J Nucl Med* (2015) 56(11):1667–73. doi: 10.2967/jnumed.115.156927
20. Kim H, Park CM, Lee M, Park SJ, Song YS, Lee JH, et al. Impact of Reconstruction Algorithms on CT Radiomic Features of Pulmonary Tumors: Analysis of Intra- and Inter-Reader Variability and Inter-Reconstruction Algorithm Variability. *PLos One* (2016) 11(10):1–11. doi: 10.1371/journal.pone.0164924
21. Forgacs A, Jonsson HP, Dahlbom M, Daver F, Difranco MD, Opposits G, et al. A Study on the Basic Criteria for Selecting Heterogeneity Parameters of F18-FDG PET Images. *PLos One* (2016) 11(10):1–14. doi: 10.1371/journal.pone.0164113
22. Placidi L, Gioscio E, Garibaldi C, Rancati T, Fanizzi A, Maestri D, et al. A Multicentre Evaluation of Dosiomics Features Reproducibility, Stability and Sensitivity. *Cancers* (2021) 13(15):3835. doi: 10.3390/cancers13153835
23. Zwanenburg A, Leger S, Agolli L, Pilz K, Troost EGC, Richter C, et al. Assessing Robustness of Radiomic Features by Image Perturbation. *Sci Rep* (2019) 9(1):1–10. doi: 10.1038/s41598-018-36938-4
24. Fiset S, Welch ML, Weiss J, Pintilie M, Conway JL, Milosevic M, et al. Repeatability and Reproducibility of MRI-Based Radiomic Features in Cervical Cancer. *Radiother Oncol* (2019) 135:107–14. doi: 10.1016/j.radonc.2019.03.001
25. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med* (2016) 15(2):155–63. doi: 10.1016/j.jcm.2016.02.012
26. van Velden FHP, Kramer GM, Frings V, Nissen IA, Mulder ER, de Langen AJ, et al. Repeatability of Radiomic Features in Non-Small-Cell Lung Cancer [18f] FDG-PET/CT Studies: Impact of Reconstruction and Delineation. *Mol Imaging Biol* (2016) 18(5):788–95. doi: 10.1007/s11307-016-0940-2
27. Bogowicz M, Riesterer O, Bundschuh RA, Veit-Haibach P, Hüllner M, Studer G, et al. Stability of Radiomic Features in CT Perfusion Maps. *Phys Med Biol* (2016) 61(24):8736–49. doi: 10.1088/1361-6560/61/24/8736
28. Ibragimov B, Toesca D, Chang D, Yuan Y, Koong A, Xing L. Development of Deep Neural Network for Individualized Hepatobiliary Toxicity Prediction After Liver SBRT. *Med Phys* (2018) 45(10):4763–74. doi: 10.1002/mp.13122
29. Ibragimov B, Toesca DAS, Chang DT, Yuan Y, Koong AC, Xing L, et al. Deep Learning for Identification of Critical Regions Associated With Toxicities After Liver Stereotactic Body Radiation Therapy. *Med Phys* (2020) 47:3721–31 47:3721–31. doi: 10.1002/mp.14235
30. 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
31. Tixier F, Hatt M, Rest CCL, Pogam AL, Corcos L, Visvikis D. Reproducibility of Tumor Uptake Heterogeneity Characterization Through Textural Feature Analysis in 18F-FDG PET. *J Nucl Med* (2012) 53(5):693–700. doi: 10.2967/jnumed.111.099127
32. Tixier F, Le Rest CC, Hatt M, Albarghach N, Pradier O, Metges J-P, et al. Intratumor Heterogeneity Characterized by Textural Features on Baseline ^{18} F-FDG PET Images Predicts Response to Concomitant Radiochemotherapy in Esophageal Cancer. *J Nucl Med* (2011) 52(3):369–78. doi: 10.2967/jnumed.110.082404

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

frontiers | Frontiers in Oncology

ORIGINAL RESEARCH
published: 30 June 2022
doi: 10.3389/fonc.2022.779030

Incremental Value of Radiomics in 5-Year Overall Survival Prediction for Stage II–III Rectal Cancer

Ke Nie[1,2], Peng Hu[1], Jianjun Zheng[3], Yang Zhang[2], Pengfei Yang[4], Salma K. Jabbour[2], Ning Yue[2], Xue Dong[1], Shufeng Xu[1], Bo Shen[1], Tianye Niu[4], Xiaotong Hu[5], Xiujun Cai[6]* and Jihong Sun[1,7]**

[1] Department of Radiology, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou, China, [2] Department of Radiation Oncology, Rutgers-Cancer Institute of New Jersey, Rutgers-Robert Wood Johnson Medical School, New Brunswick, NJ, United States, [3] Department of Radiology, Hwa Mei Hospital, Key Laboratory of Diagnosis and Treatment of Digestive System Tumors of Zhejiang Province, University of Chinese Academy of Sciences, Ningbo, China, [4] Institute of Translational Medicine, Zhejiang University, Hangzhou, China, [5] Biomedical Research Center and Key Laboratory of Biotherapy of Zhejiang Province, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou, China, [6] Department of General Surgery, Innovation Center for Minimally Invasive Techniques and Devices, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou, China, [7] Innovation Center for Minimally Invasive Techniques and Devices, Zhejiang University, Hangzhou, China

OPEN ACCESS

Edited by:
Jose Eduardo Villarreal Barajas, Royal Devon and Exeter Hospital, United Kingdom

Reviewed by:
Kevin Martell, University of Calgary, Canada
Satish E. Viswanath, Case Western Reserve University, United States

***Correspondence:**
Jihong Sun
sunjihong@zju.edu.cn
Xiujun Cai
srrsh_cxj@zju.edu.cn
Xiaotong Hu
hxt_hz@zju.edu.cn

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

Received: *17 September 2021*
Accepted: *30 May 2022*
Published: *30 June 2022*

Citation:
Nie K, Hu P, Zheng J, Zhang Y, Yang P, Jabbour SK, Yue N, Dong X, Xu S, Shen B, Niu T, Hu X, Cai X and Sun J (2022) Incremental Value of Radiomics in 5-Year Overall Survival Prediction for Stage II–III Rectal Cancer. Front. Oncol. 12:779030. doi: 10.3389/fonc.2022.779030

Although rectal cancer comprises up to one-third of colorectal cancer cases and several prognosis nomograms have been established for colon cancer, statistical tools for predicting long-term survival in rectal cancer are lacking. In addition, previous prognostic studies did not include much imaging findings, qualitatively or quantitatively. Therefore, we include multiparametric MRI information from both radiologists' readings and quantitative radiomics signatures to construct a prognostic model that allows 5-year overall survival (OS) prediction for advance-staged rectal cancer patients. The result suggested that the model combined with quantitative imaging findings might outperform that of conventional TNM staging or other clinical prognostic factors. It was noteworthy that the identified radiomics signature consisted of three from dynamic contrast-enhanced (DCE)-MRI, four from anatomical MRI, and one from functional diffusion-weighted imaging (DWI). This highlighted the importance of multiparametric MRI to address the issue of long-term survival estimation in rectal cancer. Additionally, the constructed radiomics signature demonstrated value to the conventional prognostic factors in predicting 5-year OS for stage II–III rectal cancer. The presented nomogram also provides a practical example of individualized prognosis estimation and may potentially impact treatment strategies.

Keywords: 5-year overall survival, advanced rectal cancer, prognostic model, radiomics, chemoradiation

INTRODUCTION

The mortality rate of colorectal cancer has declined steadily over the last 2 decades, with Surveillance, Epidemiology, and End Results Program (SEER) data statistics showing 23.6% observed death rate in 1992 to 12.8% death rate in 2019 (https://seer.cancer.gov/statfacts/html/colorect.html). This is widely believed to be a consequence of improvements in surgical, medical,

and supportive care (1). While substantial progress has been made, heterogeneity in survival outcomes exists (2–5). An accurate prognostication would be helpful to inform treatment decisions, determine clinical trial eligibility, and develop surveillance schedules. Tumor node metastasis (TNM) staging plays a vital role in predicting prognosis and facilitating treatment stratification, yet it is not sufficiently precise (6, 7). In the current TNM staging system, inclusion of tumor deposits (TDs) within nodal staging has given rise to worldwide discussions (8–11). While other important prognosis features, such as pretreatment serum level of carcinoembryonic antigen (CEA), extramural vascular invasion (EMVI), and circumferential resection margin (CRM), are acknowledged, they are not included in staging due to lack of standardized agreement or recommendations. Thus, a more precise survival estimation tailored to the individual patient is needed.

For colon cancer, there were established nomograms to predict recurrence or survival beyond the current TNM staging system with additional prognostic factors, both continuous and discrete, as well as nonlinear and complex mathematical relationships (12–14). In contrast, personalized prognostication of rectal cancer remains lacking despite more heterogeneous results. Valentini et al. (6) reported a nomogram incorporating clinical variables such as age, gender, TNM staging, chemoradiation, and surgical procedure to predict overall survival (OS). Similarly, van Gijn et al. (5) developed a nomogram to predict survival in patients treated with optional short-term radiotherapy by evaluating similar clinical variables and pathological TNM staging. Lately, Song et al. (15) extended the work with pretreatment/posttreatment CEA levels, Cancer Antigen (CA) 19-9 values, and combined clinical and pathological characteristics to predict OS of resected rectal cancer. However, these studies did not account for image findings, qualitatively or quantitatively.

Current improvements in survival estimation have largely been made due to advances in biologic and genomic technologies (16–20). However, the inability to obtain comprehensive information with regard to spatial and temporal heterogeneity continues to be a limitation in optimizing treatment strategy (21). Radiomics has been brought into the evolving topic, as it enables the noninvasive profiling of the disease (22, 23). Recent work in radiomics has provided insights in personalized medicine related to tumor detection, subtype classification, and therapeutic response assessment in rectal cancer (24, 25). Since imaging characteristics may also reveal underlying disease behavior and progression, in this study, we investigated whether a prognostic model that leverages the full complement of routinely available data elements and radiomics information could more accurately predict 5-year OS. We also established a radiomics nomogram to assess its incremental value to the traditional TNM system and clinical, histological, and radiological factors for individual long-term OS estimation in stage II–III rectal cancer.

METHODS

Patients' Characteristics

Institutional Review Board (IRB) approval was obtained for this retrospective analysis, and informed consent was waived. All patients with rectal cancer [American Joint Committee on Cancer (AJCC) 7th edition] were reviewed from the institutional database starting from June 2008 to July 2013. All patients received Total Mesorectal Excision (TME) as standard treatment. Postoperative chemotherapy with either Folfox- or Xelox-based regimen was applied at the physician's discretion. The criteria for inclusion in the study consisted of patients who 1) were diagnosed with stage II–III rectal cancer and no distant metastasis, 2) underwent surgical resection for curative intent, and 3) had a minimal of 5-year follow-up data. Patients with one or more of the following criteria were excluded from the study: history of inflammatory bowel disease, pelvic surgery, intestinal polyposis syndromes or colon cancer, low-quality MRI data from motion artifacts or poor contrast injection, concurrent malignancies, and missing clinical/pathological information. The patient selection flowchart is shown in **Figure 1**.

The recorded clinical and treatment data included age, gender, pretreatment CEA level, clinical tumor (cT) and node (cN) stage, postoperative TD status, surgical procedure [low anterior resection (LAR), abdominoperineal resection (APR), Hartmann's procedure, or others], adjuvant chemotherapy (no/Folfox/Xelox). The radiological reading included tumor location from MRI (low, mid, and high) and tumor distance from the anal verge measured from colonoscopy. All images and reports were reviewed by two radiologists independently (with respective 10 and 8 years of experience in abdominal MRI), and the discrepancies were revisited until consensus has been reached.

Evaluation of the surgical resection specimen for residual tumor was performed under a standard reporting protocol in a central pathological laboratory at Sir Run Run Shaw Hospital. Two experienced gastroenterology pathologists independently reviewed the specimen, and a third expert pathologist was responsible for the final decision in case of a disagreement between the two pathologists. Data from pathological readings included tumor (yT) and nodal (yN) stages, total number of examined lymph node (LN), number of positive LN, positive lymph node ratio (LNR), and histology type (I-well, II-mid, or III-poorly differentiated, and IV-mucoid and signet ring cell carcinoma). The EMVI was scored as suggested by Smith et al. (26), with 0 if there was absence of invasion while 4 if with the most overt features of invasion. The final clinical endpoint was the long-term OS, which was defined as the time from the date of surgical resection until death. The minimum follow-up time to ascertain OS was 60 months. In total, 165 cases were identified (66 women; mean age 63 ± 12 years, range 19–89 years) with survival ranging from 5 to 121 months (median of 74 months).

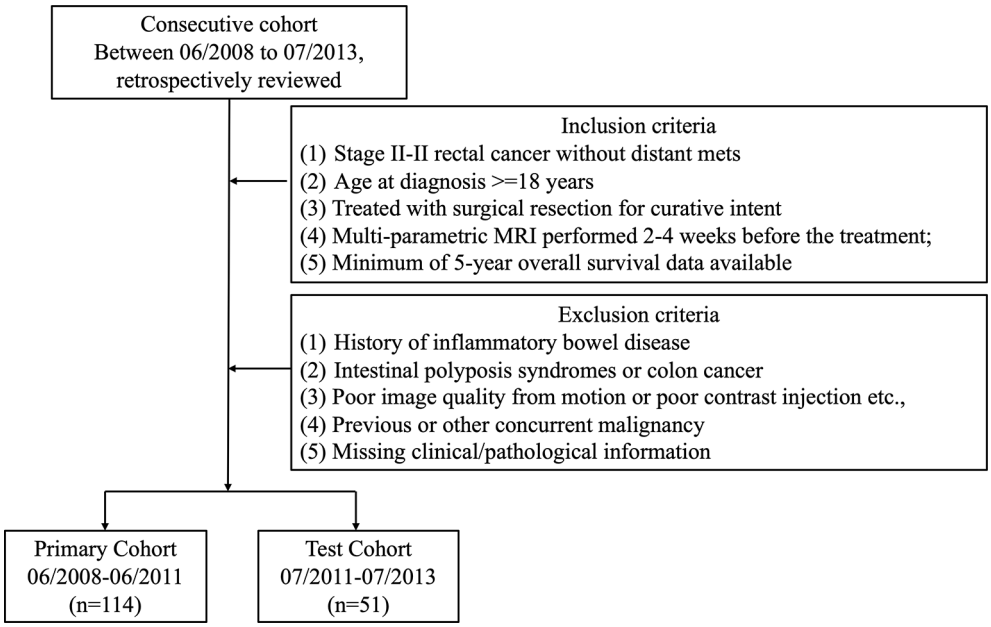


FIGURE 1 | A flowchart showing patient selection.

MRI Data Acquisition

Patients were scanned at the same institution with a 1.5 Tesla MR scanner (Signa Excite HD, GE Medical Systems) using 8-channel phased array body coil in supine position before curative treatment. Images included axial T1-weighted (T1w), high-resolution T2-weighted (T2w), four-phase dynamic contrast-enhanced MRI (DCE-MRI), and diffusion-weighted imaging (DWI) sequences.

The acquisition parameters were as follows: axial T1w (T1w spin echo sequence, repetition time as TR/echo time as TE: 460/7.4 ms) and axial T2w (T2w fast spin echo sequence, TR/TE: 2,840/131 ms, image resolution 0.49 mm × 0.49 mm × 4 mm) maps were acquired. Then, multiphase T1w were obtained using a spoiled gradient echo sequence [liver acceleration volume acquisition (LAVA)]. Scan parameters were TR/TE 4.4/1.9 ms, flip angle 12°, bandwidth 325.5 kHz, image resolution: 0.7 mm × 0.7 mm × 2 mm. All patients were injected with 0.1 mmol/kg body weight Gd-DTPA at 2.5 ml/s. Contrast injection and data acquisition were trigged simultaneously. Four repetitions were acquired with one repetition before the injection of contrast agent (L1) and three at 15 s (L2), 60 s (L3), and 120 s (L4) after the injection. Axial DWI images were obtained by using single-shot echo planar imaging sequence (SSEPI; TR/TE 5,900/69.6 ms; image resolution: 0.98 mm × 0.98 mm × 5 mm; 2-mm intersection gap) with two b-factors of 0 and 800 s/mm^2. The apparent diffusion coefficient (ADC) map was generated using these two DWI images with simple log-conversion. Patients' examples with multisequence MR images are shown in **Figure 2**. Both were 60-year-old men with mid-rectum cancer at stage of cT3N+M0. One was 5-year survival, and the other was not. No significant differences were observed from qualitative visual inspection.

Tumor Image Analysis

A region of interest (ROI) was delineated initially around the entire tumor by an experienced radiologist (10-year experience largely with colorectal MRI) using itk-SNAP software (www.itksnap.org) on each slice of T1w subtraction images (differences between the third-phase 60 s, after contrast injection and the first phase, before contrast injection), T2w, and DWI (b = 800 s/mm^2). Then, images were transferred over to a Velocity workstation (Varian Medical Health, Palo Alto, CA) by aligning all sequences under the same frame. The delineated tumor was further adjusted using all other image sequences as references. To minimize the partial volume effect and effect of contouring variance, the segmented contour was eroded by 1 mm from the border and the remaining region was used for radiomics analysis. After 3 months, 30 patients in the training set were selected randomly and segmented again by him and another radiologist (with 8 years of experience) to assess intra/inter-reader agreement of the radiomics analysis using intraclass correlation coefficients (ICCs). We interpreted an ICC of 0.81–1 as almost perfect agreement, 0.61–0.80 as substantial agreement, 0.41–0.60 as moderate agreement, 0.21–0.40 as fair, and 0–0.2 as poor agreement. An ICC of greater than 0.6 was considered as satisfactory inter- and intra-reader reproducibility.

All images were preprocessed with z-score intensity normalization and resampled into isotropic resolution (1 mm^3 × 1 mm^3 × 1 mm^3) with gray level quantized to 64 gray level. The

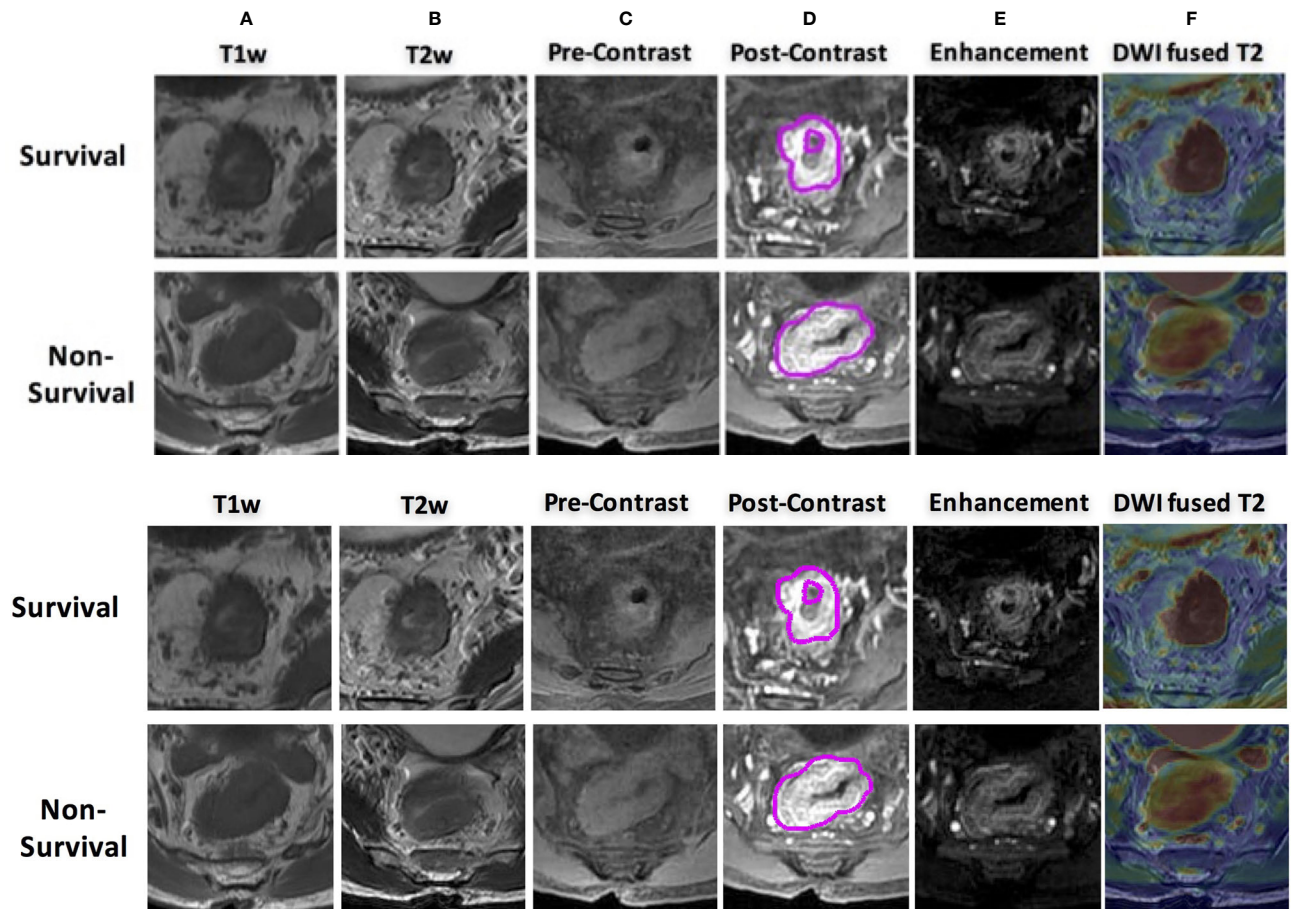


FIGURE 2 | MR images of two male patients, both at 60 years old with mid-rectum cancer at stage cT3N+M0, pretreatment **(A)** T1-weighted image, **(B)** T2-weighted image, **(C)** precontrast image, **(D)** 60 s after contrast injection image, **(E)** subtraction image showing difference between panels **(D, C, F)** the derived apparent ADC map fused to a T2-weighted image. The first row shows a case as 5-year survival patient, and the bottom shows a non-survival case.

radiomics analysis was performed using an in-house built program. Eight sequences as T1w, T2w, ADC map, L2, L3, L4, L2–L1, L3–L1, L4–L1) were analyzed. A total of 4,686 radiomics features from the category of morphology-, histogram-, texture-, and wavelet-based features were extracted from each individual patient: 1) For morphology features, 6 shape descriptors as volume, surface area, circularity, compactness, convexity, and irregularity were included; 2) 14 histogram-based features included min, mean, max, 90, 80, …, 10 percentiles, skewness, kurtosis; 3) 51 texture-based features included 7 gray-level run length matrix (GLRLM), as short run emphasis (SRE), long run emphasis (LRE), gray-level non-uniformity (GLN), run percentage (RP), run length non-uniformity (RLN), low gray-level run emphasis (LGRE), high gray-level run emphasis (HGRE). In this study, 44 gray-level co-occurrence matrix (GLCM)-based texture features were also collected with two distances as 1 and 2 pixels with each along 13 directions in 3D; 4) In addition, a discrete wavelet transform was used to decompose volumetric images into eight decomposed images, labeled as LLL, LLH, LHL, LHH, HLL, HLH, HHL, and HHH, where L and H are low- and high-frequency filters, respectively. For example, HHL represents a decomposition volume with high-pass filtering on X- and Y-directions and low-pass filtering along the z-direction. In each decomposed volume, histogram-based and texture-based features were extracted, resulting in a total of 520 (8 × 65) wavelet-transformed features for each sequence. In the end, a total of 4,686 (6 morphology+14 × 8 + 51 × 8 + 520 × 8) radiomics features were obtained from each patient.

Statistical Analysis

The patients treated from June 2008 to June 2011 (n = 114) were selected as a primary cohort, which was further randomly assigned into training and validation data sets with 4-fold cross-validation. The patients treated from July 2011 to July 2013 (n = 51) were assigned as a separate test cohort. The differences between OS vs. non-OS groups in both primary and test cohorts were compared using an analysis of variance (ANOVA) for continuous variables and Fisher's exact test for categorical variables. Multivariate analysis with a Cox regression analysis model was performed to detect independent prognostic factors for long-term survival.

Feature selection was performed in 3 steps to select the optimal survival-related features using the training cohort. Firstly, linear correlation between each pair features was evaluated using the Spearman test. Redundant features with linear correlation coefficient (arbitrary set) >0.90 were removed. Then, the least absolute shrinkage and selection operator (LASSO) regression algorithm was used for feature selection and model construction. LASSO regression shrinks the coefficient of unrelated feature toward zero, and related parameters are retained. The robustness of feature selection was tested by conducting a 4-fold cross-validation 100 times with binomial deviance minimization criteria from the primary cohort. Regression coefficients were estimated by LASSO. Lastly, the selected imaging features were then combined into a radiomics score through a linear combination of selected features weighted by their respective LASSO regression coefficients.

To further demonstrate the incremental values of radiomics signature to traditional risk factors, the discrimination performance between survival and non-survival groups was assessed with 1) TNM staging, 2) clinical-radiological (non-radiomics) features, 3) radiomics signature, and 4) combined all information. Discrimination was demonstrated by a receiver operating characteristic (ROC) curve in both primary and testing cohorts (with 1 indicating perfect prediction and 0.5 as no better concordance than chance). To provide clinicians a quantitative tool in predicting individual probability of 5-year OS, a nomogram was built based on multivariate logistic analysis in the primary cohort. Harrell's C-index was used to measure the nomogram discriminatory performance. Calibration curves accompanied by the Hosmer–Lemeshow test were used to assess the model fitting. A diagonal line along the calibration curve represented perfect agreement, and a significant p-value suggested a non-good fitting.

RESULTS

Clinical Characteristics

The present study included 165 patients with a mean age of 67 ± 13 years [standard deviation (SD)], range 19–89 years. All patients had minimal 60-month and up to 121-month follow-up. Patient characteristics are given in **Table 1**. As of the last follow-up, 123 patients (75%) experienced a confirmed 5-year (equal to or over 60-month) OS. Our data showed that OS for stage IIA was 80%, 67% for IIB, 78% for IIIA, 75% for IIIB, and 55% for IIIC.

All patients had formal rectal resections, 101 (61%) underwent LAR, 31 (19%) received APR, 23 (14%) had a Hartmann's procedure, and 10 (6%) received other types of resections. Patients who underwent LAR had the best prognosis, while those who underwent a Hartmann's procedure had the worst prognosis ($p \leq 0.05$). The presence of TDs (at the time of postsurgical evaluation) was also found to be associated with a poorer outcome as decreased OS.

A total of 95 patients (58%) had pathologically confirmed nodal involvement (43% with yN1 and 15% with yN2). Both nodal involvement and increasing node stage were significantly associated with poorer OS in both training and validation data sets ($p < 0.05$). The 5-year OS survival rate with yN0 is 84%, yN1 is 70%, and yN2 is 58% among all patient cohorts. The median number of examined lymph nodes per patient was 15 (range 3–37). Metastatic nodes were presented in 89 patients (53.9%). A positive LNR was found strongly correlated with poorer OS ($p < 0.04$).

Construction and Validation of Radiomics Signature

To avoid overfitting, the selected feature was limited to within 10:1 ratio relative to 114 patients in the primary cohorts, which means less than 11 features should be selected (27). **Figure 3** shows tuning parameter (λ) for feature selection with the values of coefficients closer to 0 with higher lambda. To keep the minimum binominal deviance, the number of features should be 32. Yet considering the number constraint of training cases, we decided to select 8 features that provided satisfactory performance and did not increase much denominal deviance. The radiomics features with a non-zero coefficient in the LASSO Cox regression model were as follows: LAVA2(Hist_Skewness), LAVA3(HLL_HistMax), LAVA3(GLCM_DiffrenceEntropy), T1w(LongRunEmphasis), T1w(HLL_HistMax), T2w(LLL_Hist10%), T2w(HLH_Hist40%), and ADC (GLCM_InfMeasCorr). Among the 8 selected features, 3 were from DCE-MRI LAVA sequences, 4 were from anatomical T1w or T2w images, and 1 was from DWI ADC map. The radiomics signature was constructed with relative weightings directly generated from LASSO regression (28). A Rad_score was calculated using the following formula:

$$\begin{aligned}\text{Rad\_score} = \text{sigmoid}[&-0.664 + 0.106 * \text{LAVA2}(\text{Hist}_{\text{Skewness}}) - 0.029\\ &* \text{LAVA3}(\text{HLL}_{\text{HistMAx}}) + 0.022 * \text{LAVA3}\left(\text{GLCM}_{\text{DifferenceEntropy}}\right) + 0.067\\ &* \text{T1w}(\text{LongRunEmphasis}) + 0.05 * \text{T1w}(\text{HLH}_{\text{HistMax}}) - 0.082\\ &* \text{T2w}(\text{LongRunEmphasis}) + 0.05 * \text{T2}(\text{HLH}_{\text{Hist40}}\,\%) - 0.092\\ &\text{ADC}(\text{GLCM}_{\text{InfMeasCorr}})]\end{aligned}$$

where $\text{sigmoid}(x) = [1+\exp(-x)]^{-1}$

Higher Rad-score patients generally had longer survival compared to those with lower scores. Distributions of the Rad_score in both primary and test cohorts are given in **Figure 4**. There was a significant difference in the Rad-score between 5-year OS vs. non-OS group in the primary cohort ($p < 0.001$*), which was also confirmed in the test cohort ($p < 0.001$*).

Incremental Prognostic Value of Radiomics to TNM Staging and Clinical-Histological-Radiological Features

The heatmap showing the correlation between the selected radiomics features and the clinical-histological-radiological factors is shown in **Figure 5**. For example, higher ADC (GLCM_InfMeasCorr) was associated with higher histology type. Most of the selected radiomics features showed a correlation with pretreatment CEA levels and adjuvant chemotherapy status. To further demonstrate the incremental

TABLE 1 | Patient Characteristics.

	Primary			Testing		
	5-yr OS (88)	**5-yr Non-OS (26)**	**P-value**	**5-yr OS (35)**	**5-yr Non-OS (16)**	**P-value**
Survival Months	82±15 [60,121]	30±15 [5,59]		82±13 [60,104]	36±14 [13,59]	
Clinical Data						
Age (years)	64±12 [29,84]	66±10 [45,80]	P=0.34	59±12 [19,89]	69±13 [40,89]	P=0.08
Gender						
Male	57	9	P=0.01*	23	10	P=0.8
Female	31	17		12	6	
Pre-tx CEA	7.3±15 [0.6,102]	26±46 [1.3,166]	P=0.04*	8.0±9.5 [1.1,43.5]	22±37 [2,116]	P=0.04*
cT stage						
T2	16	5	P=0.81	8	3	P=0.92
T3	57	17		26	13	
T4	15	4		1	0	
cN stage						
N0	54	15	P=0.32	22	8	P=0.06
N1	32	8		11	8	
N2	2	3		2	0	
Stage						
IIA	34	7	P=0.02*	18	6	P=0.07
IIB	1	1		1	0	
IIIA	3	1		4	1	
IIIB	40	9		11	8	
IIIC	10	8		1	1	
Tumor Deposit	22/50	14/23	P=0.05*	8/25	9/14	P=0.04*
Treatment Data						
Surgery						
LAR	14	10	P=0.01*	25	7	P=0.05*
APR	14	7		6	4	
Hartmann's	10	6		3	4	
Others	5	3		1	1	
Post-operative Chemo						
None	40	16	P=0.2	7	8	P=0.06
Folfox	32	6		18	5	
Xelox	16	4		10	3	
Radiological Data						
Location						
Lower	12	7	P=0.23	6	2	P=0.86
Mid	42	12		20	11	
High	34	7		9	3	
Dist from Anal Verge	7.8±2.9 [4,16]	8.9±4.1 [1.6,17]	P=0.21	8.8±4.2 [3,15]	10.5±4.5 [4,20]	P=0.21
EMVI						
0	38	6	P=0.8	1	1	P=0.9
1	2	0		2	1	
2	2	1		1	1	
3	36	11		1	0	
4	10	8		30	13	
Histo-Pathological Data						
yT stage						
T2	4	3	P=0.3	5	1	P=0.08
T3	79	21		29	12	
T4	5	2		1	3	
yN stage						
N0	40	6	P=0.04*	19	5	P=0.04*
N1	37	14		13	7	
N2	11	6		3	4	
Total LN	16±7 [3,37]	13±7 [5,28]	P=0.06	16±6 [7,33]	13±5 [8,29]	P=0.21
Positive Nodes	2±3 [0,10]	3±4 [0,13]	P=0.13	1.3±2.5 [0,8]	1.7±2.0 [0,11]	P=0.11
Positive LN Ratio	0.1±0.2 [0,0.8]	0.2±0.2 [0,0.7]	P=0.04*	0.08±0.17 [0,0.5]	0.16±0.13 [0,0.7]	P=0.04*
Histology Type						
I-well differentiated	34	9	P=0.9	14	4	P=0.6
II-mid differentiated	44	14		16	10	
III-poor differentiated	4	2		1	0	

(Continued)

TABLE 1 | Continued

	Primary			Testing		
	5-yr OS (88)	5-yr Non-OS (26)	P-value	5-yr OS (35)	5-yr Non-OS (16)	P-value
IV-mucoid or signet ring	6	1		4	2	
Radiomics Data						
Radiomics score	0.16±0.18 [-0.32,0.82]	-0.06±0.13 [-0.36,0.21]	P<0.001*	0.16±0.19 [-0.04,0.72]	-0.05±0.11 [-0.23,0.14]	P<0.001*

*OS, overall survival; Pre-tx CEA, pre-treatment carcinoembryonic level; cT-stage, clinical tumor stage; cN-stage, clinical node stage; APR, abdominoperineal resection; LAR, lower anterior resection; Dist, Distance; yT, pathological T-stage; yN, pathological N-stage; Total LN, total lymph node; Positive LN Ratio, positive lymph node ratio; EMVI, extramural vascular invasion; *: p≤0.05;*

prognostic value of radiomics to conventional prognostic features, four risk models in total were built: 1) TNM staging, 2) clinical-histological-radiological (non-radiomics), 3) radiomics signature, and 4) combined all selected features. The prognostic power in estimating 5-year OS in both primary and test data set was illustrated with ROC curves in **Figure 6**. The AJCC TNM staging system had the lowest area under the ROC curve (AUC) of 0.60 [95% confidence interval (CI): 0.50, 0.70] in the primary cohort and 0.54 (95% CI: 0.38, 0.69) in the test cohort, and the non-radiomics clinical model had a higher AUC of 0.66 (95% CI: 0.56, 0.77) and 0.56 (95% CI: 0.39, 0.73) in the primary set and test set, respectively. The radiomics signature yielded an AUC of 0.88 (95% CI: 0.82, 0.95), and a similar trend was observed in the test cohort with an AUC of 0.89 (95% CI: 0.79, 0.98). The combined model showed the best prognosis of survival outcome with an AUC of 0.91 (95% CI: 0.85, 0.96) in the primary set and 0.91 (95% CI: 0.81, 0.99) in the test set.

Furthermore, a nomogram with combined clinical-histopathological-radiological and radiomics features was developed in assessing the 5-year OS probability. The resulting nomograms can estimate outcome probability by assigning a score (upper scale) to each predictor value. The probability for 5-year OS (bottom scale) was a sum of these scores. The radiomics signature was the most important factor, followed by a positive LNR, surgery type (LAR vs. APR vs. Hartmann's vs. others), and pathological node status (yN). The final nomogram with calibration curves is illustrated in **Figure 7**. The C-index for the model was 0.898 (95% CI: 0.832, 0.947) within the primary cohort and 0.901 (95% CI: 0.803, 0.987) for the test cohort. The Hosmer–Lemeshow test yielded a nonsignificant statistic (p = 0.83), suggesting no departure from the perfect fit in the primary cohort. Good performance was also observed for the probability of pCR in the testing cohort with a nonsignificant statistic (p = 0.54). Patients were further classified into a high-risk group and a low-risk group with cutoff point identified on the ROC curve. Kaplan–Meier estimate of event rates over time showed statistically different outcomes for OS (p < 0.001). Results are also confirmed in the test data set as shown in **Figure 8**.

DISCUSSION

In this study, we extended the 5-year OS prediction beyond TNM schema to include clinical, histopathological, radiological, and high-throughput radiomics information. The combined nomogram outperformed the TNM staging or a clinical

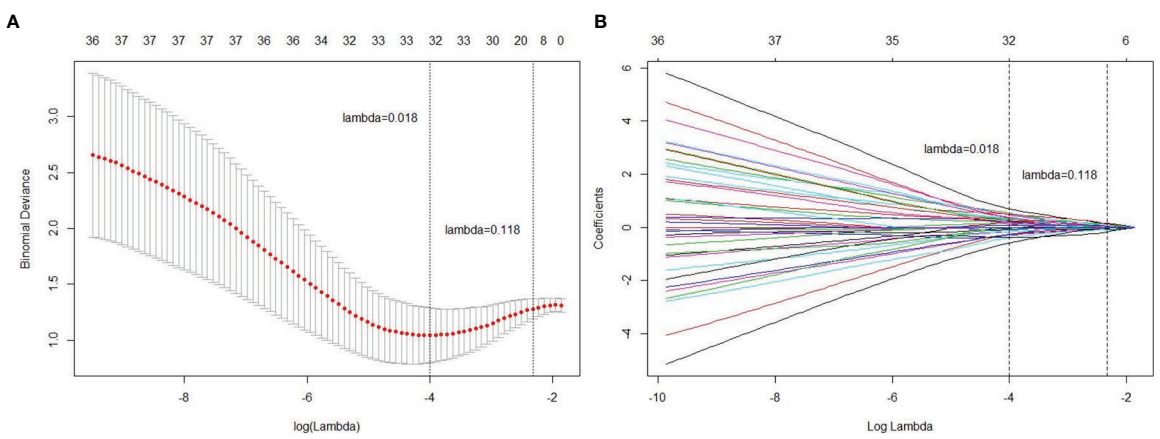


FIGURE 3 | Radiomics feature selection using the least absolute shrinkage and selection operator (LASSO) regression model. **(A)** Tuning parameter (lambda) selection in the LASSO model used 4-fold cross-validation *via* minimum criteria. **(B)** A coefficient profile plot was produced against the log (lambda) sequence. Vertical line was drawn at the value selected using 4-fold cross-validation, where optimal lambda resulted in 8 non-zero coefficients (features).

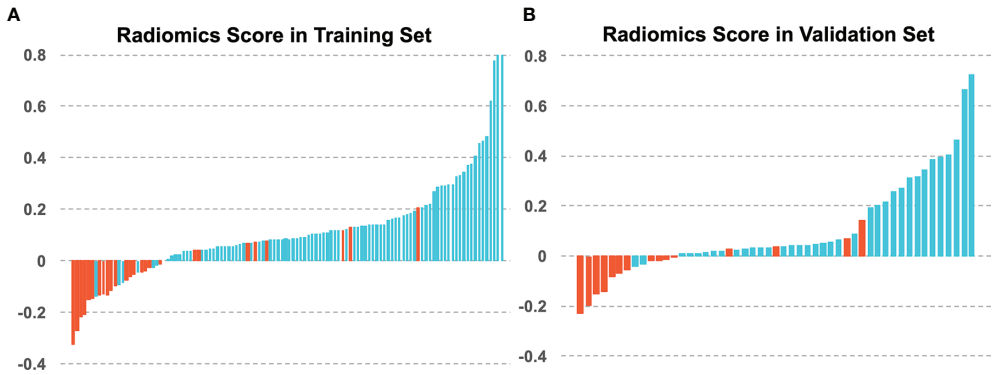


FIGURE 4 | The radiomics scores of each patient in **(A)** the training set and **(B)** the validation set, with blue for 5-year OS and red for 5-year non-OS patients.

system, which demonstrated the incremental value of the radiomics signature in individualized OS association in patients with stage II–III rectal cancer. Incorporating the radiomics signature and clinically available prognostic factors into an easy-to-use nomogram also allowed the model to support decision-making in daily practice, while further external validation is needed.

Despite rectal cancer comprising up to one-third of colorectal cancer cases and several prognosis nomograms having been established for colon cancer (13, 19), statistical tools for predicting long-term survival in rectal cancer are limited. Although MRI has become an indispensable tool for diagnosis by guiding treatment decisions for rectal cancer, the studies did not account for MRI findings. As such, in this study, we incorporated radiologists' qualitative assessment such as the depth of tumor spread beyond the muscularis propria, EMVI, and TDs into survival prediction. In addition, we extracted quantitative features from full-panel multiparametric MRIs with joint T1w, T2w, DCE-MRI, and DWI information and constructed a radiomics score for prognosis estimation. It is noteworthy that the identified radiomics signature consisted of three from dynamic MRI, four from anatomical MRI, and one from functional DWI. This study underlines the importance of multiparametric MRI in quantitative format to address the issue of long-term survival estimation in rectal cancer.

Pooling radiomics features to predict treatment outcome is still a relatively new concept. A few related studies that have been done focused on short-term treatment response as pathological complete response (pCR) to neoadjuvant chemoradiation therapy (CRT) (24, 25, 30). Although preoperative CRT has been demonstrated to improve local control, a recent meta-analysis and several clinical trials showed that there were no survival benefits for patients with stage II–III disease (2, 4, 7, 31, 32). This suggests that previously derived radiomics models may have limitations in estimating long-term clinical outcome. Instead, our study extends the "-omics"-based analysis to OS estimation, with all patients having a minimum of 60-month follow-up. In addition, unlike prior prognostic investigations that mostly analyzed patients with all stages of disease, our study focused exclusively on patients with stage II–III cancer. It is notable that when patients were stratified by clinical stage, there was a difference in OS for stage IIA vs. IIB, and stage IIIA vs. IIIB or IIIC, which suggested that heterogeneity existed in the survival outcomes. Our results showed that this radiomics signature was able to stratify patients into different risk levels beyond TNM staging or conventional clinical prognostic models. This is consistent with our current knowledge of cancer, in which malignant tumors consist of heterogeneous cell populations with distinct molecular and microenvironmental differences, increasing the likelihood of developing resistance to treatment and resulting in metastases (33, 34). In contrast, the traditional TNM staging or clinical assessment is based on gross anatomy information, with minimal regard for intratumor heterogeneity. With medical imaging, radiomics can extract features from the imaging characteristics of the entire tumor that provides a robust way to characterize the intratumor heterogeneity noninvasively.

Developing nomograms have been considered helpful in oncology prognosis (32, 35). In our study, a nomogram was built by combining selected features into a final score representing the probability of 5-year OS. Among them, the radiomics signature has the highest contribution, followed by positive LNR, types of surgery, and pathological N-staging. Increasing N-stage, as confirmed by previous studies, was found to be associated with poor survival (6, 36). The number of positive node/number of total nodes, which was revealed to be the most important risk factor of 5-year OS for colon cancer based on the well-known model from Weiser et al. (14), was only recently proposed for rectal cancer, but evidence is still limited (8). Our study confirmed that a positive LNR was an independent prognostic factor. Regarding the surgical type, our study patients with LAR operations had better survival outcomes compared to those receiving APR, as illustrated in the nomogram. There were conflicting reports regarding the prognostic effect of surgical types on OS rate. While most reported improved OS rates have been observed for LAR compared to APR (14, 36, 37), a few studies found no significant differences (38). Some previous studies also showed that the 5-year OS was higher for patients who had an LAR compared to those who had a Hartmann's procedure (36). The decision to perform a Hartmann's procedure was likely related to the individual patient characteristics such as severe

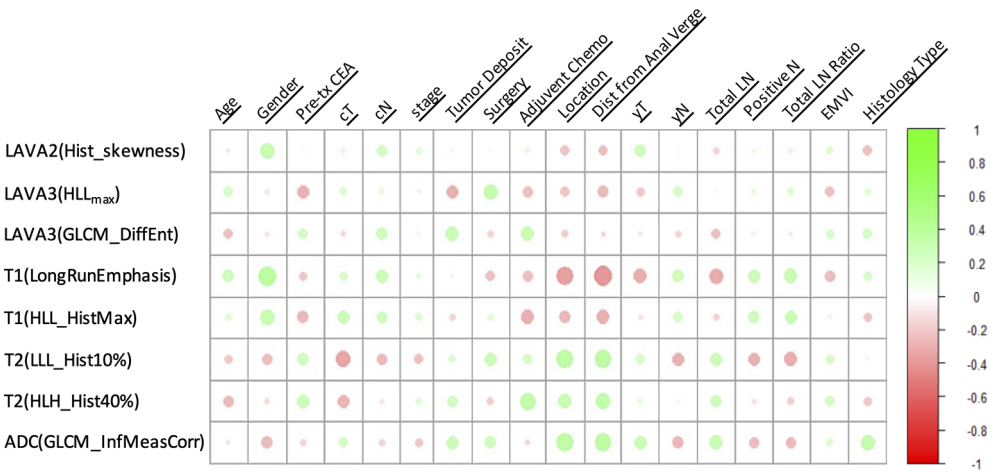


FIGURE 5 | A heatmap showing the correlation between selected radiomics features and clinical-histological-radiological parameters, with green as positive correlation and red as negative correlation.

cardiorespiratory or renal disease, which might impact survival. Interestingly, age was not found significantly correlated to OS. The reason for this is not immediately clear due to limited patient numbers. Age differences existed between good- and poor-survival group in the test cohort but not in the primary cohort. Since the model was developed using the primary cohort, age was not chosen as a predictor in this study. Other conventionally confirmed prognostic risk factors, such as pretreatment CEA level, TNM staging, TDs (39), were all found to be statistically different between better survival and worse survival groups, yet they were not selected into the final model as they were less important compared to radiomics signature.

To develop the radiomics signature, a total of 4,686 features were reduced to a set of only 8 potential descriptors using a LASSO regression model. In this method, the regression coefficients of most features were shrunk toward zero during model fitting, allowing identification of features that were strongly associated with OS. More importantly, it allowed radiomics signatures to be constructed into a generalized linear model. Recently, advanced machine learning-based methods have been used, yet the complexity of U-net-based deep learning or conventional neural network (CNN) typically requires intensive computation of the input, thus requiring a large sample set. Not surprisingly, Shi et al. (40) reported that the

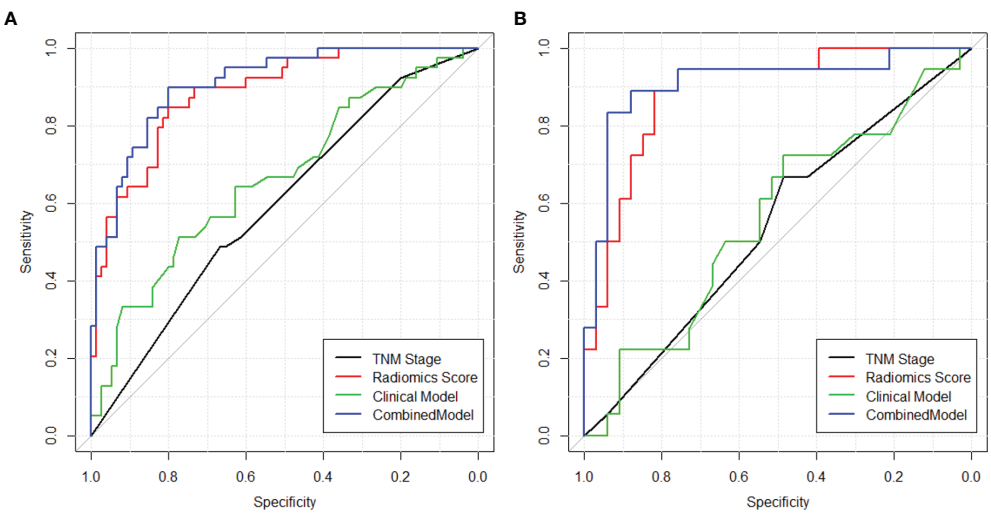


FIGURE 6 | ROC curves showing the prognostic power in estimating 5-year OS in both **(A)** training and **(B)** validation sets using (1) TNM staging (2), clinical model with clinical-histopathological-radiological (non-radiomics) risk factors, (3) radiomics score, and (4) combined model with both radiomics and non-radiomics information.

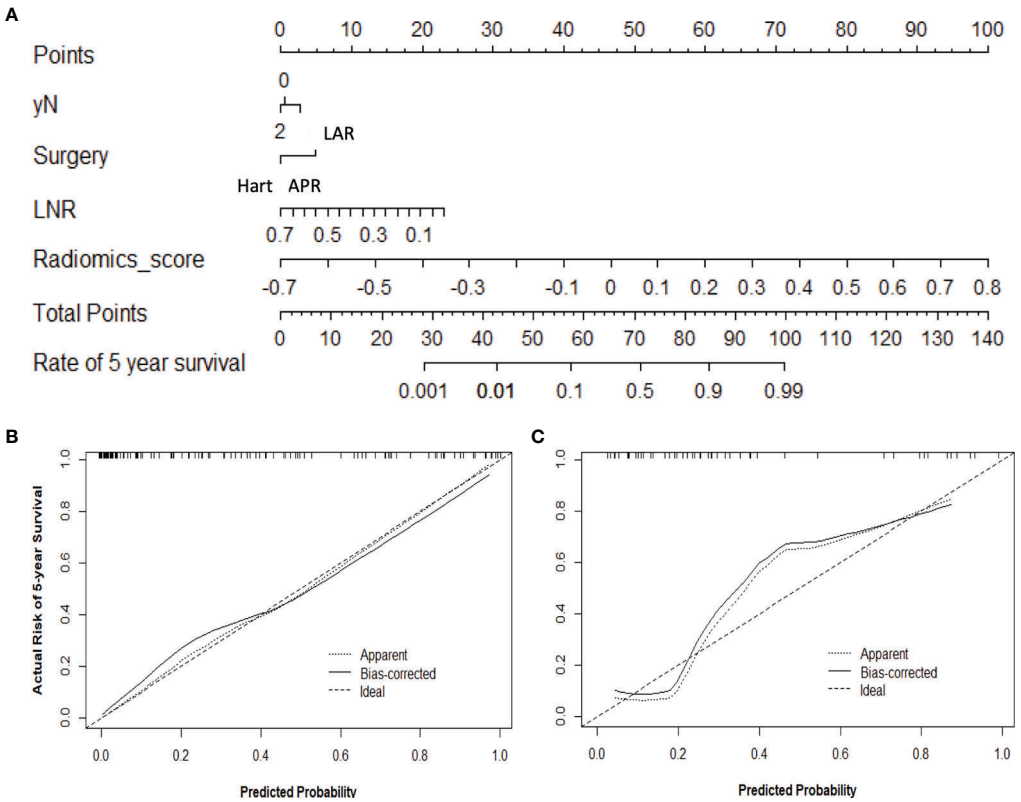


FIGURE 7 | Use of the constructed nomogram to estimate 5-year OS probability, along with the assessment of the model calibration, **(A)** nomogram with combined radiomics and non-radiomics clinical information, and calibration curves for the nomogram in the **(B)** training and **(C)** validation data sets. The y-axis represents the actual rate, and the x-axis represents the predicted probability of 5-year OS, with the diagonal line representing a perfect prediction. The solid line represents the performance of the radiomics model, of which a closer fit means better prediction.

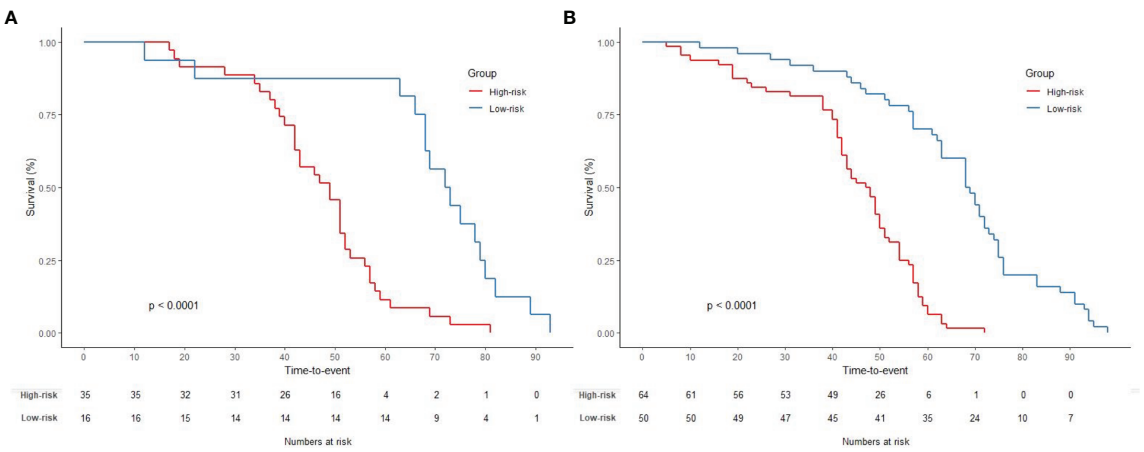


FIGURE 8 | Graphs show results of Kaplan–Meier survival analyses according to the radiomics signature for patients in **(A)** the primary data set (left) and those in **(B)** the test data set (right). A significant association of the radiomics signature with the OS was shown in the primary data set, which was then confirmed in the test data set.

prediction power of deep learning methods does not necessarily outperform conventional logistic regression in pCR prediction on a data set with 51 locally advanced rectal cancer patients. Other feature selection methods such as support vector machine or other deep learning-based algorithms, although beyond the scope of this paper, may be explored to further improve model performance (41). Moreover, by constructing multilayer nonlinear complex relationships, deep learning methods are more like a "black-box," making it difficult for physicians to interpret the association of the input images to outcome. In contrast, not only does LASSO surpass the method of choosing predictors based on the strength of their association with outcome, but it also enables the panel of selected features to be combined into a radiomics score. Nevertheless, the linear combination of selected features weighted by their respective coefficients provides an intuitive tool for the clinicians to interpret the results.

Limitations of this study included the lack of external validation with the retrospective nature of data collection. A large-scale prospective multicenter validation cohort is warranted to assess the generalizability of the reported findings; however, the protracted length of a longitudinal cohort with long-term survival data may make the research daunting. On the other hand, the statistical analysis with cross-validation used in this study justified that the identified radiomics model could hold great potential for clinical application in postoperative outcome estimation. Another limitation is that the radiological assessment by two radiologists was performed in consensus; it was impossible to assess the inter-reader agreement. Additionally, the tumor ROIs were manually performed, which was time-consuming and subjected to operators' variations. Semi- or full automatic segmentation may be needed in the future to improve the robustness of radiomics feature extraction. Moreover, due to the retrospective nature of this study, some important prognostic features such as margin positivity were not collected nor analyzed. Lastly, we only used LASSO for feature selection and model construction. It is known that L1 norm regularization can get sparse results. The variance of the model can be reduced using a higher lambda value, but it also results in a smaller number of features selected, thus leading to a biased model. Bias may be improved if using L2 norm instead. However, it will result in a large number of features selected, which leads to an overfitting. Other feature selection algorithms such as bridge and decision tree warrant to be further explored to achieve the best robust model.

Overall, the fact that clinicians naturally integrate multiple manifestations of disease to make an estimation and determine consequent therapy rather than focus on a single symptom underscores the necessity of multivariable estimation. Our study supported the suggestion that multiple variables could provide a more statistically meaningful approach to address the issue of long-term survival estimation using a multicomponent radiomics signature. The combined model with clinical-radiological-histological and radiomics signature yielded the highest AUC with 0.91 (95% CI: 0.85, 0.96) compared to any other models with TNM staging or clinical prognostic factors. The proposed nomograms were also well calibrated with nonsignificant p-values from Hosmer–Lemeshow test. Future work will include a large multi-institute prospective trial to validate the generalizability of the proposed model. Deep learning-based automatic segmentation is also under development to facilitate robust and efficient radiomics feature extraction. Nevertheless, although the usefulness of the proposed nomogram lacked external validation, the combined model, which considers multiple risk factors, was imperative. These findings and the nomogram may help patients with potential shorter OS to receive more aggressive treatment plans. Thus, our study may present a more efficient and integrated tool that enables earlier personalized treatment.

DATA AVAILABILITY STATEMENT

The software code used to analyze the current study is available upon request.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Department of Radiology, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

Conception and design: KN, YZ, XH, XC, and JS. Development of methodology: KN, YZ, PH, and PY. Analysis and interpretation of data: KN, YZ, PY, and JS. Writing, review, and/or revision of the article: KN, PH, JZ, PY, SJ, NY, XD, SX, BS, TN, XH, XC, and JS. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the National Key Research and Development Program of China (2016YFA0100900), the National Natural Science Foundation of China (81871403), the Key Research and Development Program of Zhejiang Province (2019C03014), and the Fundamental Research Funds for the Central Universities and research grant support from Varian Inc.

REFERENCES

1. Jawed I, Wilkerson J, Prasad V, Duffy AG, Fojo T. Colorectal Cancer Survival Gains and Novel Treatment Regimens: A Systematic Review and Analysis. *JAMA Oncol* (2015) 1(6):787–95. doi: 10.1001/jamaoncol.2015.1790
2. Bosset JF, Collette L, Calais G, Mineur L, Maingon P, Radosevic-Jelic L, et al. Chemotherapy With Preoperative Radiotherapy in Rectal Cancer. *New Engl J Med* (2006) 355(11):1114–23. doi: 10.1056/NEJMoa060829
3. Kapiteijn E, Marijnen CA, Nagtegaal ID, Putter H, Steup WH, Wiggers T, et al. Preoperative Radiotherapy Combined With Total Mesorectal Excision for Resectable Rectal Cancer. *New Engl J Med* (2001) 345(9):638–46. doi: 10.1056/NEJMoa010580
4. Sauer R, Liersch T, Merkel S, Fietkau R, Hohenberger W, Hess C, et al. Preoperative Versus Postoperative Chemoradiotherapy for Locally Advanced Rectal Cancer: Results of the German CAO/ARO/AIO-94 Randomized Phase III Trial After a Median Follow-Up of 11 Years. *J Clin Oncol* (2012) 30 (16):1926–33. doi: 10.1200/JCO.2011.40.1836
5. van Gijn W, van Stiphout R, van de Velde CJH, Valentini V, Lammering G, Gambacorta MA, et al. Nomograms to Predict Survival and the Risk for Developing Local or Distant Recurrence in Patients With Rectal Cancer Treated With Optional Short-Term Radiotherapy. *Ann Oncol* (2015) 26 (5):928–35. doi: 10.1093/annonc/mdv023
6. Valentini V, van Stiphout RG, Lammering G, Gambacorta MA, Barba MC, Bebenek M, et al. Nomograms for predicting local recurrence, distant metastases, and overall survival for patients with locally advanced rectal cancer on the basis of European randomized clinical trials. *J Clin Oncol* (2011) 29(23):3163–72. doi: 10.1200/JCO.2010.33.1595
7. Babaei M, Jansen L, Balavarca Y, Sjovall A, Bos A, van de Velde T, et al. Neoadjuvant Therapy in Rectal Cancer Patients With Clinical Stage II to III Across European Countries: Variations and Outcomes. *Clin Colorectal Cancer* (2018) 17(1):e129–42. doi: 10.1016/j.clcc.2017.09.002
8. Dekker JW, Peeters KC, Putter H, Vahrmeijer AL, van de Velde CJ. Metastatic Lymph Node Ratio in Stage III Rectal Cancer; Prognostic Significance in Addition to the 7th Edition of the TNM Classification. *Eur J Surg Oncol* (2010) 36(12):1180–6. doi: 10.1016/j.ejso.2010.09.007
9. Nagtegaal ID, Quirke P, Schmoll HJ. Has the New TNM Classification for Colorectal Cancer Improved Care? *Nat Rev Clin Oncol* (2011) 9(2):119–23. doi: 10.1038/nrclinonc.2011.157
10. Yagi R, Shimada Y, Kameyama H, Tajima Y, Okamura T, Sakata J, et al. Clinical Significance of Extramural Tumor Deposits in the Lateral Pelvic Lymph Node Area in Low Rectal Cancer: A Retrospective Study at Two Institutions. *Ann Surg Oncol* (2016) 23(Suppl 4):552–8. doi: 10.1245/s10434-016-5379-9
11. Ueno H, Mochizuki H, Shirouzu K, Kusumi T, Yamada K, Ikegami M, et al. Multicenter Study for Optimal Categorization of Extramural Tumor Deposits for Colorectal Cancer Staging. *Ann Surg* (2012) 255(4):739–46. doi: 10.1097/SLA.0b013e31824b4839
12. Kawai K, Nozawa H, Hata K, Kiyomatsu T, Tanaka T, Nishikawa T, et al. Nomogram Predicting Survival After Recurrence in Patients With Stage I to III Colon Cancer: A Nationwide Multicenter Study. *Dis Colon Rectum* (2018) 61(9):1053–62. doi: 10.1097/DCR.0000000000001167
13. Lemini R, Attwood K, Pecenka S, Grego J, Spaulding AC, Nurkin S, et al. Stage II-III Colon Cancer: A Comparison of Survival Calculators. *J Gastrointest Oncol* (2018) 9(6):1091–8. doi: 10.21037/jgo.2018.08.03
14. Weiser MR, Gonen M, Chou JF, Kattan MW, Schrag D. Predicting Survival After Curative Colectomy for Cancer: Individualizing Colon Cancer Staging. *J Clin Oncol* (2011) 29(36):4796–802. doi: 10.1200/JCO.2011.36.5080
15. Song J, Chen Z, Huang D, Wu Y, Lin Z, Chi P, et al. Nomogram Predicting Overall Survival of Resected Locally Advanced Rectal Cancer Patients With Neoadjuvant Chemoradiotherapy. *Cancer Manag Res* (2020) 12:7375–82. doi: 10.2147/CMAR.S255981
16. Halabi S, Lin CY, Kelly WK, Fizazi KS, Moul JW, Kaplan EB, et al. Updated Prognostic Model for Predicting Overall Survival in First-Line Chemotherapy for Patients With Metastatic Castration-Resistant Prostate Cancer. *J Clin Oncol* (2014) 32(7):671–7. doi: 10.1200/JCO.2013.52.3696
17. Salazar R, Roepman P, Capella G, Moreno V, Simon I, Dreezen C, et al. Gene Expression Signature to Improve Prognosis Prediction of Stage II and III Colorectal Cancer. *J Clin Oncol* (2011) 29(1):17–24. doi: 10.1200/JCO.2010.30.1077
18. Stirzaker C, Zotenko E, Song JZ, Qu W, Nair SS, Locke WJ, et al. Methylome Sequencing in Triple-Negative Breast Cancer Reveals Distinct Methylation Clusters With Prognostic Value. *Nat Commun* (2015) 6:5899. doi: 10.1038/ncomms6899
19. Zhang JX, Song W, Chen ZH, Wei JH, Liao YJ, Lei J, et al. Prognostic and Predictive Value of a microRNA Signature in Stage II Colon Cancer: A microRNA Expression Analysis. *Lancet Oncol* (2013) 14(13):1295–306. doi: 10.1016/S1470-2045(13)70491-1
20. Zhao SG, Chang SL, Erho N, Yu M, Lehrer J, Alshalalfa M, et al. Associations of Luminal and Basal Subtyping of Prostate Cancer With Prognosis and Response to Androgen Deprivation Therapy. *JAMA Oncol* (2017) 3(12):1663–72. doi: 10.1001/jamaoncol.2017.0751
21. Kuo MD, Jamshidi N. Behind the Numbers: Decoding Molecular Phenotypes With Radiogenomics–Guiding Principles and Technical Considerations. *Radiology* (2014) 270(2):320–5. doi: 10.1148/radiol.13132195
22. Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Carvalho S, et al. Decoding Tumour Phenotype by Noninvasive Imaging Using a Quantitative Radiomics Approach. *Nat Commun* (2014) 5:4006. doi: 10.1038/ncomms5006
23. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, et al. Radiomics: Extracting More Information From Medical Images Using Advanced Feature Analysis. *Eur J Cancer* (2012) 48(4):441–6. doi: 10.1016/j.ejca.2011.11.036
24. Nie K, Shi L, Chen Q, Hu X, Jabbour SK, Yue N, et al. Rectal Cancer: Assessment of Neoadjuvant Chemoradiation Outcome Based on Radiomics of Multiparametric MRI. *Clin Cancer Res an Off J Am Assoc Cancer Res* (2016) 22 (21):5256–64. doi: 10.1158/1078-0432.CCR-15-2997
25. Zhou X, Yi Y, Liu Z, Cao W, Lai B, Sun K, et al. Radiomics-Based Pretherapeutic Prediction of Non-Response to Neoadjuvant Therapy in Locally Advanced Rectal Cancer. *Ann Surg Oncol* (2019) 26(6):1676–84. doi: 10.1245/s10434-019-07300-3
26. Smith NJ, Barbachano Y, Norman AR, Swift RI, Abulafi AM, Abulafi AM, et al. Prognostic Significance of Magnetic Resonance Imaging-Detected Extramural Vascular Invasion in Rectal Cancer. *Br J Surg.* 95(2):229–36. doi: 10.1002/bjs.5917
27. Nie K, Al-Hallaq H, Li XA, Benedict SH, Sohn JW, Moran JM, et al. NCTN Assessment on Current Applications of Radiomics in Oncology. *Int J Radiat Oncol Biol Phys* (2019) 104(2):302–15. doi: 10.1016/j.ijrobp.2019.01.087
28. 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 an Off J Am Assoc Cancer Res* (2017) 23(23):7253–62. doi: 10.1158/1078-0432.CCR-17-1038
29. Guyon I, Elisseeff A. An Introduction to Variable and Feature Selection. *J Mach Learn Res* (2003) 3:1157–82. doi: 10.1162/153244303322753616
30. 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
31. Quasar Collaborative G, Gray R, Barnwell J, McConkey C, Hills RK, Williams NS, et al. Adjuvant Chemotherapy Versus Observation in Patients With Colorectal Cancer: A Randomised Study. *Lancet (London England)* (2007) 370 (9604):2020–9. doi: 10.1016/S0140-6736(07)61866-2
32. Sebag-Montefiore D, Stephens RJ, Steele R, Monson J, Grieve R, Khanna S, et al. Preoperative Radiotherapy Versus Selective Postoperative Chemoradiotherapy in Patients With Rectal Cancer (MRC CR07 and NCIC-CTG C016): A Multicentre, Randomised Trial. *Lancet (London England)* (2009) 373(9666):811–20. doi: 10.1016/S0140-6736(09)60484-0
33. Park H, Lim Y, Ko ES, Cho HH, Lee JE, Han BK, et al. Radiomics Signature on Magnetic Resonance Imaging: Association With Disease-Free Survival in Patients With Invasive Breast Cancer. *Clin Cancer Res an Off J Am Assoc Cancer Res* (2018) 24(19):4705–14. doi: 10.1158/1078-0432.CCR-17-3783
34. Zhang B, Tian J, Dong D, Gu D, Dong Y, Zhang L, et al. Radiomics Features of Multiparametric MRI as Novel Prognostic Factors in Advanced Nasopharyngeal Carcinoma. *Clin Cancer Res an Off J Am Assoc Cancer Res* (2017) 23(15):4259–69. doi: 10.1158/1078-0432.CCR-16-2910

35. Shariat SF, Capitanio U, Jeldres C, Karakiewicz PI. Can Nomograms be Superior to Other Prediction Tools? *BJU Int* (2009) 103(4):492–495; discussion 495-497. doi: 10.1111/j.1464-410X.2008.08073.x
36. Bown EJ, Lloyd GM, Boyle KM, Miller AS. Rectal Cancer: Prognostic Indicators of Long-Term Outcome in Patients Considered for Surgery. *Int J Colorectal Dis* (2014) 29(2):147–55. doi: 10.1007/s00384-013-1772-z
37. Temple LK, Romanus D, Niland J, Veer AT, Weiser MR, Skibber J, et al. Factors Associated With Sphincter-Preserving Surgery for Rectal Cancer at National Comprehensive Cancer Network Centers. *Ann Surg* (2009) 250 (2):260–7. doi: 10.1097/SLA.0b013e3181ae330e
38. Law WL, Chu KW. Abdominoperineal Resection is Associated With Poor Oncological Outcome. *Br J Surg* (2004) 91(11):1493–9. doi: 10.1002/bjs.4723
39. Nagtegaal ID, Knijn N, Hugen N, Marshall HC, Sugihara K, Tot T, et al. Tumor Deposits in Colorectal Cancer: Improving the Value of Modern Staging-A Systematic Review and Meta-Analysis. *J Clin Oncol* (2017) 35 (10):1119–27. doi: 10.1200/JCO.2016.68.9091
40. Shi L, Zhang Y, Nie K, Sun X, Niu T, Yue N, et al. Machine Learning for Prediction of Chemoradiation Therapy Response in Rectal Cancer Using Pre-Treatment and Mid-Radiation Multi-Parametric MRI. *Magnet Res Imaging* (2019) 61:33–40. doi: 10.1016/j.mri.2019.05.003
41. Wang Q, Zhang Y, Zhang E, Xing X, Chen Y, Su MY, et al. Prediction of the Early Recurrence in Spinal Giant Cell Tumor of Bone Using Radiomics of Preoperative CT: Long-Term Outcome of 62 Consecutive Patients. *J Bone Oncol* (2021) 27:100354. doi: 10.1016/j.jbo.2021.100354

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 Technology and Code
PUBLISHED 11 August 2022
DOI 10.3389/fonc.2022.798460

OPEN ACCESS

EDITED BY
Jose Eduardo Villarreal Barajas,
Royal Devon and Exeter Hospital,
United Kingdom

REVIEWED BY
Ayşegul Yurt,
Dokuz Eylül University, Turkey
James Chow,
University of Toronto, Canada

*CORRESPONDENCE
Massimiliano Porzio
massimiliano.porzio@gmail.com

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

RECEIVED 20 October 2021
ACCEPTED 19 July 2022
PUBLISHED 11 August 2022

CITATION
Porzio M and Anam C (2022) Real-time fully automated dosimetric computation for CT images in the clinical workflow: A feasibility study. *Front. Oncol.* 12:798460. doi: 10.3389/fonc.2022.798460

Real-time fully automated dosimetric computation for CT images in the clinical workflow: A feasibility study

Massimiliano Porzio[1]* and Choirul Anam[2]

[1]Department of Fisica Sanitaria, Azienda Sanitaria Locale Cuneo1 (ASL CN1), Cuneo, Italy, [2]Department of Physics, Faculty of Sciences and Mathematics, Diponegoro University, Semarang, Indonesia

Background: Currently, the volume computed tomography dose index ($CTDI_{vol}$), the most-used quantity to express the output dose of a computed tomography (CT) patient's dose, is not related to the real size and attenuation properties of each patient. The size-specific dose estimates (SSDE), based on the water-equivalent diameter (D_W) overcome those issues. The proposed methods found in the literature do not allow real-time computation of D_W and SSDE.

Purpose: This study aims to develop a software to compute D_W and SSDE in a real-time clinical workflow.

Method: In total, 430 CT studies and scans of a water-filled funnel phantom were used to compute accuracy and evaluate the times required to compute the D_W and SSDE. Two one-sided tests (TOST) equivalence test, Bland–Altman analysis, and bootstrap-based confidence interval estimations were used to evaluate the differences between actual diameter and D_W computed automatically and between D_W computed automatically and manually.

Results: The mean difference between the D_W computed automatically and the actual water diameter for each slice is –0.027% with a TOST confidence interval equal to [–0.087%, 0.033%]. Bland–Altman bias is –0.009% [–0.016%, –0.001%] with lower limits of agreement (LoA) equal to –0.0010 [–0.094%, –0.068%] and upper LoA equal to 0.064% [0.051%, 0.077%]. The mean difference between D_W computed automatically and manually is –0.014% with a TOST confidence interval equal to [–0.056%, 0.028%] on phantom and 0.41% with a TOST confidence interval equal to [0.358%, 0.462%] on real patients. The mean time to process a single image is 13.99 ms [13.69 ms, 14.30 ms], and the mean time to process an entire study is 11.5 s [10.62 s, 12.63 s].

Conclusion: The system shows that it is possible to have highly accurate D_W and SSDE in almost real-time without affecting the clinical workflow of CT examinations.

KEYWORDS

automatic, software, water-equivalent, SSDE, CT, dose, real-time

Introduction

A computed tomography (CT) scanner is often used to accurately diagnose cancer. In addition, CT is used in the treatment planning of radiotherapy (1). A CT image accurately provides the shape and position of cancer and surrounding healthy tissues, along with organs at risk. CT provides a map of the electron density information for the various tissues to accurately and precisely calculate the dose delivered to the patients during radiotherapy. However, CT employs ionizing radiation, and it potentially induces new cancer in the future (2–6). CT delivers higher ionizing radiation doses than common radiographic examinations (7).

Currently, the quantities used to express CT dose are the volume computed tomography dose index ($CTDI_{vol}$) and the dose-length product (DLP; the $CTDI_{vol}$ multiplied by the scan length) (8). Those values do not directly estimate the patient dose but are measured from the output of the X-ray tube on cylindrical plastic phantoms representing an average patient head (16 cm in diameter and 15 cm thick) and/or torso (32 cm in diameter and 15 cm thick).

Those phantoms do not represent the actual patient dimensions since the dose received by the patient varies according to the size and attenuation of the patient. A small patient would receive a higher radiation dose than a bigger one, even if their $CTDI_{vol}$ is the same (9, 10). To overcome those issues, the American Association of Physicists in Medicine (AAPM) in 2011 introduced a new metric called the size-specific dose estimates (SSDE) (11).

The computation of SSDE relies on the water-equivalent diameter (D_W), a quantity that accounts for patient composition and attenuation properties (12). It is possible to compute D_W for every CT image using manual (13) or automated techniques (14, 15).

Many studies (16–18) proposed fully automated methods. Özsoykal et al. (19) proposed an automated patient contour after the exclusion of irrelevant objects such as clothes and the CT table from the image. The threshold value was determined through trial and error until a complete successful segmentation of the body contour was obtained. Anam et al. (14) also proposed an automated approach to calculate D_W in human anatomic regions and phantoms using a region of interest (ROI) that is automatically fitted to the patient border. The automated approach produced an excellent correlation with the manual one ($R^2 = 0.999$). Gharbi et al. (20) also successfully proposed an automated approach to measure D_W based on the Fuzzy C-means classification and edge detection algorithms. Juszczyk et al. (21) proposed an automated segmentation approach to calculate D_W using a convolutional neural network (CNN) and reported that the proposed method produces accurate results. However, they are retrospective techniques and do not allow for real-time assessment of D_W and SSDE. Real-time measurements could be useful for assessing whether the correct technique is being applied to the patient, or whether it is necessary for some changes to account for the patient's size [e.g., tube current modulation (TCM)].

Furthermore, a rapid SSDE computational system would benefit many oncological patients who undergo numerous CT exams and need their exposures tailored to their specific sizes (22) and reduced as low as possible to reduce the risk of relapse and/or radiation-induced new cancer (2–4).

The aim of this study is to develop a fully automated method of D_W and SSDE measurement for real-time and to investigate the feasibility of its implementation in a clinical workflow, evaluating its accuracy and the time required to do it for a complete CT study.

Method

Main algorithm description

A previously published work was the basis of the algorithm (14). The workflow is described in Figure 1. The language used was Java, and the main third-party library is ImageJ (RRID: SCR_003070) (23). The input was the path on the filesystem of a series of CT images and the path of the text file where the user wants to save the computed dosimetric quantities.

Images were first ordered by the image number read in the Digital Imaging and Communications in Medicine (DICOM) header attribute. The middle slice[1] then becomes the reference for other study-related information read from the DICOM header (i.e., accession number, station name, series number, pixel dimensions, $CTDI_{vol}$, protocol name, sex, and age). We read the anatomical region from the protocol name attribute.

D_W and SSDE computations

The algorithm implemented a combination of basic segmentation techniques and specific information about the border of the patient's body using the "Analyze particles" (24) plugin.

For every image, we computed D_W and SSDE using the following steps: *First*, reading the image—ImageJ opens the images and converts them into Hounsfield units (HU). *Second*, applying the threshold (−410 HU, using the Analyze particles [24) plugin)] in order to detect border regions of interest (ROIs) with an area greater than 3,000 mm^2. We selected this value (−410 HU) because the border between the patient and its surroundings is skin, with pixel values of approximately 0 HU, while the surroundings outside the patient are air or other materials with pixel values lower than −410 HU. The thresholding produces binary images. However, thresholding alone could not contour the patient completely because of the

[1] The first slice may not contain a useful body region.

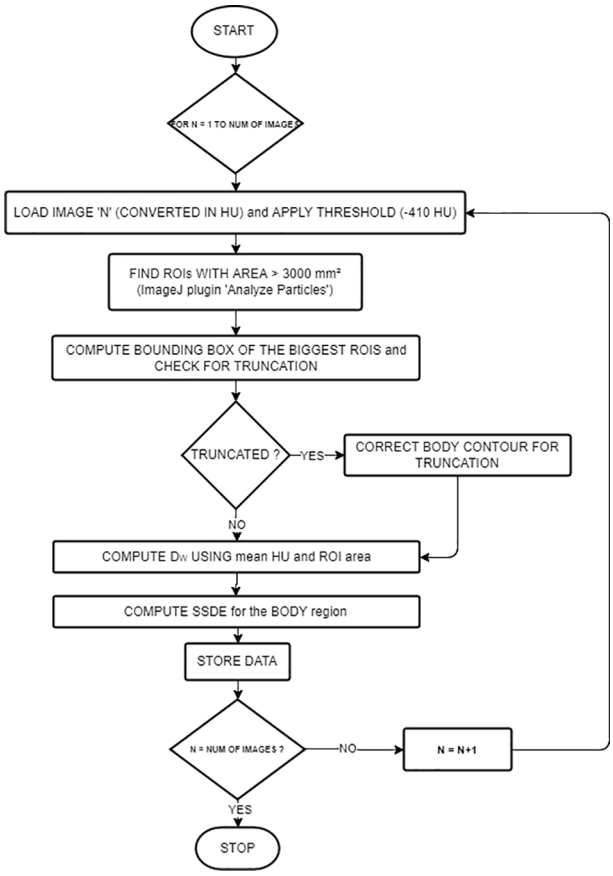


FIGURE 1
Workflow of the main algorithm for D_W and SSDE computations.

presence of other objects inside the patient with pixel values lower than −410 HU. To overcome this problem, the plugin implemented edge detection to identify these objects and label them using their areas. We considered the largest area identified to be the border of the patient or phantom. If the plugin finds multiple ROIs, we choose the ROI with the centroid nearest to the center of the image.

Third, computing the bounding rectangle of the ROI to see if the patient's body is truncated as in (25). If so, we first applied a correction to the ROI for letting the border of the patient not follow the lung contour but go straight along the image sides (see Figure 2), and then we applied the correction presented in (25), based on the percentage of truncation of the patient's body (25) to estimate more correctly the water-equivalent diameter, based on the anatomical region we previously found.

Fourth, computing means HU, ROI area, and then D_W using:

$$D_W = 2\sqrt{\left[\frac{1}{1,000}\overline{\mathbf{HU}(x,y)_{\mathbf{ROI}}}\right]\frac{A_{\mathbf{ROI}}}{\pi}} \quad (1)$$

Fifth, SSDE was computed according to the anatomical region (head or "other") and the reconstruction diameter. The SSDE for head CTs was computed with

$$\mathbf{SSDE} = 1.9852e^{-0.0486D_W} \quad (2)$$

The SSDE for body CTs, based on the reconstructed diameter previously read from the DICOM header along with the anatomical region inferred from the protocol name, was computed with:

$$\mathbf{SSDE} = \begin{cases} 1.877\mathbf{e}^{-0.039D_W} \text{ if reconstr . diam . } \leq 320\,\text{mm} \\ 3.7055e^{-0.037D_W} \text{ if } 320\,\text{mm} < \text{reconstr . diam . } \leq 400\,\text{mm} \end{cases} \quad (3)$$

Equation (3) separates CT images with $CTDI_{vol}$ computed for a 16-cm diameter and those with $CTDI_{vol}$ for a 32-cm phantom.

For every image, we also computed the time elapsed from the reading of the DICOM file to the end of computations of all the quantities.

We wrote on a text file the accession number, series number, protocol name, station name, D_W, SSDE, $CTDI_{vol}$, the

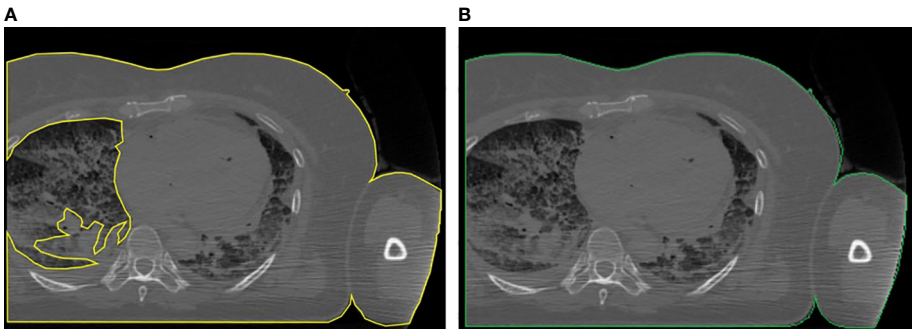


FIGURE 2
Example of the correction applied to truncated images: **(A)** body region not properly located and **(B)** body region after the correction.

percentage of truncation—computed as in (25)—and the time elapsed for analyzing the image.

Real-time software architecture

The described algorithm was at the core of the system. We built a software architecture that runs this algorithm without affecting the communication between CT scanners and the Picture Archive and Communication System (PACS). Figure 3 shows the architecture of our system.

We decoupled the dosimetric computation from the clinical workflow, using a routing rule of the institutional PACS archive. It sent every received CT series to another PACS installed in our Medical Physics Department (Orthanc DICOM storage server (26)).

The Orthanc server uses Lua (27) scripts to monitor DICOM instances and react accordingly. A Lua script filtered the CT series and, if the related metadata (read from the DICOM header) were suitable for the study, then it saved the images on disk and writes a message to an ActiveMQ queue. The message contains the path where the Lua script saved the CT images of each series.

ActiveMQ is an open-source message broker (28) for managing messages received from other software components and letting other components react to each message's arrival.

Another Java software component is listening for messages on the queue and, for every message, it takes the path of CT images and runs the main algorithm, saving dosimetric information on a text file.

Algorithm validation

We validated our algorithm using a homemade funnel phantom used in a previous study (29) filled with water (Figure 4). It was composed of plastic, and its diameter range was 10 ÷ 34 cm with an effective length of 14 cm (total length: 36 cm). We scanned the phantom on a Philips Brilliance 40 CT scanner and computed the manual D_W for a very thin slice, to compare automatic D_W computation with the actual diameter and with D_W computed manually. The manual computation involved body contouring and collecting statistics (mean CT number and area of the contoured ROI). We repeated the manual calculation on five actual patient series and compared the D_W to the automatic computation.

Data collection

We collected data for a week on CT scans of patients with an age of >18 years and with scanning protocols related to head, head and neck, thorax, abdomen, and/or pelvis anatomical regions. We computed D_W and SSDE from five CT scanners (see Table 1). The number of CT series, after the removal of outliers[2] in the elapsed time, was 1,789 (1,013 abdomen/pelvis, 432 head, and 344 thorax). Summing the elapsed time for images and series with the same accession number, we gathered the time to process a whole CT study. The number of slices collected ranged from 14 to 4,460, with a median value of 585 slices per patient. The total number of CT studies was 430.

To show the potential of our approach, we draw the SSDE and $CTDI_{vol}$ vs. slice number for one patient, as well as the D_W vs. slice number, using the collected data, in order to show the type of information a final user can get with little further effort.

Statistical analysis

The R language (30) was used to perform all the statistical analysis. We evaluated the differences between the actual

[2] A processing time greater than 1.5 times the interquartile range of all processing times.

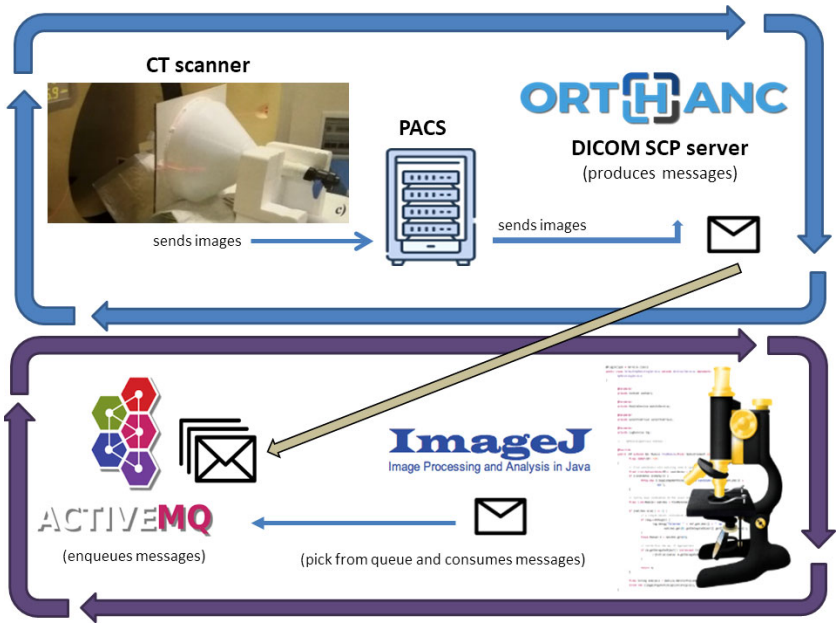


FIGURE 3
Software architecture of our system: Images received from the Institutional PACS archive are sent for storage to an Orthanc DICOM server. The server sends messages on an ActiveMQ queue using the Lua script. A Java software built on ImageJ API does D_W and SSDE computations for every message it reads from the queue.

FIGURE 4
Our funnel phantom (plastic filled with water).

TABLE 1 CT scanners used in this study.

Manufacturer	Model
Philips	Brilliance CT 6-slice
Philips	Brilliance CT 16-slice
Philips	Brilliance CT 16-slice
Philips	Brilliance CT 40-slice
Toshiba	Aquilion 64-slice

diameter and D_W computed automatically for the funnel phantom with two one-sided test (TOST) (31) (equivalence bounds: ± 7%) and Bland–Altman analysis (32). We also evaluated the difference between manual and automatic D_W computation using TOST (equivalence bounds: ± 10% for the funnel phantom). On five patient series, we compared automatic and manual D_W using TOST with 12% as equivalence bounds. We chose equivalence bounds based on the International

Electrotechnical Commission (IEC) norm IEC 62985:2019 (Methods for calculating size-specific dose estimates (SSDE) for computed tomography).

We analyzed patients' data in terms of mean processing time for a single image and for a single study. We computed confidence intervals (CI) for mean times with non-parametric bootstrap using the "nptest" package (33). We set the significance level to $\alpha = 0.01$ for TOST analysis and to $\alpha = 0.05$ for the Bland–Altman confidence interval significance level. All CIs are reported in square brackets, i.e., [lower CI limit, upper CI limit].

Finally, we computed the mean percentage of truncation, following (25), along with its bootstrapped CI using the "nptest" package (33).

Results

D_W of water-filled funnel phantom

Figure 5 shows the Bland–Altman plot for the D_W of the water-filled funnel computed and the actual funnel diameter. Bland–Altman bias is −0.009% [−0.016%, −0.001%] with lower limits of agreement (LoA) equal to −0.0010 [−0.094%, −0.068%] and upper LoA equal to 0.064% [0.051%, 0.077%]. Figure 6A displays the TOST plot of the D_W of the water-filled funnel computed automatically and the actual funnel diameter. The D_W of the water-filled funnel is computed automatically, and the actual funnel diameter is statistically not different from zero ($p < 0.001$) and statistically equivalent to zero ($p < 0.001$), given the ±7% equivalence bounds. The mean difference between the D_W computed automatically and the actual funnel diameter for each slice is −0.027% with a TOST confidence interval equal to [−0.087%, 0.033%].

Figure 6B displays the TOST plot of the D_W of the water-filled funnel computed automatically and manually. The D_W of the water-filled funnel computed automatically and manually is statistically not different from zero ($p < 0.001$) and statistically equivalent to zero ($p < 0.001$), given the ±10% equivalence bounds. The mean difference between the D_W of the water-filled funnel computed automatically and manually is −0.014% with a TOST confidence interval equal to [−0.056%, 0.028%].

D_W of patients

Figure 6C shows the TOST plot of the D_W of patients computed automatically and manually. The D_W of patients computed automatically and manually is statistically different from zero ($p < 0.001$) with a mean difference equal to 0.41% and statistically equivalent to zero ($p < 0.001$) given the ±12%

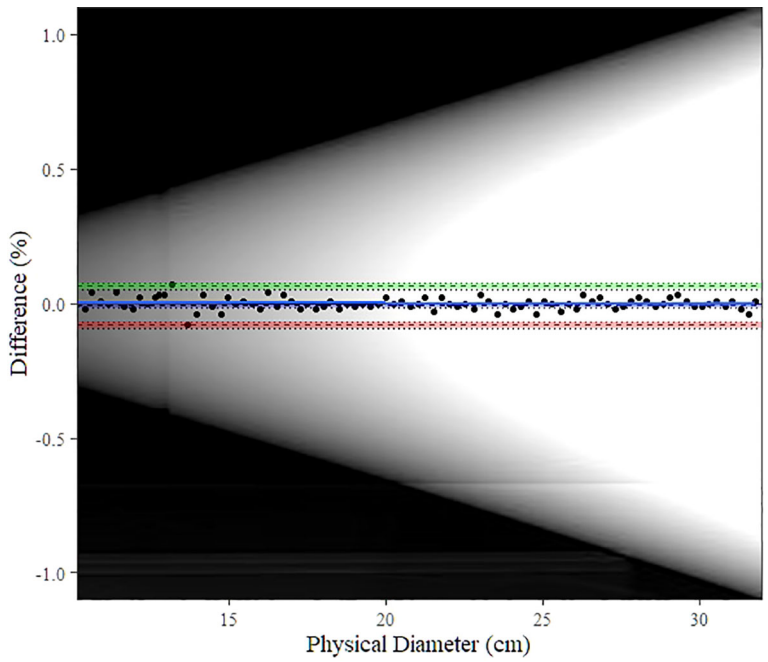


FIGURE 5
Bland–Altman plot for automatic D_W computation vs. physical (actual) diameter. The blue line shows the bias estimate with confidence interval (CI) as a blue band. Upper and lower limits of agreement (LoA) are depicted as dashed lines with colored bands as CI (upper LoA's CI: green, lower LoA's CI: red). The background image is the scout image of our water-filled funnel phantom.

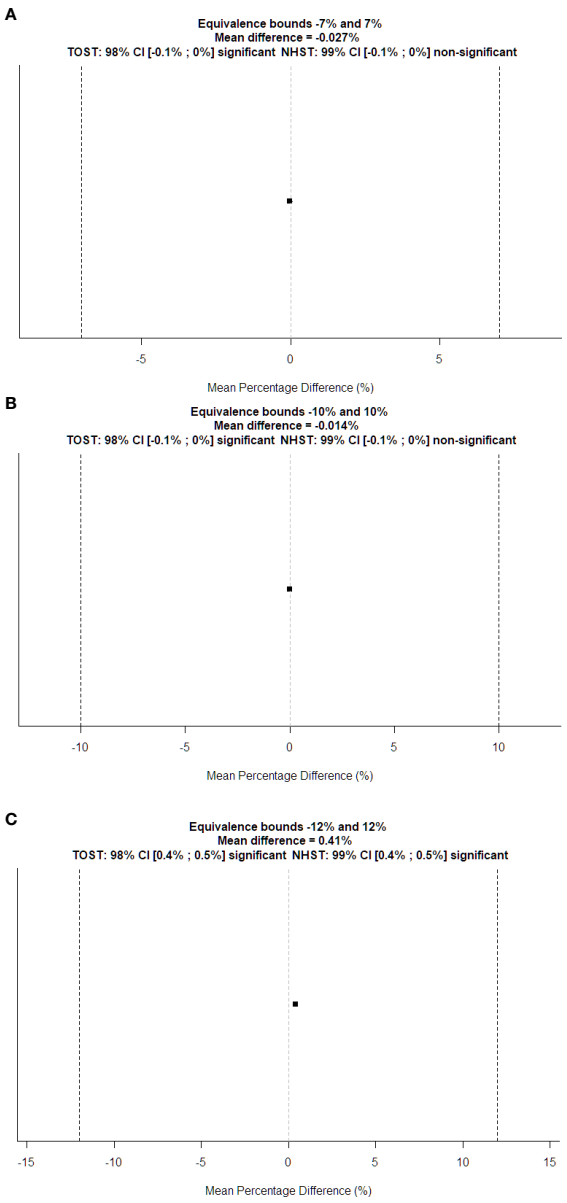


FIGURE 6
TOST plots for equivalence tests between automatically computed D_W vs. **(A)** physical diameter and **(B)** manually computed D_W of the water-filled funnel phantom; **(C)** manually computed D_W of the patient's CT series. Dotted vertical lines represent the boundary of equivalence. The squared dot is the mean percentage difference, and the small line through it shows the confidence interval for the standard *t*-test.

equivalence bounds with a TOST confidence interval equal to [0.358%, 0.462%].

The mean percentage of truncation is 3.58% [3.27%, 3.91%]. Those (small) percentages show that the field of view encompassed nearly the whole-body contour.

Figure 7 shows the graphs of SSDE and $CTDI_{vol}$ vs. slice number. Figure 8 displays the plot of D_W vs. slice number. We created these plots using the package "ggplot2" (34). This type of graph shows the values of D_W and/or SSDE and CTDI along the slice number, i.e., showing the values along the patient's anatomy.

Time for processing

Figure 9 shows histograms of time for processing of a single CT slice and a patient study. The mean time to process a single image is 13.99 ms [13.69 ms, 14.30 ms], and the mean time to process an entire study is 11.5 s [10.62 s, 12.63 s]. Figure 9 shows histograms of the times elapsed during the processing. Table 2 reports the summary of the distribution of times to process a whole CT study.

Discussion

We developed a software for real-time computation of size-specific dosimetric quantities: D_W and SSDE. We also evaluated the accuracy of the computation and the usability in a clinical workflow, in terms of time for processing single images and a whole patient study.

The mean difference in the D_W of the water-filled funnel phantom between the automatically computed and the actual diameter of the funnel phantom is small (−0.027%), as well as the difference between automatically and manually computed D_W on that phantom (−0.014%). Furthermore, the mean difference between the automatic computation of the D_W and the manually computed value of a patient's data is also small (0.41%). TOST analysis revealed that the D_W values of the water-filled funnel phantom computed with our algorithm are statistically equivalent to actual water-phantom diameters and to manually computed D_W values slice by slice, as prescribed by the IEC 62985:2019 norm. Bland–Altman analysis corroborates those findings, giving a smaller value for the LoA (<0.1%) than the one the IEC requires.

The processing times (see Table 2) are compatible with a clinical workflow. In particular, the maximum time for processing a study is about 1.5 min (99 s) for 4,460 slices. Moreover, 75% (third quartile) of the studies required a processing time below 16 s. The mean time to process a single study is 11.52 s (the mean number of slices is 789).

These results can lead to the building of a real-time D_W and SSDE and D_W computation system, allowing users (radiologists, radiographers, and medical physicists mainly) to obtain size-specific dosimetric information in short times after the very first scan of a patient. This should permit checking if the protocol used on the patient is tailored to its build or needs optimization.

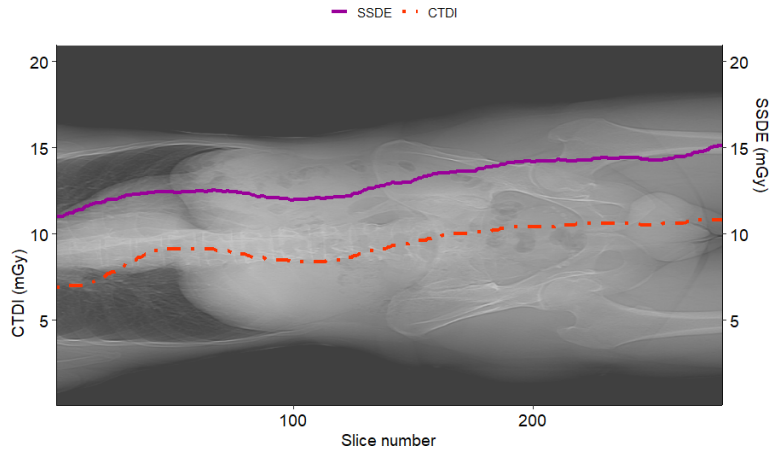


FIGURE 7
Example of SSDE and $CTDI_{vol}$ plot vs. slice number for a sample patient. The solid curve represents the SSDE computed for each slice, and the dotted one is the $CTDI_{vol}$ read from the DICOM header of the slices. It is worth noting that SSDE is different from $CTDI_{vol}$.

Having decoupled the computation from the DICOM storage, we are sure that our architecture does not interfere with the existing DICOM network environment.

The accuracy achieved in the SSDE and D_W computations and the very small time for processing an entire CT study allow us to compute the water-equivalent diameter averaged over the slice D_{w_ave} (17), a quantity favorable for protocol optimization (17). Based on the authors' knowledge, there are no previous studies showing such small computational times allowing for clinical real-time computation of size-specific dosimetric quantities.

We computed D_W and corrected it for truncated images, and the data we saved contain the percentage of truncation, so the final user can relate the accuracy of D_W with the presence of an amount of truncation.

This study has, however, some limitations. First, we computed times without accounting for time latencies in DICOM networks and/or for writing on files. Further studies should investigate these issues, and they should use a database instead of a text file. Second, our software relies massively on the Italian-named CT Protocol Names and thus is not immediately expandable to other clinical realities. Furthermore, we have not tested it on all CT vendors and all the protocols available for CT. Moreover, we applied our algorithm only to CT studies of the head, thorax, and abdomen/pelvis regions. Further studies require testing (or changing/adapting) the algorithm on

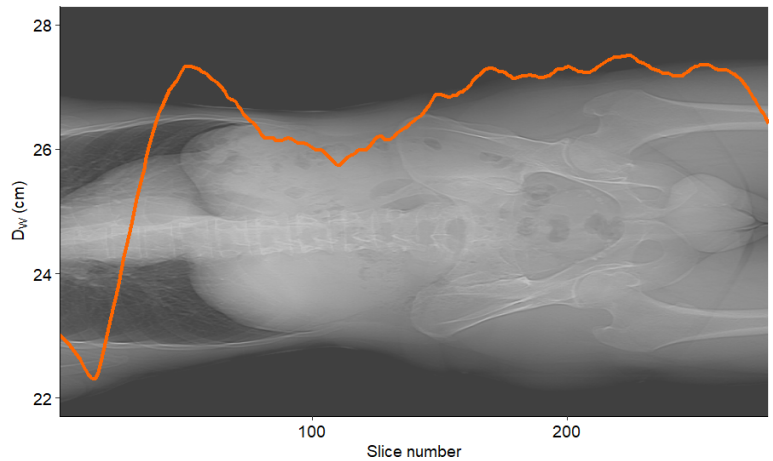


FIGURE 8
Example of D_W vs. slice number for a sample patient. The curve represents the water-equivalent diameter D_W computed for each slice.

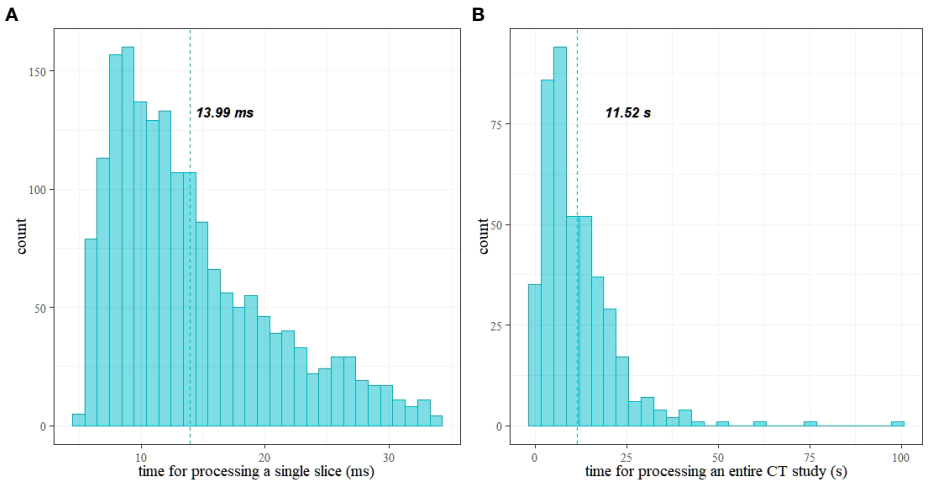


FIGURE 9
Histograms of time for processing: **(A)** a single CT slice and **(B)** a patient study; vertical lines show the mean values.

TABLE 2 Summary of time (in seconds) required to process a whole CT study.

Min.	1st Quartile	Median	Mean	3rd Quartile	Max
0.16	4.57	8.53	11.52	15.56	99.2

anatomical districts with multiple body regions, such as extremities, wrist, and shoulders.

Conclusion

A fully automated method of D_W measurement in a clinical workflow has been successfully developed. The system shows that it is possible to have highly accurate D_W and SSDE in almost real time, without affecting the clinical workflow of CT examinations.

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 below: https://github.com/massimilianoporzio/computeSSDE.

Ethics statement

The studies involving human participants were reviewed and approved by Comitato Etico Interaziendale dell'A.O. S. Croce e Carle di Cuneo e delle AA.SS.LL. CN1, CN2 e Asti. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

Author contributions

MP conceived the study design and developed the software. MP and CA performed data analysis, took part in discussions and preparation of the manuscript, and drafted the manuscript. All authors read, discussed, and approved the final manuscript.

Funding

This work was funded by the World Class Research University (WCRU), Diponegoro University, No. 118-08/UN7.6.1/PP/2021.

Acknowledgments

We would like to thank ASL CN1 for the effort in approving of this study in such a brief time.

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

References

1. Alaei P, Spezi E. Commissioning kilovoltage cone-beam CT beams in a radiation therapy treatment planning system. *J Appl Clin Med Phys* (2012) 13 (6):19–33. doi: 10.1120/jacmp.v13i6.3971

2. Brink JA, Amis ES. Image wisely: A campaign to increase awareness about adult radiation protection. *Radiology* (2010) 257(3):601–2. doi: 10.1148/radiol.10101335

3. Lange I, Alikhani B, Wacker F, Raatschen HJ. Intraindividual variation of dose parameters in oncologic CT imaging. *PloS One* (2021) 16(4):e0250490. doi: 10.1371/journal.pone.0250490

4. Kwee TC, Dijkstra H, Knapen DG, de Vries EGE, Yakar D. Which patients are prone to undergo disproportionate recurrent CT imaging and should we worry? *Eur J Radiol* (2020) 125:108898. doi: 10.1016/j.ejrad.2020.108898

5. Chow JCL, Leung MKK, Islam MK, Norrlinger BD, Jaffray DA. Evaluation of the effect of patient dose from cone beam computed tomography on prostate IMRT using Monte Carlo simulation: Patient dose from CBCT on prostate IMRT. *Med Phys* (2007) 35(1):52–60. doi: 10.1118/1.2815625

6. Mututantri-Bastiyange D, Chow JCL. Imaging dose of cone-beam computed tomography in nanoparticle-enhanced image-guided radiotherapy: A Monte Carlo phantom study. *AIMS Bioengin* (2020) 7(1):1–11. doi: 10.3934/bioeng.2020001

7. Frija G, Damilakis J, Paulo G, Loose R, Vano EEuropean Society of Radiology (ESR). Cumulative effective dose from recurrent CT examinations in Europe: proposal for clinical guidance based on an ESR EuroSafe imaging survey. *Eur Radiol* (2021) 31(8):5514–23. doi: 10.1007/s00330-021-07696-1

8. Huda W, Mettler FA. Volume CT dose index and dose-length product displayed during CT: what good are they? *Radiology* (2011) 258(1):236–42. doi: 10.1148/radiol.10100297

9. McCollough CH, Leng S, Yu L, Cody DD, Boone JM, McNitt-Gray MF, et al. And patient dose: They are not the same thing. *Radiology* (2011) 259(2):311–6. doi: 10.1148/radiol.11101800

10. Brady SL, Kaufman RA. Investigation of American association of physicists in medicine report 204 size-specific dose estimates for pediatric CT implementation. *Radiology* (2012) 265(3):832–40. doi: 10.1148/radiol.12120131

11. Boone J, Strauss K, Cody D, McCollough C, McNitt-Gray M, Toth T, et al. *Size-specific dose estimates (SSDE) in pediatric and adult body CT examinations [Internet]*. College Park, MD: AAPM (2011). Available at: https://www.aapm.org/pubs/reports/detail.asp?docid=143.

12. McCollough C, Bakalyar D, Bostani M, Brady S, Boedeker K, Boone J, et al. *Use of water equivalent diameter for calculating patient size and size-specific dose estimates (SSDE) in CT [Internet]*. College Park, MD: AAPM (2014). Available at: https://www.aapm.org/pubs/reports/detail.asp?docid=146.

13. Mihailidis D, Tsapaki V, Tomara P. A simple manual method to estimate water-equivalent diameter for calculating size-specific dose estimate in chest computed tomography. *Br J Radiol* (2021) 94(1117):20200473. doi: 10.1259/bjr.20200473

14. Anam C, Haryanto F, Widita R, Arif I, Dougherty G. Automated calculation of water-equivalent diameter (DW) based on AAPM task group 220. *J Appl Clin Med Phys* (2016) 17(4):320–33. doi: 10.1120/jacmp.v17i4.6171

15. Anam C, Mahdani FR, Dewi WK, Sutanto H, Triadyaksa P, Haryanto F, et al. An improved method for automated calculation of the water-equivalent diameter for estimating size-specific dose in CT. *J Appl Clin Med Phys* (2021) 22 (9):313–23. doi: 10.1002/acm2.13367

16. Leng S, Shiung M, Duan X, Yu L, Zhang Y, McCollough CH. Size-specific dose estimates for chest, abdominal, and pelvic CT: Effect of intrapatient variability in water-equivalent diameter. *Radiology* (2015) 276(1):184–90. doi: 10.1148/radiol.15142160

17. Sarmento S, Mendes B, Gouvêa M.Automatic calculation of patient size metrics in computed tomography: What level of computational accuracy do we need? *J Appl Clin Med Phys* (2018) 19(1):218–27. doi: 10.1002/acm2.12240

18. Schuirmann DJ. A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *J Pharmacokinet Biopharm* (1987) 15(6):657–80. doi: 10.1007/BF01068419

19. Ozsoykal I, Yurt A, Akgungor K. Size-specific dose estimates in chest, abdomen, and pelvis CT examinations of pediatric patients. *Diagn Interv Radiol* (2018) 24(4):243–8. doi: 10.5152/dir.2018.17450

20. Gharbi S, Labidi S, Mars M. Automatic brain dose estimation in computed tomography using patient dicom images. *Radiat Prot Dosimet* (2020) 188(4):536–42. doi: 10.1093/rpd/ncaa006

21. Juszczyk J, Badura P, Czajkowska J, Wijata A, Andrzejewski J, Bozek P, et al. Automated size-specific dose estimates using deep learning image processing. *Med Imag Anal* (2021) 68:101898. doi: 10.1016/j.media.2020.101898

22. Larson DB. Optimizing CT radiation dose based on patient size and image quality: the size-specific dose estimate method. *Pediatr Radiol* (2014) 44(S3):501–5. doi: 10.1007/s00247-014-3077-y

23. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* (2012) 9(7):671–5. doi: 10.1038/nmeth.2089

24. *Analyze particles. [ImageJ documentation wiki] [Internet]*. Available at: https://imagejdocu.tudor.lu/gui/analyze/analyze_particles.

25. Anam C, Haryanto F, Widita R, Arif I, Dougherty G. The size-specific dose estimate (ssde) for truncated computed tomography images. *Radiat Prot Dosimet* (2017) 175(3):313–20. doi: 10.1093/rpd/ncw326

26. Jodogne S. The orthanc ecosystem for medical imaging. *J Digit Imag* (2018) 31(3):341–52. doi: 10.1007/s10278-018-0082-y

27. Ierusalimschy R, de Figueiredo LH, Celes W. *Lua 5.1 reference manual*. Rio de Janeiro (2006). 103.

28. Ionescu VM. (2015). The analysis of the performance of RabbitMQ and ActiveMQ, in: *2015 14th RoEduNet International Conference - Networking in Education and Research (RoEduNet NER)*, 132–7.

29. Becci D, Gentile L, Sartor W. Dosimetric and qualitative evaluation of CT longitudinal automatic tube current modulation using in-house phantoms. *Phys Med* (2016) 32:72. doi: 10.1016/j.ejmp.2016.01.245

30. R Core Team. *R: A language and environment for statistical computing [Internet]*. Vienna, Austria: R Foundation for Statistical Computing (2021). Available at: https://www.R-project.org/.

31. Lakens D. Equivalence tests: A practical primer for t tests, correlations, and meta-analyses. *Soc psychol Pers Sci* (2017) 8(4):355–62. doi: 10.1177/1948550617697177

32. Giavarina D. Understanding bland altman analysis. *Biochem Med (Zagreb)* (2015) 25(2):141–51. doi: 10.11613/BM.2015.015

33. Helwig NE. *Nptest: Nonparametric bootstrap and permutation tests [Internet]* (2021). Available at: https://CRAN.R-project.org/package=nptest.

34. Wickham H. *ggplot2: Elegant graphics for data analysis [Internet]*. New York: Springer-Verlag (2016). Available at: https://ggplot2.tidyverse.org.

35. *massimilianoporzio/computeSSDE [Internet]* (2021). Available at: https://github.com/massimilianoporzio/computeSSDE.

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