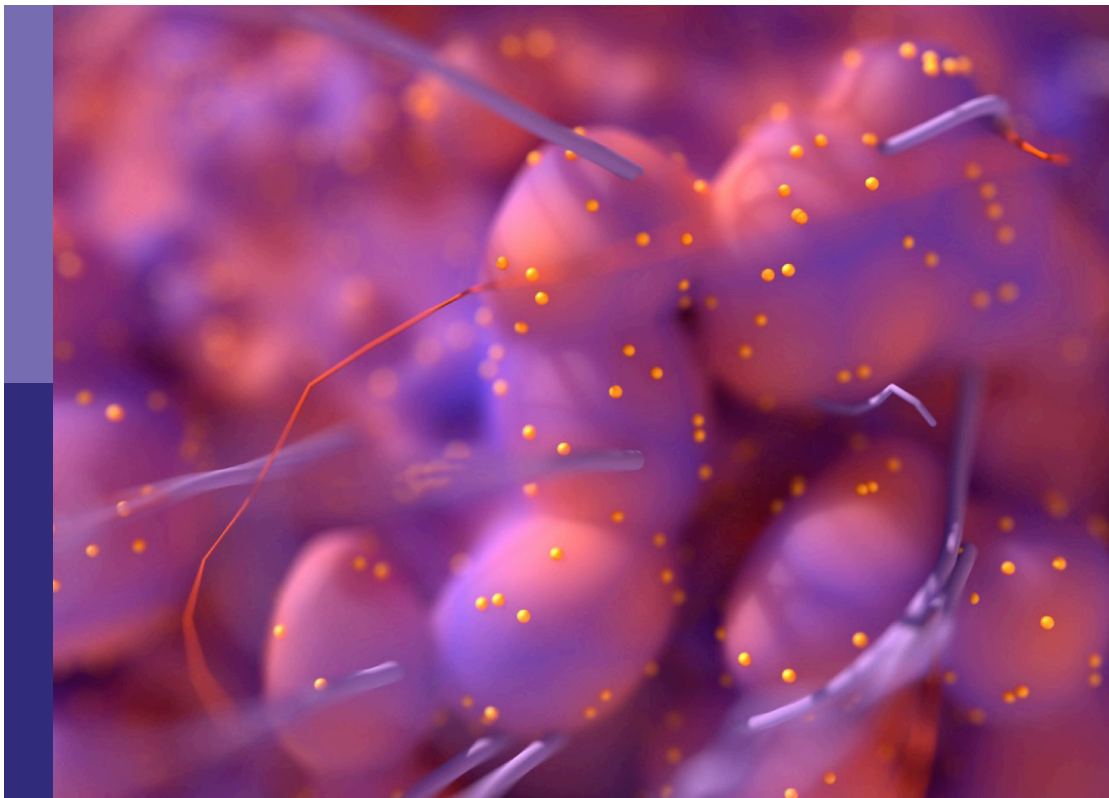
frontiers | Research Topics

Methods in cancer imaging and image-directed interventions

Edited by
Bahram Mohajer and Jasper Nijkamp

Published in
Frontiers in Oncology



ISSN 1664-8714
ISBN 978-2-8325-3835-7
DOI 10.3389/978-2-8325-3835-7

About Frontiers

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

Frontiers journal series

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

Dedication to quality

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

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Methods in cancer imaging and image-directed interventions

Topic editors

Bahram Mohajer — Department of Radiology and Radiological Science, School of Medicine, Johns Hopkins Medicine, United States

Jasper Nijkamp — Aarhus University, Denmark

Citation

Mohajer, B., Nijkamp, J., eds. (2023). *Methods in cancer imaging and image-directed interventions*. Lausanne: Frontiers Media SA.

doi: 10.3389/978-2-8325-3835-7

Table of contents

05 Editorial: Methods in cancer imaging and image-directed interventions
Jasper Nijkamp and Bahram Mohajer

08 Multiparametric MRI-based nomograms in predicting positive surgical margins of prostate cancer after laparoscopic radical prostatectomy
Shuang Meng, Lihua Chen, Qinhe Zhang, Nan Wang and Ailian Liu

21 Frontiers and hotspots of ^{18}F-FDG PET/CT radiomics: A bibliometric analysis of the published literature
Xinghai Liu, Xianwen Hu, Xiao Yu, Pujiao Li, Cheng Gu, Guosheng Liu, Yan Wu, Dandan Li, Pan Wang and Jiong Cai

38 Novel small ^{99m}Tc-labeled affibody molecular probe for PD-L1 receptor imaging
Zhigang Liang, Xianwen Hu, Hongyu Hu, Pan Wang and Jiong Cai

48 Visualising spatial heterogeneity in glioblastoma using imaging habitats
Mueez Waqar, Petra J. Van Houdt, Eline Hessen, Ka-Loh Li, Xiaoping Zhu, Alan Jackson, Mudassar Iqbal, James O'Connor, Ibrahim Djoukhadar, Uulke A. van der Heide, David J. Coope and Gerben R. Borst

65 Applications of T_1 and T_2 relaxation time calculation in tissue differentiation and cancer diagnostics—a systematic literature review
Marta Micek, David Aebisher, Justyna Surówka, Dorota Bartusik-Aebisher and Michał Madera

83 Diagnostic value of [^{68}Ga]Ga-FAPI-04 in patients with colorectal cancer in comparison with [^{18}F]F-FDG PET/CT
Xinfeng Lin, Yingjie Li, Shuailiang Wang, Yan Zhang, Xuetao Chen, Maomao Wei, Hua Zhu, Aiwen Wu, Zhi Yang and Xuejuan Wang

96 Magnetic resonance imaging radiomics to differentiate ovarian sex cord-stromal tumors and primary epithelial ovarian cancers
Meiying Cheng, Shifang Tan, Tian Ren, Zitao Zhu, Kaiyu Wang, Lingjie Zhang, Lingsong Meng, Xuhong Yang, Teng Pan, Zhexuan Yang and Xin Zhao

106 Dose accumulation for MR-guided adaptive radiotherapy: From practical considerations to state-of-the-art clinical implementation
Brigid A. McDonald, Cornel Zachiu, John Christodouleas, Mohamed A. Naser, Mark Ruschin, Jan-Jakob Sonke, Daniela Thorwarth, Daniel Létourneau, Neelam Tyagi, Tony Tadic, Jinzhong Yang, X. Allen Li, Uffe Bernchou, Daniel E. Hyer, Jeffrey E. Snyder, Edyta Bubula-Rehm, Clifton D. Fuller and Kristy K. Brock

120 A CT-based radiomics approach to predict immediate response of radiofrequency ablation in colorectal cancer lung metastases
Haozhe Huang, Dezhong Zheng, Hong Chen, Chao Chen, Ying Wang, Lichao Xu, Yaohui Wang, Xinhong He, Yuanyuan Yang and Wentao Li

134 Lesion-specific 3D-printed moulds for image-guided tissue multi-sampling of ovarian tumours: A prospective pilot study
Maria Delgado-Ortet, Marika A. V. Reinius, Cathal McCague, Vlad Bura, Ramona Woitek, Leonardo Rundo, Andrew B. Gill, Marcel Gehrung, Stephan Ursprung, Helen Bolton, Krishnayan Haldar, Pubudu Pathiraja, James D. Brenton, Mireia Crispin-Ortuzar, Mercedes Jimenez-Linan, Lorena Escudero Sanchez and Evis Sala

149 Correlation between ADC, ADC ratio, and Gleason Grade group in prostate cancer patients undergoing radical prostatectomy: Retrospective multicenter study with different MRI scanners
Johan Bengtsson, Erik Thimansson, Erik Baubeta, Sophia Zackrisson, Pia Charlotte Sundgren, Anders Bjartell and Despina Flondell-Sité

160 Segmentation stability of human head and neck cancer medical images for radiotherapy applications under de-identification conditions: Benchmarking data sharing and artificial intelligence use-cases
Jaakko Sahlsten, Kareem A. Wahid, Enrico Glerean, Joel Jaskari, Mohamed A. Naser, Renjie He, Benjamin H. Kann, Antti Mäkitie, Clifton D. Fuller and Kimmo Kaski

170 Intrafraction tumor motion monitoring and dose reconstruction for liver pencil beam scanning proton therapy
Saber Nankali, Esben Schjødt Worm, Jakob Borup Thomsen, Line Bjerregaard Stick, Jenny Bertholet, Morten Høyer, Britta Weber, Hanna Rahbek Mortensen and Per Rugaard Poulsen

181 Breaking down the RECIST 1.1 double read variability in lung trials: What do baseline assessments tell us?
Antoine Iannessi and Hubert Beaumont

195 The synergized diagnostic value of VTQ with chemokine CXCL13 in lung tumors
Xu Zhang, Yejian Lu, Kenan Huang, Qingfang Pan, Youchao Jia, Baoshuan Cui, Peipei Yin, Jianhui Li, Junping Ju, Xiangyu Fan and Rui Tian

204 A novel approach for automatic segmentation of prostate and its lesion regions on magnetic resonance imaging
Huipeng Ren, Chengjuan Ren, Ziyu Guo, Guangnan Zhang, Xiaohui Luo, Zhuanqin Ren, Hongzhe Tian, Wei Li, Hao Yuan, Lele Hao, Jiacheng Wang and Ming Zhang

frontiers | Frontiers in Oncology

TYPE Editorial
PUBLISHED 17 October 2023
DOI 10.3389/fonc.2023.1293020

OPEN ACCESS

EDITED AND REVIEWED BY
Lizhi Liu,
Sun Yat-sen University Cancer Center (SYSUCC), China

*CORRESPONDENCE
Jasper Nijkamp
jaspernijkamp@clin.au.dk

RECEIVED 12 September 2023
ACCEPTED 10 October 2023
PUBLISHED 17 October 2023

CITATION
Nijkamp J and Mohajer B (2023) Editorial: Methods in cancer imaging and image-directed interventions. *Front. Oncol.* 13:1293020. doi: 10.3389/fonc.2023.1293020

Editorial: Methods in cancer imaging and image-directed interventions

Jasper Nijkamp[1,2]* and Bahram Mohajer[3]

[1]Department of Clinical Medicine, Aarhus University, Aarhus, Denmark, [2]Danish Center for Particle Therapy, Aarhus University Hospital, Aarhus, Denmark, [3]Russell H. Morgan Department of Radiology and Radiological Science, Johns Hopkins University School of Medicine, Baltimore, MD, United States

KEYWORDS

methods, cancer imaging, image-directed interventions, image-guidance, radiotherapy, radiology, radiomics

Editorial on the Research Topic
Methods in cancer imaging and image-directed interventions

Introduction

The methods we employ to visualize, analyze, and intervene in cancerous tissues rapidly evolve. These methods, which include various imaging modalities and image-guided therapies, play a vital role in the accurate diagnosis, precise treatment planning, and targeted interventions for cancer patients.

We are glad to see a total of 16 high-quality papers published in this Research Topic, of which twelve original articles, three reviews, and one methods paper, from various countries and regions. As can be expected from the broad topic, a wide range of papers is presented.

Molecular, magnetic resonance, and ultrasound imaging

For molecular imaging, Liang et al. presented a novel molecular probe for imaging programmed death ligand 1 (PD-L1) with single-photon emission computed tomography (SPECT). Their PD-L1 targeting affibody with ^{99m}Tc labelling showed high *in vivo* affinity and quick blood clearance.

Lin et al. performed a prospective study comparing ^{18}FDG-PET/CT with ^{68}Ga-FAPI-04 PET/CT on colorectal cancer staging. Differences in staging between both scans were investigated in 61 patients. FAPI-PET led to TNM upstaging in 10, and downstaging in 5 patients, which consecutively altered the treatment course in 13 patients, primarily due to better nodal and metastasis detection.

Three studies assessed use of MRI in tumor grading and tissue characterization. Apparent diffusion coefficient (ADC) is often suggested as biomarker of prostate tumor aggressiveness. Bengtsson et al. addressed the challenges in generalizing ADC cut-off values between scanners and centers with regard to tumor aggressiveness. They showed that ADC

(and its ratios) did not correlate with tumor Gleason Grade, and should therefore be used with caution as biomarker.

Secondly, Waqar et al. focused their narrative review on current status of habitat MR-imaging in glioblastoma. With habitat imaging, tumors can be subdivided into regions that express similar imaging characteristics, assuming that these regions represent different tumor biology. They provide an overview of the current work, challenges of producing relevant and reproducible habitats, and clinical applications for personalized treatment.

Finally, Micek et al. reviewed the current applications of non-contrast-enhanced quantitative MRI parameters, specifically T1 and T2 relaxation time, for tissue characterization. They focused on epithelium in breast cancer, but also reviewed studies on lung, prostate, skin, kidney, and liver malignancies. They concluded that in most studies only T1 or T2 relaxation times were reported, resulting in limited groundwork to the value of combining the two in differentiating tissues.

In another study, Zhang et al. show the diagnostic value of virtual touch tissue quantification (VTQ) ultrasonography in combination with pleural effusion chemokine expression quantification (CXCL13) for detection of malignant lung nodules. They show a synergistic diagnostic potential of CXCL13 and VTQ for differentiating malignant and benign pleural effusions.

Image segmentation and validation

Ren et al. provide a novel prostate and prostate lesion segmentation model, based on UNet, but using dense blocks, a convolution block attention module, and group norm-Atrous spatial pyramidal pooling. Their input data was DWI imaging only, and they show that their approach outperforms other well-known architectures.

Sahlsten et al. investigated the effect of defacing CT and MRI scans on segmentation accuracy of organs at risk in the head and neck area. Defacing imaging data without affecting the segmentation accuracy could help to overcome challenges in data sharing. Of the eight available defacing models, only 3 performed acceptably on MRI, none did well on CT. Segmentation accuracy was influenced by the defacing tools, and improvements are needed to assure safe data sharing.

With an innovative design and with the aim of correlating imaging findings with histology in ovarian tumors, Delgado-Ortet et al. presented a method for designing and 3D printing molds. The work is inspired by similar workflows presented in imaging validation of prostate cancer but applied in the challenging setting of ovarian tumors.

Response assessment, radiomics, and prediction models

Iannessi and Beaumont evaluated the effect of observer variation in standardized response assessment (RECIST) in clinical trials of patient with lung cancer with blinded independent central review. They show that even with standardization, there are substantial variabilities at baseline and provide evidence of observer dependent reading patterns.

In the field of radiomics, Liu et al. provide an bibliometric overview of the research that is being conducted in the area of FDG-PET/CT radiomics. They included 361 original articles and 96 reviews, and provide which countries, institutes, and individuals are the drivers of the research in this area.

In a diagnostic application, Cheng et al. developed an MRI radiomics model to differentiate between ovarian sex cord-stromal tumors and epithelial ovarian cancers. They showed best performance when combining clinical characteristics, conventional MRI parameters (e.g. ADC), and a radiomics score based on 9 features. For ease of use, they also developed a nomogram.

Huang et al. aimed to predict response to radiofrequency ablation (RFA) of colorectal lung metastasis using radiomics. Their model contained combinations of CT-radiomics features of the pre-RFA metastasis and the ablation zone, plus cancer antigen 19-9 and location of the metastasis as clinical features. With an accuracy of 82.6%, this work shows potential to guide personalized follow-up in these patients.

In another study, using multiparametric (mp)MRI, without radiomics, Meng et al. provided an approach to predict the surgical resection margins after laparoscopic prostatectomy. This approach could help choosing optimal treatment for prostate cancer patients. They presented nomograms based on clinical parameters, mpMRI, and a combination, with the latter as most accurate (AUC=0.756).

Radiotherapy

In the field of radiotherapy, McDonald et al. provided a comprehensive review of dose accumulation for MR-guided adaptive radiotherapy. They cover the whole range from practical considerations to state-of-the-art clinical implementation.

Finally, Nankali et al. describe a method to monitor intrafraction tumor motion and accumulate the fraction dose during liver pencil beam scanning proton therapy. The illustration of the method on three patient cases is impressive, but the pathway to broad clinical implementation is challenged by integration needs.

This Research Topic provides a cross-sectional snapshot of a rapidly evolving field. We are excited to observe where the field progresses in coming years.

Author contributions

JN: Writing – original draft, Writing – review & editing. BM: Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

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

Publisher's note

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 12 September 2022
DOI 10.3389/fonc.2022.973285

OPEN ACCESS

EDITED BY
Jasper Nijkamp,
Aarhus University, Denmark

REVIEWED BY
Stefano Luzzago,
European Institute of Oncology (IEO), Italy
Ugo Giovanni Falagario,
University of Foggia, Italy

*CORRESPONDENCE
Ailian Liu
liuailian@dmu.edu.cn

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

RECEIVED 20 June 2022
ACCEPTED 11 August 2022
PUBLISHED 12 September 2022

CITATION
Meng S, Chen L, Zhang Q, Wang N and Liu A (2022) Multiparametric MRI-based nomograms in predicting positive surgical margins of prostate cancer after laparoscopic radical prostatectomy. *Front. Oncol.* 12:973285. doi: 10.3389/fonc.2022.973285

Multiparametric MRI-based nomograms in predicting positive surgical margins of prostate cancer after laparoscopic radical prostatectomy

Shuang Meng, Lihua Chen, Qinhe Zhang, Nan Wang and Ailian Liu*

Department of Radiological, First Affiliated Hospital of Dalian Medical University, Dalian, China

Background: Positive surgical margins (PSMs) are an independent risk factor of biochemical recurrence in patients with prostate cancer (PCa) after laparoscopic radical prostatectomy; however, limited MRI-based predictive tools are available. This study aimed to develop a novel nomogram combining clinical and multiparametric MRI (mpMRI) parameters to reduce PSMs by improving surgical planning.

Methods: One hundred and three patients with PCa (55 patients with negative surgical margins [NSMs] and 48 patients with PSMs) were included in this retrospective study. The following parameters were obtained using GE Functool post-processing software: diffusion-weighted imaging (DWI); intravoxel incoherent motion model (IVIM); and diffusion kurtosis imaging (DKI). Patients were divided into different training sets and testing sets for different targets according to a ratio of 7:3. The least absolute shrinkage and selection operator (LASSO) regression algorithm was used to analyze the data set to select the optimal MRI predictors. Preoperatively clinical parameters used to build a clinical nomogram (C-nomogram). Multivariable logistic regression analysis was used to build an MRI nomogram (M-nomogram) by introducing the MRI parameters. Based on the MRI and clinical parameters, build an MRI combined with clinical parameters nomogram (MC-nomogram). Comparisons with the M-nomogram and MC-nomogram were based on discrimination, calibration, and decision curve analysis (DCA). A 3-fold cross-validation method was used to assess the stability of the nomogram.

Results: There was no statistical difference in AUC between the C-nomogram (sensitivity=64%, specificity=65% and AUC=0.683), the M-nomogram (sensitivity=57%, specificity=88% and AUC=0.735) and the MC-nomogram (sensitivity= 64%, specificity=82% and AUC=0.756). The calibration curves of the three nomograms used to predict the risk of PSMs

in patients with PCa showed good agreement. The net benefit of the MC-nomogram was higher than the others (range, 0.2-0.7).

Conclusions: The mpMRI-based nomogram can predict PSMs in PCa patients. Although its AUC (0.735) is not statistically different from that of the clinical-based nomogram AUC (0.683). However, mpMRI-based nomogram has higher specificity (88% VS. 63%), model stability, and clinical benefit than clinical-based nomogram. And the predictive ability of mpMRI plus clinical parameters for PSMs is further improved.

KEYWORDS

prostate cancer, positive surgical margins, MRI, nomogram, laparoscopic radical prostatectomy

Introduction

Laparoscopic radical prostatectomy (LRP) has been widely used in clinical practice and is currently the main way to treat localized prostate cancer (PCa) (1). Of note, 29.1%-34% of patients who undergo LRP have positive surgical margins (PSMs) (2–4), which is an independent risk factor for biochemical recurrence (BCR) in patients after prostatectomy (5, 6). And the results of a cohort study showed that PSMs poses a substantial financial burden (7). Therefore, it is necessary to predict PSMs so that optimal treatment strategies can be implemented.

In previous studies it was reported that there are some preoperative parameters correlate with PSMs after prostatectomy, including age, clinical stage, free prostate specific antigen (FPSA)/total PSA (TPSA), Gleason score, percent of positive cores (PPC), and extra-prostatic extension (3, 8). To predict the risk of advanced PCa, clinicians also often use staging nomograms, such as D'AMICO or CAPRA (9, 10). However, most studies only included the clinical characteristics, and a lack of knowledge about the predictive value of multiparametric MRI (mpMRI).

An mpMRI is considered a common examination for the diagnosis of PCa; specifically, reduces false-negative biopsies (11) and identify risk factors associated with PSMs (12). An mpMRI mainly consists of T2-weighted imaging (T2WI), diffusion-weighted imaging (DWI), and dynamic contrast-enhanced MRI. DWI reflects the PCa histopathologic tissue composition (13) and has the potential to predict abnormal pathologic features after prostatectomy. Over the last two decades, several advanced DWI models have been developed to improve assessment of PCa, including (14). No systematic research studies have shown if advanced DWI models can help clinicians predict PSMs after LRP are lacking. Therefore, the aim of the present study was to develop a novel nomogram combining clinical and mpMRI parameters to predict PSMs after LRP to guide decision-making.

Materials and methods

Study population

This single-center, retrospective study included 1055 consecutive male inpatients who underwent pelvic mpMRI between January 2016 and November 2021.

Participants who met the following inclusion criteria were included in our study: ① confirmed diagnosis of PCa by systemic biopsies; ② patients with complete clinical data, including age, Prostate Imaging Reporting and Data System version 2 (PI-RADS v2) score, TPSA, biopsy-based Gleason score, PPC, clinical TNM (cTNM), postoperative Gleason score, pathologic TNM (pTNM), and PSMs locations; and ③ LRP performed on PCa patients by urologists who have performed 500 radical prostatectomies within 3 months after MRI and systemic biopsies. Urologists recommend systematic biopsy of patients before surgery based on PI-RADS and PSA levels. And referring to the mpMRI report to formulate a surgical plan.

Of the participants, those who met at least one of the following criteria were excluded: ① history of PCa treatment; ② incomplete MRI sequences; ③ prostate lesions with poorly-defined boundaries on T2WI and apparent diffusion coefficient (ADC) images, according to PI-RADS v2 (15).

PSMs were defined by cancer cells involving the inked surface of the specimen (16) and divided into negative surgical margins (NSMs) and PSMs groups according to marginal status. One hundred three patients with PCa (55 patients with NSMs and 48 patients with PSMs) were included in our study (Figure 1).

The study was approved by the Institutional Review Board of First Affiliated Hospital of Dalian Medical University, informed consent was waived.

MRI protocols

A 3.0 T MRI scanner (GE-Signa HDXT; GE, Milwaukee, WI, USA) with an eight-channel phased-array body-coil was used in this study. MRI scan was done before the biopsy. The MRI scanning protocol included T1-weighted imaging (T1WI), T2WI, DWI, intravoxel incoherent motion model (IVIM), and diffusion kurtosis imaging (DKI). Sagittal and axial high-definition T2WI sequences were used for prostate tumor location. DWI was performed with high *b* values (up to a maximum of 1000 s/mm^2). IVIM with *b* values of 0, 20, 50, 100, 150, 200, 400, 800, 1200, 2000, and 3000 s/mm^2 was performed with a single-shot echo planar (SS-EPI) sequence. DKI with *b* values of 0 and 1500 s/mm^2 were performed in the oblique axial plane using a SS-EPI sequence with comparable parameters. The diffusion gradients were applied simultaneously along with 15 orthogonal directions. The MRI scanning parameters (DWI, IVIM, DKI, and T2WI) are shown in Table 1, supporting information. The images were transferred to an AW 4.4 workstation (GE Healthcare) and reconstructed using GE Functool post-processing software.

For analysis of images obtained with DWI, parameter maps were generated by fitting the following models to the pixel signal intensities at the different b values, as follows.

For the mono-exponential DWI model (17),

$$S_b/S_0 = exp(-b \times ADC)$$

where S_b is the mean signal intensity with diffusion gradient b, S_0 is the mean signal intensity without a diffusion gradient, the b value of ADC_{ME} is 0 and 1,000 s/mm^2, and the b value of ADC_{BE} is (0, 20, 50, 100, 150, 200, 400, 800, 1200, 2000, and 3000 s/mm^2).

For the bi-exponential DWI model (18),

$$S_b/S_0 = (1-f)exp(-bD) + f\,exp(-bD*)$$

where $S_b$represents the mean signal intensity with diffusion gradient b and S_0 is the mean signal intensity. When b = 0 s/mm^2, D (D_{mono}, D_{Bi}) is the true molecular diffusion coefficient. D* (D^*_{mono}, D^*_{Bi}) is the pseudodiffusion coefficient and *f* (f_{mono}, f_{Bi}) is the perfusion fraction.

For the stretched exponential DWI model (19),

$$S_b/S_0 = exp^{[-(b\cdot DDC)]\,\alpha}$$

where α represents an anomalous exponential term of the intravoxel water molecule ranging and DDC represents a mean intravoxel diffusion coefficient.

For the DKI model (20),

$$1n\,(S_b) = 1n\,S_0 - b \bullet D + 1/6 \bullet b^2 \bullet D^2 \bullet K$$

where S_b is the MR signal intensity at the particular b value used, S_0 is the MR signal intensity without a diffusion gradient, K (FAk, MK, Ka, Kr) is the apparent diffusion kurtosis, and D (FA, Da, Dr, MD) is the ADC revised for non-Gaussian behavior.

MRI measurements

MRI measurements were performed by two experienced radiologists (with 5 and 6 years of experience in abdominal

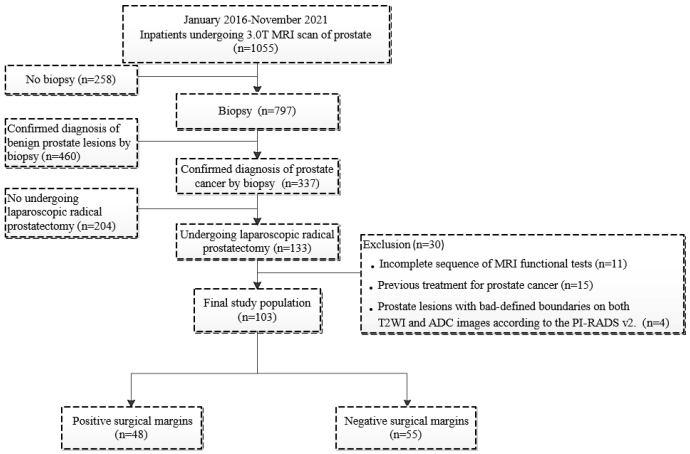


FIGURE 1
Flowchart of the patient population.

TABLE 1 MR sequences parameters of DWI, IVIM, DKI and T2WI.

Parameter	DWI	IVIM	DKI	Axial T2WI	Coronal T2WI	Sagittal T2WI
Pulse sequence name	EPI	EPI	EPI	FRFSE	FRFSE	FRFSE
TR / TE (ms)	4200 / 95	2800 / 90	2500 / 80	5140 / 139	2460 / 128	2660 / 118
Flip angle (°)	90	90	90	90	90	90
FOV (cm)	30 × 30	35 × 31	35 × 35	30 × 30	30 × 30	30 × 30
Voxel (mm)	1.2 × 1.2	1.4 × 1.4	1.4 × 1.4	0.6 × 0.6	0.6 × 0.6	0.6 × 0.6
Matrix	128 × 128	128 × 128	128 × 128	320 × 224	320 × 224	320 × 224
Slice/Thickness (mm)	4.0 / 1.0	7.0 / 1.0	7.0 / 1.0	4.0 / 1.0	5.0 / 1.0	4.0 / 1.0
ETL	–	–	–	23	18	19
Scan Duration(s)	109	151	178	161	101	101
NEX	8	2	2	4	4	4

radiology) using the double-blinding method. Prostate MRI interpretation was based on the PI-RADS v2 (15). The region of interest (ROI) was placed by observers in the slice of the largest prostate cancer lesion, and covered the entire lesion while avoiding obvious necrotic or fibrotic areas. The ROI locations on the IVIM and DKI pseudo-color maps were consistent with T2WI and ADC to the greatest extent possible (Figures 2 and 3).

Statistical analysis

All statistical analyses were performed using SPSS (version 25.0; IBM Corp., Armonk, NY, USA), MedCalc (version 15.2.2; Digimizer, Belgium), and R software (version 3.6.1; https://www.R-project.org).

The normality of the data was tested using the Shapiro-Wilk test. Normally distributed continuous variables are expressed as the means ± standard deviations. Non-normally distributed continuous variables are expressed as medians and ranges (25^{th} and 75^{th} percentiles). Nominal variables are expressed as frequencies with percentages.

The differences between two groups were analyzed using two-sided t-tests or the non-parametric Mann-Whitney U test for normally or non-normally distributed data for continuous variables and the Wilcoxon rank-sum test for categorical variables. The inter-observer agreement of the MRI measurements was analyzed by calculating the intraclass correlation coefficient (ICC).

The least absolute shrinkage and selection operator (LASSO) regression algorithm was used to analyze the data set to select the optimal predictors among the mpMRI quantitative parameters. Then, preoperatively clinical parameters used to build a clinical nomogram (C-nomogram), multivariable logistic regression analysis was used to build an MRI nomogram (M-nomogram) and MRI combined with the clinical nomogram (MC-monogram). Further, several kinds of validation methods were used to estimate the accuracy of the nomograms. The receiver operating characteristic (ROC) curve was used to evaluate the discrimination abilities. The area under the curve (AUC) was compared using the DeLong test. The calibration curve was used to evaluate the calibration of the nomogram, and decision curve analysis (DCA) was used to assess the net benefit of nomogram-assisted decisions. A 3-fold cross-validation method was used to randomly split the training cohort into 3 sets, where every two sets were the training sets and the remaining set was the validation set. The average AUC values of the 3 results were used to assess the stability of the nomogram. A two-tailed $P< 0.05$ was considered statistically significant.

Results

Participant characteristics

One hundred and three PCa patients with mean age of 71 years (range, 54-83 years) were included in our study. The overall PSMs incidence was 46.6% (48/103). There were significant differences between the two groups with respect to TPSA, PPC and pathological extra-prostatic extension ($P< 0.05$), but not differences in age, PI-RADS v2 score, prostate volume, lesion diameter, biopsy-based Gleason score, cTNM, postoperative Gleason score and pTNM ($P > 0.05$). The clinical characteristics are shown in Table 2. A stratified sampling method was used to divide the data into the training set and testing set at a ratio of 7:3. Of the 103 patients in this study, 72 were assigned to the training set, and 31 were assigned to the testing set. There were no significant differences in clinical characteristics between the training and testing sets. The details are shown in Tables 3 and 4.

Consistency analysis

As shown in Table 5, supporting information, the ICC values were > 0.9, which suggested excellent inter-observer agreement.

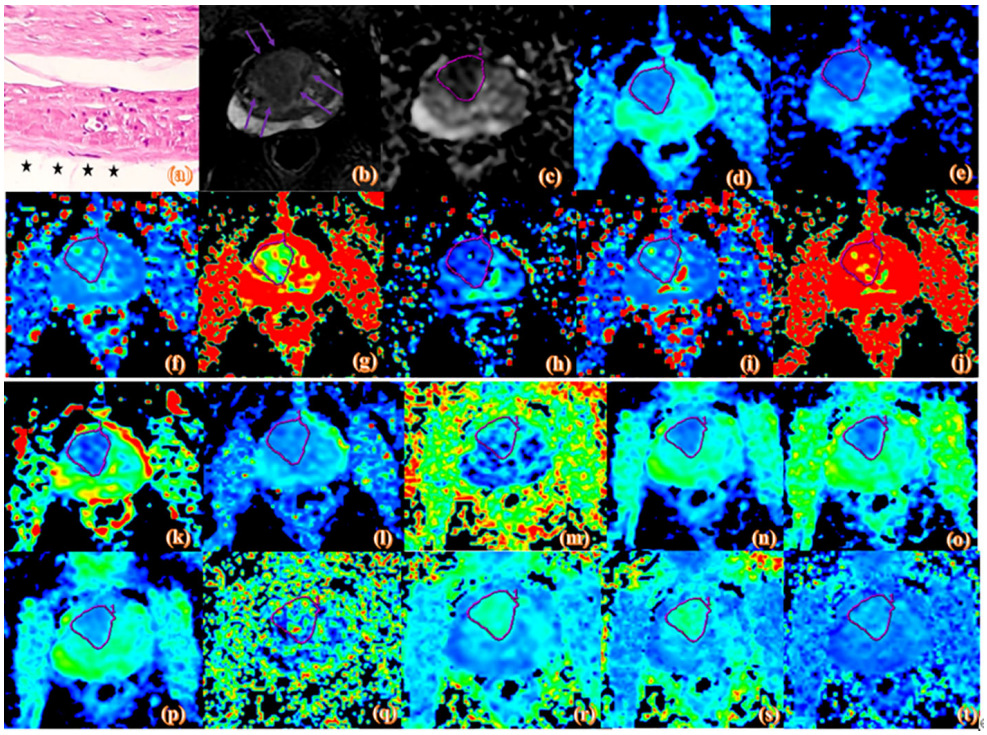


FIGURE 2
Pathologic and MR parametric maps of a 77-year-old patient (PSA level, 46.4 ng/mL; postoperative Gleason score, 3 + 4 = 7) with negative surgical margins. No neoplastic cells are seen at the ink mark of the margin indicated by the black star on the histologic map (20 × 10 magnification). **(A)**. T2WI map is shown **(B)**, lesion is indicated by pink arrows. ADC map is shown **(C)**, IVIM maps are shown **(D–L)** and DKI maps are shown **(M–T)**. Lesions are indicated by pink ROI. ADC_{ME} value is $0.989\times10^{-3}mm^2/s$, ADC_{BE} value is $0.716\times10^{-3}mm^2/s$, D_{mono} value is $0.462\times10^{-3}mm^2/s$, D^*_{mono} value is 0.0046 mm^2/s, f_{mono} value is 0.426%, D_{Bi} value is $0.417\times10^{-3}mm^2/s$, D^*_{Bi} value is 0.0044 mm^2/s, f_{Bi} value is 0.486%, DDC value is $0.975\times10^{-3}mm^2/s$, α value is 0.671, FA value is 0.261, MD value is 1.22 um^2/ms, Da value is 1.55 Am^2/ms, Dr value is 1.06 um^2/ms, FAk value is 0.404, MK value is 0.791, Ka value is 0.91, and Kr value is 0.659.

Correlations between mpMRI parameters and PSMs

There were four potential predictors selected on the basis of the data from the 103 patients with non-zero coefficients in the LASSO regression model, including Da, D_{mono}, D_{Bi}, and DDC (Figure 4).

Prediction model development

Introducing preoperatively clinical parameters TPSA, PPC, and cTNM as independent predictors, a C-nomogram was developed and is presented in Figure 5A. Introducing the Da, D_{mono}, D_{Bi}, and DDC as independent predictors, an M-nomogram was developed and is presented in Figure 5B. Introducing the MRI and preoperatively clinical parameters as independent predictors, an MC-nomogram was developed and is presented in Figure 5C.

Comparison of predictive model performance

The ROC of the C-nomogram (sensitivity=64%, specificity=65% and AUC=0.683), the M-nomogram (sensitivity=57%, specificity=88% and AUC=0.735) and the MC-nomogram (sensitivity= 64%, specificity=82% and AUC=0.756) were shown in Figures 5D, H, L. The models had moderately good performance. Although there was no statistical difference in AUC between the three models (C-nomogram VS. M-nomogram, p=0.71, C-nomogram VS. CM-nomogram, p=0.05, M-nomogram VS. CM-nomogram, p=0.1). However, the specificity of M-nomogram was better than that of C-nomogram, and the model was optimized after MRI parameters was combined with clinical parameters, and the optimal sensitivity and specificity were obtained (sensitivity= 64%, specificity=82%).

The calibration curves of the three nomograms used to predict the risk of PSMs in patients with PCa also showed good agreement (Figures 5E, I, M).

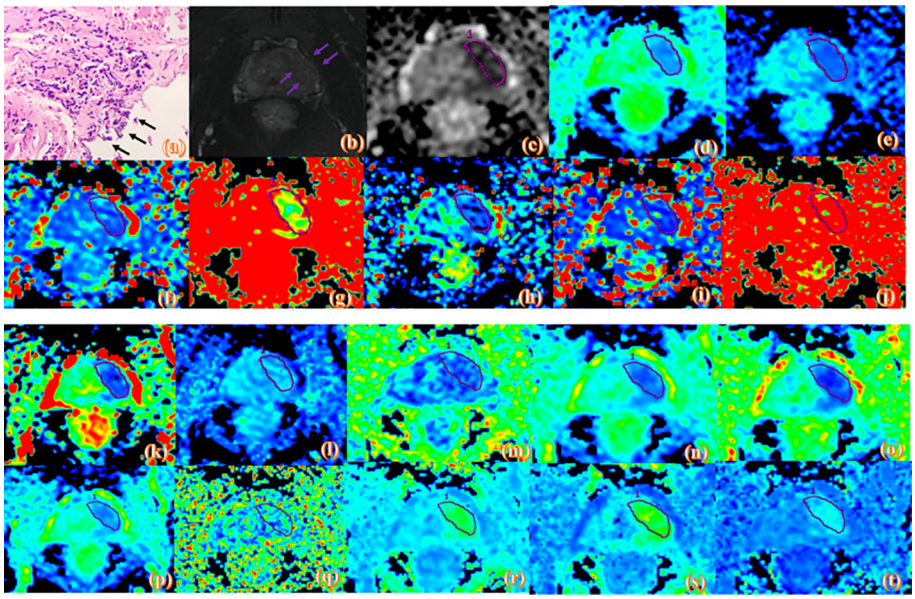


FIGURE 3
Pathologic and MR parametric maps of a 70-year-old patient (PSA level, 23.2 ng/mL; postoperative Gleason score, 4 + 5 = 9) with positive surgical margins. Neoplastic cells are indicated by black arrows breaking through the edge of the ink blot on the histologic map (20 × 10 magnification). **(A)**. T2WI map is shown **(B)**, lesion is indicated by pink arrows. ADC map is shown **(C)**, IVIM maps are shown **(D–L)** and DKI maps are shown **(M–T)**. Lesions are indicated by pink ROI. ADC_{ME} value is $0.788\times10^{-3}mm^2/s$, ADC_{BE} value is $0.592\times10^{-3}mm^2/s$, D_{mono} value is $0.439\times10^{-3}mm^2/s$, D^*_{mono} value is $0.0039mm^2/s$, f_{mono} value is 0.303%, D_{Bi} value is $0.285\times10^{-3}mm^2/s$, D^*_{Bi} value is $0.008\ mm^2/s$, f_{Bi} value is 0.476%, DDC value is $0.692\times10^{-3}mm^2/s$, α value is 0.743, FA value is 0.134, MD value is 1.06 um^2/ms, Da value is 1.25 um^2/ms, Dr value is 0.963 um^2/ms, FAk value is 0.25, MK value is 1.03, Ka value is 1.11, and Kr value is 0.948.

The decision curve analysis for the PSMs risk nomograms were presented in (Figures 5F, J, N). The decision curve showed that it would be more accurate to use MC-nomogram in the current study to predict the risk of PSMs in the range from 0.2 - 0.7.

The 3-fold cross-validation of the nomograms demonstrated its stability in predicting PSMs (Figures 5G, K, O).

Discussion

PSMs in PCa patients are commonly associated with BCR and a higher risk for secondary treatment (21, 22). Therefore, identifying predictive factors may help urologists select the at-risk patients who are more likely to benefit from LRP therapy

TABLE 2 Clinical characteristics of patients with negative and positive margins.

Characteristics	NSMs (n = 55)	PSMs (n = 48)	*P value*
Age (year), median [IQR]	72 (67–77)	71 (67–76)	0.743
TPSA (ng/ml), median [IQR]	15.53 [10.17-26.35]	25.52 [14.85-56.22]	**0.004***
PI-RADS v2, n (%)			0.336
4	54 (98)	45 (94)	
5	1 (2)	3 (6)	
Prostate volume (ml)	34.36(IQR,47.18-23.38)	34.16(IQR,54.05-27.71)	0.248
Lesion diameter (cm)	1.45 (IQR,1.97-1.00)	1.10 (IQR,1.80-0.60)	0.061
Biopsy Gleason score, n (%)			0.133
ISUP 1	18 (33)	7 (15)	
ISUP 2	4 (7)	5 (10)	
ISUP 3	6 (11)	5 (10)	
ISUP 4	19 (35)	16 (33)	
ISUP 5	8 (15)	15 (32)	

(Continued)

TABLE 2 Continued

Characteristics	NSMs (n = 55)	PSMs (n = 48)	*P value*
Percent of positive cores, median [IQR]	0.33 [0.17-0.58]	0.63 [0.33-0.83]	**< 0.001***
cTNM, n (%)			0.280
T2a	47 (85)	39 (81)	
T2b	2 (4)	0	
T2c	6 (11)	9 (19)	
Postop Gleason score, n (%)	PSM	NSM	0.083
ISUP 1	7 (13)	6 (12.5)	
ISUP 2	12 (22)	6 (12.5)	
ISUP 3	4 (7)	11 (23)	
ISUP 4	10 (18)	13 (27)	
ISUP 5	22 (40)	12 (25)	
pTNM, n (%)			0.334
pT2	43 (78)	39 (81)	
pT3a	3 (6)	0	
pT3b	9 (16)	9 (19)	
Pathological extra-prostatic extension			**0.001***
No	30 (62.5)	52 (94.5)	
Yes	18 (37.5)	3 (5.5)	
Positive margin position, n (%)			—
Peripheral margin	—	31	
Tip incisal margin	—	19	
Basal margin	—	21	

NSMs, negative surgical margins; PSMs, positive surgical margins, TPSA, total prostate specific antigen. *P value is statistically significant.

and the optimal surgical procedure can be planned. For high-risk patients, who can choose to retain one side or not to retain the neurovascular bundle during surgery, or to combine ADT before surgery to reduce the rate of PSMs. Previous studies have shown that several clinical and pathologic factors have the potential of predicting PSMs (4, 8, 23). Zhang et al. (8) conducted a comprehensive meta-analysis and systematic review with a sample of 50,014 patients, and showed that TPSA, biopsy-based Gleason score, postoperative Gleason score, pTNM, positive lymph nodes, extra-prostatic extension, and seminal vesicle invasion are independent prognostic factors for PSMs. None of the studies, however, have systematically predicted the

TABLE 3 Clinical characteristics of the training and testing sets for MRI nomogram predicting PSMs.

Variable	Training set	Testing set	*p*-value
NSMs	38 (52.8)	17 (54.8)	0.847
PSMs	34 (47.2)	14 (45.2)	
Age, y	71 ± 7	71 ± 6	0.816
TPSA, ng/ml	17.7 (IQR,44.3-11.2)	19.1 (IQR,44.9-14.0)	0.326
Biopsy Gleason score, n (%)			0.542
ISUP 1	17 (23.6)	8 (25.8)	
ISUP 2	5 (6.9)	4 (12.9)	
ISUP 3	7 (9.7)	4 (12.9)	
ISUP 4	24 (33.3)	11 (35.5)	
ISUP 5	19 (26.4)	4 (12.9)	
Percent of positive cores, median [IQR]	0.46 (IQR,0.81-0.17)	0.42 (IQR,0.67-0.17)	0.610
Postop Gleason score, n (%)			0.463
ISUP 1	9 (12.5)	4 (12.9)	

(Continued)

TABLE 3 Continued

Variable	Training set	Testing set	*p*-value
ISUP 2	12 (16.7)	6 (19.4)	
ISUP 3	11 (15.3)	4 (12.9)	
ISUP 4	13 (18.1)	10 (32.3)	
ISUP 5	27 (37.5)	7 (22.6)	
pTNM, n (%)			0.196
pT2	54 (75.0)	28 (90.3)	
pT3a	3 (4.2)	0 (0.0)	
pT3b	15 (20.8)	3 (9.7)	
cTNM, n (%)			0.210
T2a	58 (80.6)	28 (90.3)	
T2b	2J (2.8)	0 (0.0)	
T2c	12 (16.7)	3 (9.7)	
PI-RADS			1.000
4	69 (95.8)	30 (96.8)	
5	3 (4.2)	1 (3.2)	

NSMs, negative surgical margins; PSMs, positive surgical margins, TPSA, total prostate specific antigen.

TABLE 4 Clinical characteristics of the training and testing sets for MRI combined with clinical parameters nomogram predicting PSMs.

Variable	Training set	Testing set	*p*-value
NSMs	38 (53)	17 (55)	0.847
PSMs	34 (47)	14 (45)	
Age, y	72 ± 6	70 ± 7	0.248
TPSA, ng/ml	19.4 (IQR,53.1-12.5)	17.3 (IQR,30.6-8.5)	0.052
Biopsy Gleason score, n (%)			0.602
ISUP 1	17 (23.6)	8 (25.8)	
ISUP 2	6 (8.3)	3 (9.7)	
ISUP 3	8 (11.1)	3 (9.7)	
ISUP 4	22(30.6)	13(41.9)	
ISUP 5	19(26.4)	4(12.9)	
Percent of positive cores, median [IQR]	0.50 (IQR,0.83-0.19)	0.42 (IQR=0.58-0.17)	0.178
Postop Gleason score, n (%)			0.542
ISUP 1	9 (12.5)	4 (12.9)	
ISUP 2	11 (15.3)	7 (22.6)	
ISUP 3	11 (15.3)	4 (12.9)	
ISUP 4	14 (19.4)	9 (29.0)	
ISUP 5	27 (37.5)	7 (22.6)	
pTNM, n (%)			0.068
pT2	53 (73.6)	29 (93.5)	
pT3a	3 (4.2)	0 (0.0)	
pT3b	16 (22.2)	2 (6.5)	
cTNM, n (%)			0.889
T2a	60 (83.3)	26 (83.9)	
T2b	2 (2.8)	0 (0.0)	
T2c	10 (13.9)	5 (16.1)	
PI-RADS			1.000
4	69 (95.8)	30 (96.8)	
5	3 (4.2)	1 (3.2)	

NSMs, negative surgical margins; PSMs, positive surgical margins, TPSA, total prostate specific antigen.

TABLE 5 Two-observer measurement consistency.

Variable	NSMs (n = 55)			PSMs (n = 48)		
	Observer 1	Observer 2	ICC	Observer 1	Observer 2	ICC
ADC_{ME} ($\times10^{-3}mm^2/s$)	0.943 ± 0.03	0.941 ± 0.028	0.993	0.876 ± 0.021	0.875 ± 0.0209	0.993
IVIM						
ADC_{BE}	0.725 ± 0.019	0.723 ± 0.02	0.993	0.636 ± 0.014	0.638 ± 0.014	0.99
D_{mono} ($\times10^{-3}mm^2/s$)	0.517 ± 0.014	0.517 ± 0.015	0.991	0.455 ± 0.009	0.457 ± 0.009	0.984
D^*_{mono} (mm^2/s)	0.0192 ± 0.006	0.0187 ± 0.006	0.998	0.01 ± 0.003	0.009 ± 0.002	0.993
f_{mono} (%)	0.385 ± 0.013	0.384 ± 0.013	0.992	0.338 ± 0.009	0.337 ± 0.009	0.973
D_{Bi} ($\times10^{-3}mm^2/s$)	0.478 ± 0.025	0.479 ± 0.026	0.985	0.343 ± 0.02	0.37 ± 0.027	0.875
D^*_{Bi} (mm^2/s)	0.0254 ± 0.007	0.0248 ± 0.007	0.997	0.0183 ± 0.004	0.0178 ± 0.004	0.996
f_{Bi} (%)	0.444 ± 0.018	0.448 ± 0.019	0.991	0.436 ± 0.017	0.437 ± 0.017	0.985
DDC ($\times10^{-3}mm^2/s$)	0.966 ± 0.043	0.947 ± 0.046	0.96	0.754 ± 0.032	0.762 ± 0.034	0.982
α	0.705 ± 0.012	0.71 ± 0.013	0.99	0.706 ± 0.012	0.705 ± 0.011	0.995
DKI						
FA	0.214 ± 0.009	0.214 ± 0.008	0.993	0.213 ± 0.008	0.213 ± 0.008	0.996
MD (um^2/ms)	1.171 ± 0.029	1.177 ± 0.03	0.972	0.415 ± 0.131	0.414 ± 0.137	0.996
Da (um^2/ms)	1.426 ± 0.035	1.43 ± 0.036	0.972	0.415 ± 0.131	0.414 ± 0.137	0.996
Dr (um^2/ms)	1.03 ± 0.029	1.43 ± 0.036	0.972	0.415 ± 0.131	0.414 ± 0.137	0.996
FAk	0.30 ± 0.013	0.30 ± 0.013	0.992	0.31 ± 0.015	0.31 ± 0.014	0.985
MK	0.904 ± 0.021	0.904 ± 0.021	0.987	0.969 ± 0.024	0.969 ± 0.024	0.991
Ka	0.973 ± 0.023	0.971 ± 0.023	0.994	1.081 ± 0.03	1.079 ± 0.031	0.989
Kr	0.809 ± 0.02	0.809 ± 0.02	0.993	0.87 ± 0.022	0.868 ± 0.021	0.992

NSMs, negative surgical margins; PSMs, positive surgical margins.

post-LRP margin status based on mpMRI. Currently, mpMRI is widely recommended for detection and localization of PCa, and studies have indicated that mpMRI improves the predictions of preoperative clinical nomograms (24, 25). And the application of artificial intelligence (26) and mpMRI-3D model (27) also provides more possibilities to reduce the rate of PSMs. Herein, we developed three nomograms based on clinical, mpMRI and mpMRI combined with clinical parameters. Further, the performance of the three prediction models was compared.

First, the PSMs rate was 46.6%, which was higher than the results reported by Qu et al. (4) (PSMs rate = 34%). The reason for the difference may be that the patients in our study were diagnosed at a later stage; specifically, 56.3% of the patients had a biopsy-based Gleason score ≥4 compared to 38% of patients in their study.

Moreover, we showed that Da, D_{mono}, D_{Bi}, and DDC were associated with PSMs risk in PCa patients based on LASSO regression analysis. The M-nomogram achieved a higher AUC (0.735) compared to C-nomogram (0.683) for PSMs prediction, and suggested that lower Da, Dmono, DBi, and DDC were the key parameters that determined the risk of PSMs for PCa patients. Because as tumors grow, both cell overcrowding and changes in stroma production alter cell-stroma and cell-cell associations in an ongoing dynamic process that disturbs the microarchitecture (28). These microstructural changes promote the proliferation of tumor cells and the interstitial transition of tumor cells, resulting in restricted diffusion of water molecules. Moreover, Da, D_{mono}, D_{Bi} and DDC reflect the diffusion of water molecules in and out of cells. Therefore, the decrease in Da, D_{mono}, D_{Bi}, and DDC values in the PSMs group may be related to an increase in cell number and the loss of interstitial matrix. This is consistent with the previous parameters of Alessi et al. (29), who also showed that ADC performs well in PSMs prediction, with lower ADC values observed in PSMs patients. The ADC model, however, tends to oversimplify the complexity of prostate tissue while ignoring the biological specificity of PCa, which results in poor predictive performance (30). Bourne et al. (31) compared the information content of four phenomenologic diffusion models in whole prostate tissue *ex-vivo* using the Akaike information criterion. Bourne et al. (31) found the biexponential and DKI model to have a higher information content than the mono-exponential DWI model. Therefore, we attempted to construct a preoperative prediction model of PSMs using a more complex model that might provide a richer informative description of DWI signals in PCa and obtain better predictive power.

Recent studies have also found that preoperative mpMRI can be used to predict PSMs and appears to have a significant favorable impact on surgical planning. A retrospective study of 179 patients with in-house robotic assisted LRP, M. Quentin et al. found that length of capsular tumor contact was the best MRI predictor for PSMs at the capsule and distance to the membranous urethra for tumors with PSMs at the apical urethra (32). Irini Youssef et al. found that pathologic T-stage, anteroposterior pelvic outlet and pelvic

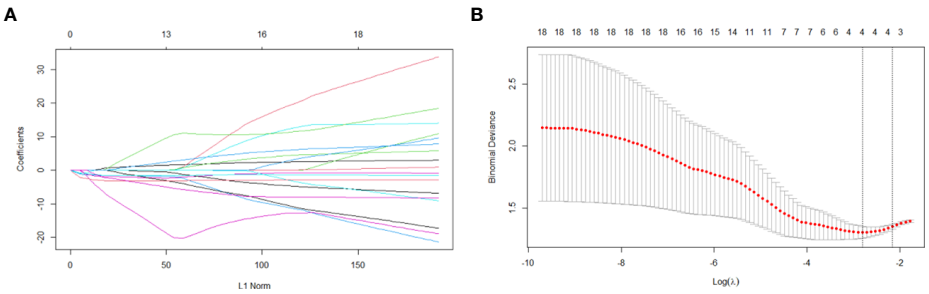


FIGURE 4
Variable selection based on the LASSO regression model. A coefficient profile plot was produced against the log(lambda) sequence **(A)**. Four variables with non-zero coefficients were selected by optimal lambda. By verifying the optimal parameter (lambda) in the LASSO model, the partial likelihood deviance (binomial deviance) curve was plotted versus log(lambda) and dotted vertical lines were drawn based on 1 standard error criterion **(B)**.

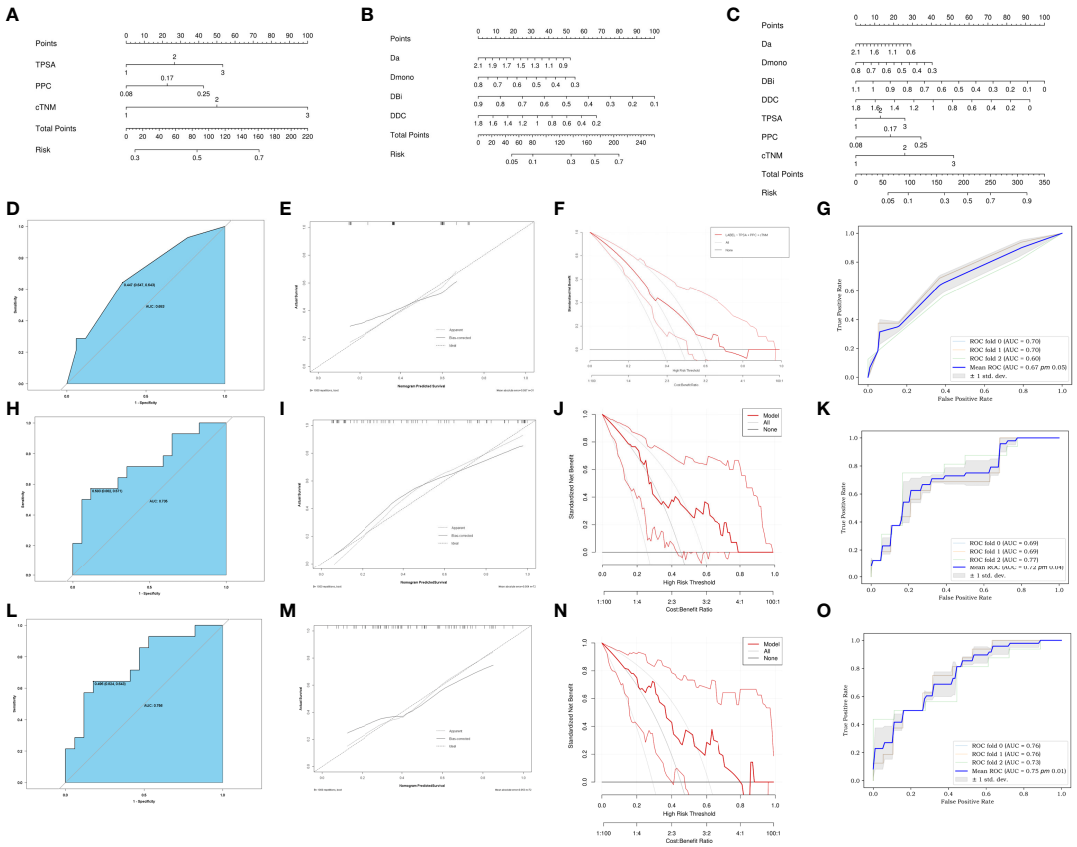


FIGURE 5
Development of C-nomogram **(A)**, M-nomogram **(B)** and MC-nomogram **(C)** predicting PSMs after laparoscopic radical prostatectomy. **(D, H, L)** ROC validation of the PSMs risk nomogram prediction. The blue area represented the performance of the nomogram. **(E, I, M)** Calibration curves of the PSMs risk nomogram prediction. The y-axis represents the actual diagnosed PSMs. The x-axis represents the predicted risk of PSMs. The diagonal dotted line represents a perfect prediction by an ideal model. The solid line represented the performance of the C-nomogram **(D)**, M-nomogram **(H)** and MC-nomogram **(L)**, which indicated that a closer fit to the diagonal dotted line represented a better prediction. **(F, J, N)** Decision curve of the PSMs risk nomogram prediction. The y axis represents the net benefit and the x axis represents the risk threshold. The thick solid line represents the assumption that all patients had no PSMs. The thin solid line represents the assumption that all patients had PSMs. The red line represents the risk nomogram. **(F)** From C-nomogram, **(J)** the M-nomogram and **(N)** MC-nomogram. The net benefit of the MC-nomogram is highest in the range from 0.2-0.7.). **(G, K, O)** Mean ROC curve of the nomogram to predict PSMs after 3-fold cross-validation. TPSA, total prostate specific antigen; PPC, percent of positive cores.

depth were risk factors for positive margins (33). The findings were consistent with our study. However, our study highlights the application of functional sequences. We combine different DWI models to screen for optimal functional parameters. The focus is on the predictive power of the biological behavior of the tumors themselves for positive margins, which has rarely been addressed in previous studies.

In addition, by incorporating preoperative clinical indicators, a combined risk MC-nomogram was created. The MC-nomogram also had a higher AUC (0.756). Comparison of the calibration curve and DCA, showed that MC-nomogram demonstrated relatively good calibration power and clinical net benefit. This finding indicates that the ability of MRI-based nomogram to predict PSMs was optimized after combining clinical indicators. He et al. (34) reported that radiomics signatures based on ADC predict PSMs (AUC=0.733), and when combined with clinical parameters, improves the model efficiency (AUC=0.766). This finding is consistent with our parameters. Preoperative mpMRI can be used to predict PSMs and appears to have a significant favorable impact on surgical planning (35, 36). However, previous studies mostly used mpMRI to predict pathologic extra-prostatic extension (37), and there were limited studies to predict PSMs. Model performance was improved when clinical parameters were included in the mpMRI prediction model. A recent study showed that the nomogram described by Gandaglia et al. (38) using an MRI combined with clinical parameters as a staging method improved discrimination in predicting postoperative adverse pathologic factors (38). Therefore, it is difficult to predict PSMs only using preoperative mpMRI or clinical parameters, and a combination of the them is necessary.

Our study had some significant limitations. First, this was a single-center retrospective study, and thus, the limited sample size may cause selection bias and other confounding factors. Second, the MC-nomogram lacked external validation. Third, we constructed predictive models based on mpMRI parameters obtained from complex DWI models, which may have limited the application and multicenter generalization of the MC-nomogram. And more our study did not include robot-assisted LRP, so results could be not applicable in robotic-assisted LRP.

Conclusions

The mpMRI-based nomogram can predict PSMs in PCa patients. Although its AUC (0.735) is not statistically different from that of the clinical-based nomogram AUC (0.683). However, mpMRI-based nomogram has higher specificity (88% VS. 63%), model stability, and clinical benefit than clinical-based nomogram. And the predictive ability of mpMRI plus clinical parameters for PSMs is further improved.

Data availability statement

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

Ethics statement

The studies involving human participants were reviewed and approved by the institutional review board of First Affiliated Hospital of Dalian Medical University. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

Author contributions

SM and AL conceived of the presented idea. SM and LC performed the measurements. AL supervised the work. SM and QZ processed the experimental data and performed the analysis. SM drafted the manuscript and NW aided in working on the manuscript. All authors contributed to the article and approved the submitted version.

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. Mottet N, van den Bergh RCN, Briers E, Van den Broeck T, Cumberbatch MG, De Santis M, et al. EAU-EANM-ESTRO-ESUR-SIOG guidelines on prostate cancer-2020 update. part 1: Screening, diagnosis, and local treatment with curative intent. *Eur Urol.* (2021) 79(2):243–62. doi: 10.1016/j.eururo.2020.09.042

2. Yang R, Cao K, Han T, Zhang YF, Zhang GT, Xu LF, et al. Perineural invasion status, Gleason score and number of positive cores in biopsy pathology are predictors of positive surgical margin following laparoscopic radical prostatectomy. *Asian J Androl* (2017) 19(4):468–72. doi: 10.4103/1008-682X.173444

3. Tian XJ, Wang ZL, Li G, Cao SJ, Cui HR, Li ZH, et al. Development and validation of a preoperative nomogram for predicting positive surgical margins after laparoscopic radical prostatectomy. *Chin Med J (Engl)* (2019) 132(8):928–34. doi: 10.1097/CM9.0000000000000161

4. Qu W, Yu S, Tao J, Dong B, Fan Y, Du H, et al. Evaluating incidence, location, and predictors of positive surgical margin among Chinese men undergoing robot-assisted radical prostatectomy. *Cancer Control* (2021) 28:10732748211055265. doi: 10.1177/10732748211055265

5. Zhang L, Wu B, Zha Z, Zhao H, Jiang Y, Yuan J. Positive surgical margin is associated with biochemical recurrence risk following radical prostatectomy: A meta-analysis from high-quality retrospective cohort studies. *World J Surg Oncol* (2018) 16(1):124. doi: 10.1186/s12957-018-1433-3

6. Celik S, Eker A, Bozkurt İH, Bolat D, Basmacı İ, Şefik E, et al. Factors affecting biochemical recurrence of prostate cancer after radical prostatectomy in patients with positive and negative surgical margin. *Prostate Int* (2020) 8(4):178–84. doi: 10.1016/j.prnil.2020.08.003

7. Martini A, Marqueen KE, Falagario UG, Waingankar N, Wajswol E, Khan F, et al. Estimated costs associated with radiation therapy for positive surgical margins during radical prostatectomy. *JAMA Netw Open* (2020) 3(3):e201913. doi: 10.1001/jamanetworkopen.2020.1913

8. Zhang L, Zhao H, Wu B, Zha Z, Yuan J, Feng Y. Predictive factors for positive surgical margins in patients with prostate cancer after radical prostatectomy: A systematic review and meta-analysis. *Front Oncol* (2021) 10:539592. doi: 10.3389/fonc.2020.539592

9. D'Amico AV, Whittington R, Malkowicz SB, Schultz D, Blank K, Broderick GA, et al. Biochemical outcome after radical prostatectomy, external beam radiation therapy, or interstitial radiation therapy for clinically localized prostate cancer. *JAMA* (1998) 280(11):969–74. doi: 10.1001/jama.280.11.969

10. Cooperberg MR, Hilton JF, Carroll PR. The CAPRA-s score: A straightforward tool for improved prediction of outcomes after radical prostatectomy. *Cancer* (2011) 117(22):5039–46. doi: 10.1002/cncr.26169

11. Bryant RJ, Hobbs CP, Eyre KS, Davies LC, Sullivan ME, Shields W, et al. Comparison of prostate biopsy with or without prebiopsy multiparametric magnetic resonance imaging for prostate cancer detection: An observational cohort study. *J Urol* (2019) 201(3):510–9. doi: 10.1016/j.juro.2018.09.049

12. McEvoy SH, Raeside MC, Chaim J, Ehdaie B, Akin O. Preoperative prostate MRI: A road map for surgery. *AJR Am J Roentgenol* (2018) 211(2):383–91. doi: 10.2214/AJR.17.18757

13. Hectors SJ, Semaan S, Song C, Lewis S, Haines GK, Tewari A, et al. Advanced diffusion-weighted imaging modeling for prostate cancer characterization: Correlation with quantitative histopathologic tumor tissue composition-a hypothesis-generating study. *Radiology* (2018) 286(3):918–28. doi: 10.1148/radiol.2017170904

14. American College of Radiology. *Prostate imaging– reporting and data system, PI-RADS v2.1* . American College of Radiology. Available at: www.acrorg/Quality-Safety/Resources/PIRADS2019 (Accessed 6th of January 2020).

15. Weinreb JC, Barentsz JO, Choyke PL, Cornud F, Haider MA, Macura KJ, et al. And data system: 2015, version 2. *Eur Urol* (2016) 69(1):16–40. doi: 10.1016/j.eururo.2015.08.052

16. van der Kwast TH, Amin MB, Billis A, Epstein JI, Griffiths D, Humphrey PA, et al. International society of urological pathology (ISUP) consensus conference on handling and staging of radical prostatectomy specimens. working group 2: T2 substaging and prostate cancer volume. *Mod Pathol* (2011) 24(1):16–25. doi: 10.1038/modpathol.2010.156

17. Le Bihan D. Apparent diffusion coefficient and beyond: What diffusion MR imaging can tell us about tissue structure. *Radiology* (2013) 268(2):318–22. doi: 10.1148/radiol.13130420

18. Le Bihan D, Breton E, Lallemand D, Aubin ML, Vignaud J, Laval-Jeantet M. Separation of diffusion and perfusion in intravoxel incoherent motion MR imaging. *Radiology* (1988) 168(2):497–505. doi: 10.1148/radiology.168.2.3393671

19. Bennett KM, Schmainda KM, Bennett RT, Rowe DB, Lu H, Hyde JS. Characterization of continuously distributed cortical water diffusion rates with a stretched-exponential model. *Magn Reson Med* (2003) 50(4):727–34. doi: 10.1002/mrm.10581

20. Jensen JH, Helpern JA, Ramani A, Lu H, Kaczynski K. Diffusional kurtosis imaging: The quantification of non-gaussian water diffusion by means of magnetic resonance imaging. *Magn Reson Med* (2005) 53(6):1432–40. doi: 10.1002/mrm.20508

21. Preisser F, Coxilha G, Heinze A, Oh S, Chun FK, Sauter G, et al. Impact of positive surgical margin length and Gleason grade at the margin on biochemical recurrence in patients with organ-confined prostate cancer. *Prostate* (2019) 79 (16):1832–6. doi: 10.1002/pros.23908

22. John A, John H, Catterwell R, Selth LA, Callaghan MO. Primary Gleason grade and Gleason grade group at positive surgical margins: A systematic review and meta-analysis. *BJU Int* (2021) 127 Suppl 1:13–22. doi: 10.1111/bju.15316

23. Tuliao PH, Koo KC, Komninos C, Chang CH, Choi YD, Chung BH, et al. Number of positive preoperative biopsy cores is a predictor of positive surgical margins (PSM) in small prostates after robot-assisted radical prostatectomy (RARP). *BJU Int* (2015) 116(6):897–904. doi: 10.1111/bju.12888

24. Muehlematter UJ, Burger IA, Becker AS, Schawkat K, Hötker AM, Reiner CS, et al. Diagnostic accuracy of multiparametric MRI versus 68Ga-PSMA-11 PET/MRI for extracapsular extension and seminal vesicle invasion in patients with prostate cancer. *Radiology* (2019) 293(2):350–8. doi: 10.1148/radiol.2019190687

25. Diamand R, Ploussard G, Roumiguié M, Oderda M, Benamran D, Fiard G, et al. External validation of a multiparametric magnetic resonance imaging-based nomogram for the prediction of extracapsular extension and seminal vesicle invasion in prostate cancer patients undergoing radical prostatectomy. *Eur Urol* (2021) 79(2):180–5. doi: 10.1016/j.eururo.2020.09.037

26. Tătaru OS, Vartolomei MD, Rassweiler JJ, Virgil O, Lucarelli G, Porpiglia F, et al. Artificial intelligence and machine learning in prostate cancer patient management-current trends and future perspectives. *Diagnostics (Basel).* (2021) 11(2):354. doi: 10.3390/diagnostics11020354

27. Martini A, Falagario UG, Cumarasamy S, Jambor I, Wagaskar VG, Ratnani P, et al. The role of 3D models obtained from multiparametric prostate MRI in performing robotic prostatectomy. *J Endourol* (2022) 36(3):387–93. doi: 10.1089/end.2021.0541

28. Nia HT, Munn LL, Jain RK. Physical traits of cancer. *Science* (2020) 370 (6516):eaaz0868. doi: 10.1126/science.Aaz0868

29. Alessi S, Maggioni R, Luzzago S, Colombo A, Pricolo P, Summers PE, et al. Apparent diffusion coefficient and other preoperative magnetic resonance imaging features for the prediction of positive surgical margins in prostate cancer patients undergoing radical prostatectomy. *Clin Genitourin Cancer* (2021) 19(6):e335–45. doi: 10.1016/j.clgc.2021.04.004

30. Wichtmann BD, Zöllner FG, Attenberger UI, Schönberg SO. Multiparametric MRI in the diagnosis of prostate cancer: Physical foundations, limitations, and prospective advances of diffusion-weighted MRI. *Rofo* (2021) 193 (4):399–409. doi: 10.1055/a-1276-1773

31. Bourne RM, Panagiotaki E, Bongers A, Sved P, Watson G, Alexander DC. Information theoretic ranking of four models of diffusion attenuation in fresh and fixed prostate tissue *ex vivo*. *Magn Reson Med* (2014) 72(5):1418–26. doi: 10.1002/mrm.25032

32. Quentin M, Schimmöller L, Ullrich T, Valentin B, Demetrescu D, Al-Monajjed R, et al. Pre-operative magnetic resonance imaging can predict prostate cancer with risk for positive surgical margins. *Abdom Radiol (NY).* (2022) 47 (7):2486–93. doi: 10.1007/s00261-022-03543-z

33. Youssef I, Poch M, Raghunand N, Pow-Sang J, Johnstone PAS. Analysis of MRI radiomic pelvimetry and correlation with margin status after robotic prostatectomy. *Can J Urol* (2022) 29(1):10976–8.

34. He D, Wang X, Fu C, Wei X, Bao J, Ji X, et al. MRI-Based radiomics models to assess prostate cancer, extracapsular extension and positive surgical margins. *Cancer Imaging.* (2021) 21(1):46. doi: 10.1186/s40644-021-00414-6

35. Jäderling F, Akre O, Aly M, Björklund J, Olsson M, Adding C, et al. Preoperative staging using magnetic resonance imaging and risk of positive surgical margins after prostate-cancer surgery. *Prostate Cancer Prostatic Dis* (2019) 22 (3):391–8. doi: 10.1038/s41391-018-0116-z

36. Park MY, Park KJ, Kim MH, Kim JK. Preoperative MRI-based estimation of risk for positive resection margin after radical prostatectomy in patients with prostate cancer: development and validation of a simple scoring system. *Eur Radiol* (2021) 31(7):4898–907. doi: 10.1007/s00330-020-07569-z

37. Guerra A, Negrão E, Papanikolaou N, Donato H. Machine learning in predicting extracapsular extension (ECE) of prostate cancer with MRI: A protocol for a systematic literature review. *BMJ Open* (2022) 12(5):e052342. doi: 10.1136/bmjopen-2021-052342

38. Gandaglia G, Ploussard G, Valerio M, Mattei A, Fiori C, Roumiguié M, et al. The key combined value of multiparametric magnetic resonance imaging, and magnetic resonance imaging-targeted and concomitant systematic biopsies for the prediction of adverse pathological features in prostate cancer patients undergoing radical prostatectomy. *Eur Urol* (2020) 77(6):733–41. doi: 10.1016/j.eururo.2019.09.005

Glossary

ADC	apparent diffusion coefficient
ADCME	Monoexponential apparent diffusion coefficient
ADCBE	bi-exponential apparent diffusion coefficient
AUC	area under the curve
BCR	biochemical recurrence
cTNM	clinical TNM
Da	axial diffusion
Dr	radial diffusion
DCA	decision curve analysis
DDC	mean intra-voxel diffusion coefficient
DKI	diffusion kurtosis imaging
DWI	diffusion-weighted imaging
FA	fractional anisotropy
FAK	fractional anisotropy kurtosis
FPSA	free prostate specific antigen
ICC	intraclass correlation coefficient
IVIM	intravoxel incoherent motion model
Ka	axial kurtosis
Kr	radial kurtosis
LASSO	least absolute shrinkage and selection operator
LRP	laparoscopic radical prostatectomy
M-nomogram	MRI nomogram
MC-monogram	MRI combined with the clinical nomogram
MD	mean diffusion
MK	mean kurtosis
mpMRI	multiparametric MRI
NSM	negative surgical margin
PCa	prostate cancer
PPC	percent of positive cores
PSMs	positive surgical margins
pTNM	pathologic TNM
ROC	receiver operating characteristic curve
SS-EPI	single-shot echo planar
T1WI	T1-weighted imaging
T2WI	T2-weighted imaging
TPSA	total prostate specific antigen1

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 13 September 2022
DOI 10.3389/fonc.2022.965773

OPEN ACCESS

EDITED BY
Jasper Nijkamp,
Aarhus University, Denmark

REVIEWED BY
Shaoli Song,
Fudan University, China
Kezheng Wang,
Harbin Medical University Cancer Hospital, China

*CORRESPONDENCE
Jiong Cai
jiong_cai@163.com
Pan Wang
1298178828@qq.com
Dandan Li
807442003@qq.com

†These authors have contributed equally to this work

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

RECEIVED 10 June 2022
ACCEPTED 12 August 2022
PUBLISHED 13 September 2022

CITATION
Liu X, Hu X, Yu X, Li P, Gu C, Liu G, Wu Y, Li D, Wang P and Cai J (2022) Frontiers and hotspots of ^{18}F-FDG PET/CT radiomics: A bibliometric analysis of the published literature. *Front. Oncol.* 12:965773. doi: 10.3389/fonc.2022.965773

Frontiers and hotspots of ^{18}F-FDG PET/CT radiomics: A bibliometric analysis of the published literature

Xinghai Liu[1,2†], Xianwen Hu[1†], Xiao Yu[1,2†], Pujiao Li[1,2], Cheng Gu[1,2], Guosheng Liu[1,2], Yan Wu[1], Dandan Li[3*], Pan Wang[1*] and Jiong Cai[1*]

[1]Department of Nuclear Medicine, Affiliated Hospital of Zunyi Medical University, Zunyi, China, [2]The First Clinical College, Zunyi Medical University, Zunyi, China, [3]Department of Obstetrics, Zunyi Hospital of Traditional Chinese Medicine, Zunyi, China

Objective: To illustrate the knowledge hotspots and cutting-edge research trends of ^{18}F-FDG PET/CT radiomics, the knowledge structure of was systematically explored and the visualization map was analyzed.

Methods: Studies related to ^{18}F-FDG PET/CT radiomics from 2013 to 2021 were identified and selected from the Web of Science Core Collection (WoSCC) using retrieval formula based on an interview. Bibliometric methods are mainly performed by CiteSpace 5.8.R3, which we use to build knowledge structures including publications, collaborative and co-cited studies, burst analysis, and so on. The performance and relevance of countries, institutions, authors, and journals were measured by knowledge maps. The research foci were analyzed through research of keywords, as well as literature co-citation analysis. Predicting trends of ^{18}F-FDG PET/CT radiomics in this field utilizes a citation burst detection method.

Results: Through a systematic literature search, 457 articles, which were mainly published in the United States (120 articles) and China (83 articles), were finally included in this study for analysis. Memorial Sloan-Kettering Cancer Center and Southern Medical University are the most productive institutions, both with a frequency of 17. ^{18}F-FDG PET/CT radiomics–related literature was frequently published with high citation in *European Journal of Nuclear Medicine and Molecular Imaging* (IF9.236, 2020), *Frontiers in Oncology* (IF6.244, 2020), and *Cancers* (IF6.639, 2020). Further cluster profile of keywords and literature revealed that the research hotspots were primarily concentrated in the fields of image, textural feature, and positron emission tomography, and the hot research disease is a malignant tumor. Document co-citation analysis suggested that many scholars have a co-citation relationship in studies related to imaging biomarkers, texture analysis, and immunotherapy simultaneously. Burst detection suggests that adenocarcinoma studies are frontiers in ^{18}F-FDG PET/CT radiomics, and the landmark literature put emphasis on the reproducibility of ^{18}F-FDG PET/CT radiomics features.

Conclusion: First, this bibliometric study provides a new perspective on ^{18}F-FDG PET/CT radiomics research, especially for clinicians and researchers providing scientific quantitative analysis to measure the performance and correlation of countries, institutions, authors, and journals. Above all, there will be a continuing growth in the number of publications and citations in the field of ^{18}F-FDG PET/CT. Second, the international research frontiers lie in applying ^{18}F-FDG PET/CT radiomics to oncology research. Furthermore, new insights for researchers in future studies will be adenocarcinoma-related analyses. Moreover, our findings also offer suggestions for scholars to give attention to maintaining the reproducibility of ^{18}F-FDG PET/CT radiomics features.

KEYWORDS

bibliometric analysis, ^{18}F-FDG, PET/CT, radiomics, hot topics

Introduction

Fluorine-18 fluorodeoxyglucose positron emission computed tomography (^{18}F-FDG PET/CT) is one of the most widely used metabolic imaging methods in nuclear medicine, which can reflect the heterogeneity of glucose metabolism between tumor cells and normal cells in an early and quantitative manner (1). Generally, this technique is widely used to evaluate the differential diagnosis of benign and malignant tumors, the staging and restaging of malignant tumors, the evaluation of treatment efficacy, prognosis prediction, and so on (2, 3). Radiomics is a relatively broad concept, and its application includes traditional imaging techniques, such as computed tomography (CT), magnetic resonance imaging (MRI), and ultrasound. As the frontier of molecular imaging, nuclear medicine, represented by single-photon emission CT (SPECT) and PET/CT, has significant advantages in a series of clinical problems including efficacy evaluation of refractory malignant tumors, differential diagnosis of heterogeneous cells, and disease prognosis and survival evaluation (4, 5). Recent studies have demonstrated the advantages or superior performance of radiomics over traditional manual readings (6–8). As far as manual reading is concerned, the accuracy and reliability of the image report largely depend on the clinical experience and expertise of the doctor. Doctors who are inexperienced and unskilled in using image analysis tools often produce low-quality inspection reports, which may lead to misdiagnosis or even missed diagnosis. That might be the reason why the result of PET/CT diagnosis varies from doctor to doctor. For senior doctors, it is a basic literacy to accurately describe the examination results according to the visible imaging changes at the lesions. Unfortunately, human eyes cannot distinguish the subtle changes specific to each pixel in an image, whereas this can be done entirely through formal machine learning and machine recognition. In previous clinical studies, CT and MRI radiomics analysis has shown high accuracy in distinguishing benign and malignant tumors and in evaluating the efficacy and prognosis after treatment (9–12). In recent years, radiomics has also been used to assist in the diagnosis of endoscopic ultrasound images and to identify whether it is a COVID-19 infection by chest x-ray (13, 14). Up to now, radiomics-related literature has been analyzed by knowledge visualization maps from a macroscopic perspective. Nonetheless, bibliometric methods have not yet been used to summarize the literature on specific metabolic imaging technology.

PET/CT imaging can clearly and intuitively reflect the metabolic changes of tumor cells and detailed metabolic image changes in the establishment of treatment plans for malignant tumors, identification of benign and malignant pulmonary nodules (15), identification of radiation necrosis and tumor recurrence in glioma patients (16), non-invasive prediction of epidermal growth factor receptor (EGFR) mutations in lung adenocarcinoma (17), and other clinical problems; the diagnostic accuracy of PET/CT radiomics analysis is expected to be superior to traditional methods. As the process of precision medicine deepens, scholars have been interested in analyzing ^{18}F-FDG PET/CT metabolic images according to the core technologies of radiomics to assist in clinical diagnosis, the establishment of treatment decisions and prediction of prognosis, etc. and have achieved promising results with the number of related research publications increasing year by year (18–22).

Bibliometric is a science that uses quantitative methods such as mathematics and statistics to study the distribution laws and quantitative relationships of documents. In order to encourage researchers from various disciplines to actively and creatively participate in the practice of ^{18}F-FDG PET/CT radiomics, this

present study mainly uses the CiteSpace software to visualize and analyze the related research of ^{18}F-FDG PET/CT radiomics, sorting out the current situation and future development trend of this field. The aim of this study is to offer clinicians an objective summary of the development processes and research hotspots and provide reference and scientific bases for scholars to refine the research direction.

Methods

Data retrieval strategy

Before searching the literature, we contacted the Information Department of the Zunyi Medical University Library to interview professional workers about data search strategies. Based on the interview, we expanded the synonyms and subordinate terms based on the MeSH database (https://www.ncbi.nlm.nih.gov/mesh, USA), and, finally, we determined the search terms and the combined search form. The retrieval formula can be described as follows: TS=((^{18}F-FDG OR FDG OR Fluorodeoxyglucose OR Fluorine 18 fluorodeoxyglucose OR 2 Fluoro 2 deoxy D glucose) AND (Positron Emission Computed Tomography OR PET OR PET/CT OR PET-CT) AND Radiomic*). The Web of Science (WOS) database was used to search the Science Citation Index Expanded (SCI-E) from its core collection (WoSCC) database, all data were searched by our corresponding authors independently (Prof. Pan Wang and Dr. Jiong Cai, Affiliated Hospital of Zunyi Medical University, Zunyi, China). Relevant data published between 2013 and 2021 were obtained. The process of data downloading and literature searching was all completed on 1/3/2022, in order to preclude potential bias caused by frequently updated data. The proofreading notices, editing materials, conference papers, and retraction notices were excluded, and, by reading the titles and abstracts, obviously irrelevant literatures were excluded as well. Original articles and reviews were included for utilization for the bibliometric analysis. Details of data searching were presented in Supplementary Figure 1, and general data from WoSCC were shown in Supplementary Tables 1, 2. In case that a disagreement occurs during the execution of the above steps, either the two corresponding authors discussed with each other to resolve problems or they consulted the third author to assist in judgment.

Analysis method

All WoSCC data were transformed into text containing information of authors, research institutions, subjects, years, keywords, abstracts, journals, volumes, and page numbers. We created a spider map and bar charts to identify publication volumes and annual trends by using OriginPro 2021 (OriginLab Corporation, USA). CiteSpace 5.8.R3 (Drexel University, USA) was applied to identify visualized map: 1) Analyzation map of the overall research status was generated by illustrating the distribution of disciplines, journals, countries/regions, institutions, and the contribution of authors; 2) keyword co-occurrence and clustering analysis were performed to anchor research hotspots; and 3) co-citation and citation burst analysis were performed to target frontiers in the field. VOSviewer 1.6.16 (Leiden University, Netherlands) was used to assist in constructing visualization maps of countries/regions and keyword co-occurrence. Scimago Graphica 1.0.18 (https://www.graphica.app/, USA) was applied to form a world map depicting the publication counts of each country. All software used in this study was performed on Windows 10 (64-bit) Chinese version operating system (Microsoft, USA). When importing the data into bibliometric analysis, we set the parameters as default to obtain original maps. We checked the original maps to see if the intellectual structure matched the results of data searching. After that, adjusted parameters to ensure the generated graphs were clear enough to highlight key points. After all analyses were finished, we invited three nuclear medicine physicians to review if the results are objective and fit the developing trends in this field based on their professional knowledge. The flow chart of this bibliometric analysis is shown in Figure 1.

Results

Temporal trends of publications

A total of 552 publications were initially generated through the WoSCC search (Figure 2A). After excluding 95 publications, 457 publications were finally included in this study, comprising 361 original articles and 96 reviews. From 2013 to 2015, there were few relative studies, indicating low research interest in this area. From 2016 onward, the number of papers on ^{18}F-FDG PET/CT radiomics began to increase rapidly. In 2021, the importance of ^{18}F-FDG PET/CT radiomics research has been noticed, with 146 publications annually—an increase of more than 40% over the same period in the previous year. It is the first high-yield stage in this field, showing a significant trend of steady growth in the number of publications (Figure 2B).

Distribution of disciplines and productive journals

All articles in this study were published in 134 journals, 13 of which were with more than 10 articles published (Figure 2C). The impact factor (IF) and journal quartile were obtained from Journal Citation Reports 2020 (23). The top 3 prolific journals were *European Journal of Nuclear Medicine and Molecular Imaging* (IF9.236), *Frontiers in Oncology* (IF6.244), and

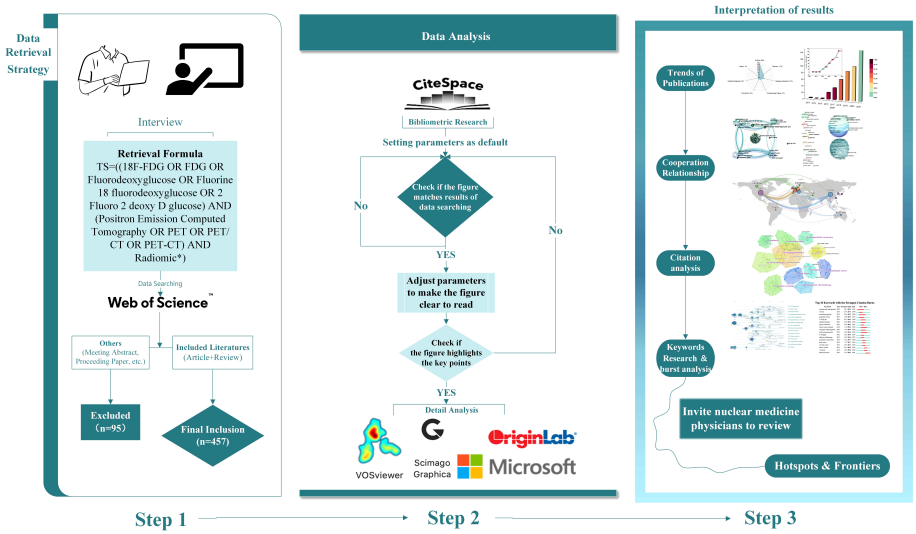


FIGURE 1
Flow chart of this bibliometric analysis. (Step 1) Data retrieval strategy of published literatures, showing the detailed process of screening and inclusion. (Step 2) Data analysis of this study, including bibliometric analysis and data processing. (Step 3) Interpretation of results, containing a brief view of results and the final process of drawing conclusions.

Cancers (IF6.639). The top journal with the highest IF was *Journal of Nuclear Medicine* (IF10.057). In addition, journals with more than 10 publications were all classified in Q1 or Q2. Specifically, with 53.84% (7/13) being Q1 and 46.15% (6/13) being Q2 among these journals, the results indicate that the above journals have strong academic performance in ^{18}F-FDG PET/CT radiomics research.

Academic journals are an important mass media for the academic communication of disciplinary knowledge, and the journal distribution of literature reflects the information related to the research status to a certain extent. Scholars can gain a simple understanding of the current research status by considering the number of publications in different journals. In the dual-visualization map (Figure 3A), there are two main links identified, with green color: Medicine/Medical/Clinical journals frequently cite studies published in Health/Nursing/Medicine journals and studies published in Medicine/Medical/Clinical journals are also cited in the studies published in Molecular/Biology/Genetics journals. More clear details can be seen in Figure 3B, showing the connection between disciplines.

Cooperation between countries/regions and institutions

All publications were distributed among 45 countries (or regions) and 214 institutions. The nodes in the visualization map represent countries/regions/institutions, and the connection line represents the strength of the relationship between the two nodes. The larger the circle of the node, the greater the volume of published documents. The thicker the connection, the closer the cooperative relationship.

Active countries

The top contributor was the United States (120 publications, 26.26%), followed by China (100 publications, 21.88%), France (58 publications, 12.69%), Italy (56 publications, 12.25%), and Netherlands (43 publications, 9.40%) (Figure 4A). Although neither China (centrality = 0.28) nor Germany (centrality = 0.25) have the highest publication frequency, we still identify the three countries as critical nodes according to their high centrality (Figure 4B), which means that they play a significant role in the field of ^{18}F-FDG PET/CT radiomics. The above results demonstrate that ^{18}F-FDG PET/CT radiomics is increasingly receiving widespread attention from global scholars and that extensive research has been conducted recently, especially in the United States. Other high-quality publications were mainly completed by Chinese, French, German, and Dutch researchers. The cooperation between countries/regions across the world was shown in Figure 4C.

Productive institutions

The visualization map shows the 214 institutes which made contributions to ^{18}F-FDG PET/CT radiomics research

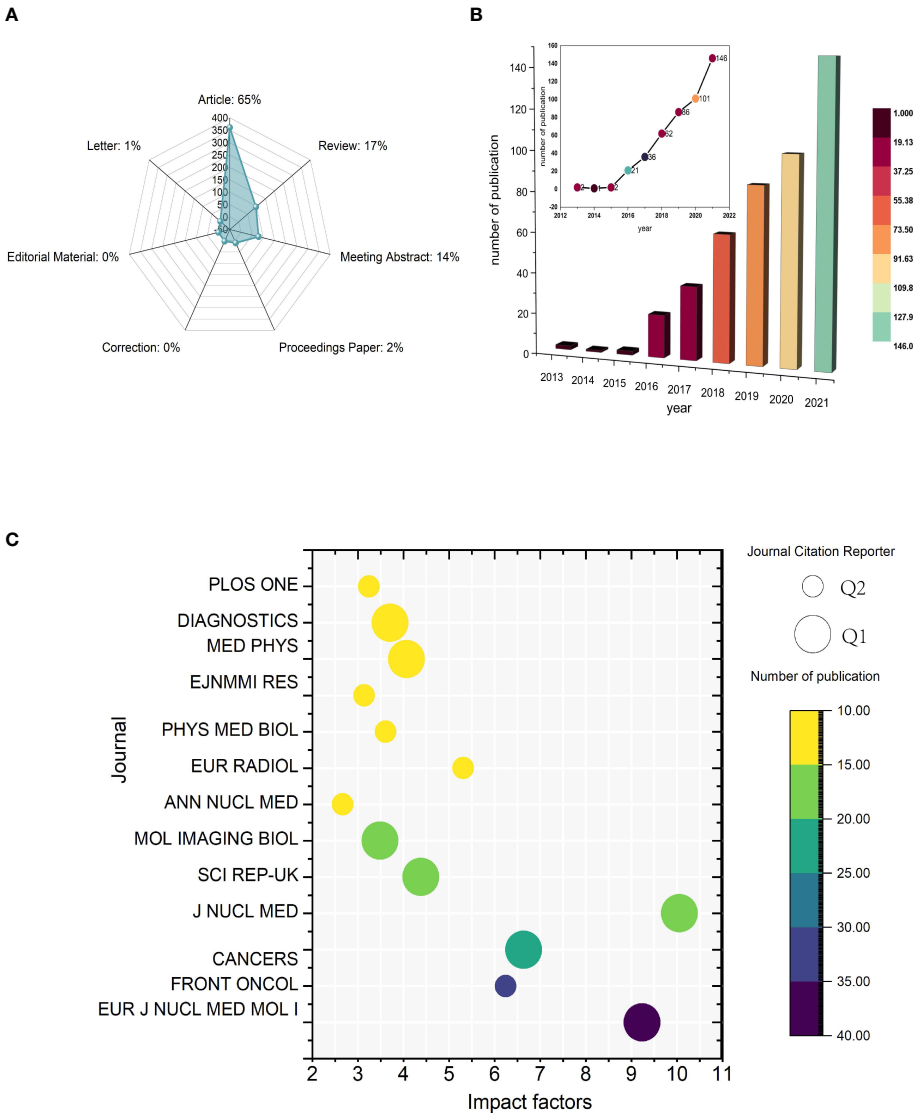


FIGURE 2
Temporal trends of publications of ^{18}F-FDG PET/CT radiomics research. **(A)** Category of all published literatures. **(B)** Annual publications and temporal trends of ^{18}F-FDG PET/CT radiomics research. **(C)** Bubble plot of journals with more than 10 publications. The size of bubbles represents classification from Journal Citation Reports and the colors are related to the number of publications.

(Figure 5)**.** Institutions with the highest frequency were shown in Table 1, and full names of the institutions are shown in Supplementary Table 3. The Memorial Sloan-Kettering Cancer Center and Southern Medical University are the most productive institutions with 17 publications, followed by Maastricht University (frequency = 14), H. Lee Moffitt Cancer Center and Research Institute (frequency = 13), and Humanitas University (frequency = 12). We investigated the possible impact of research published by institutions according to the centrality of the visualization map. The German Cancer Research Center (centrality = 0.13), the University of Groningen (centrality = 0.13), and the Memorial Sloan-Kettering Cancer Center (centrality = 0.11) tend to be the leading driving force and still dominate in this research field among top institutions.

Contribution of authors

Focusing on the research contributions of authors helps us to quickly classify scholars who are active in the discipline at this stage. A total of 288 authors have made contributions to ^{18}F-FDG PET/CT radiomics research. The top 10 prolific authors and top 10 cited authors were listed in Supplementary Tables 4, 5. Mathieu Hatt, the most productive author from Australia, published 16 articles,

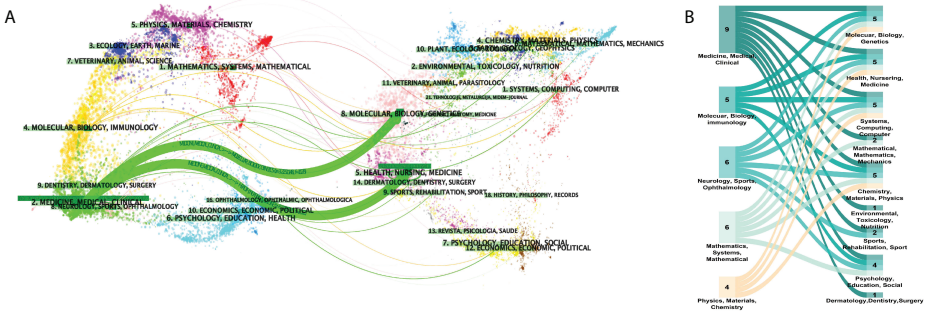


FIGURE 3
Distribution of disciplinaries. **(A)** Dual-visualization map. The left part was the targeted literature, whereas the right part was the source literature. Each dot represents one journal. On the left, there are the citing journals of this field and, on the right, lay the cited journals in this field. The waves linking to two sides mean that the publications on the journals on the left side may cite publications from the journals on the right-hand side. **(B)** Sankey diagram. The numbers on the graph show the amounts of publications in each disciplinary. Disciplinaries shown on the left are the subjects of citing literature and on the right-hand side lays the subjects of cited ones. The color represents citation relationships of corresponding disciplinaries.

followed by Dimitris Visvikis (11 articles, France), Lijun Lu (9 articles, China), Arman Rahmim (9 articles, USA), and Martina Sollini (8 articles, Italy) (Figure 6A). Among the authors with a relatively larger number of publications, Mathieu Hatt and Esther G. C. Troost have greater influence in the field, with the centrality ≥0.1 in the visualized map of our bibliometric analysis (Figure 6B).

Research of keyword

Keywords are a high-level summary and condensation of the topic in an article. Those with high frequency in different periods reflect the core issues in the field, whereas their clustering view highlights key nodes and important connections, revealing the research theme of a knowledge field and its evolution process.

Keyword co-occurrence

The visualized/visualization map constructed based on the literature includes 298 nodes, and the nodes are closely related to form a complex relationship network, indicating that the literature has a wide range of research (Supplementary Figure 2A). According to the density map, keyword nodes represented by “radiomics”, “CT”, “images”, “diagnosis”, and “features” are closer to red color (Supplementary Figure 2B), suggesting that they are the frequently occurring keywords of ^{18}F-FDG PET/CT radiomics.

Keyword clusters

The high-frequency keywords and their clusters at different times are shown in Figure 7. Analysis of keywords found that the content of the literature mostly focused on the research methodology of radiomics (keywords: positron emission tomography, FDG PET, F 18 FDG PET/CT, computed tomography, etc.), and scholars were very concerned about the core analysis content of PET/CT radiomics (keywords: texture analysis and textural feature). Data related to diagnosis and prognosis are valued in terms of evaluation indicators (keywords: heterogeneity, survival, and prognostic value). On the basis of the co-occurrence network, the log-likelihood ratio (LLR) algorithm was used to cluster the keywords. In the timeline view clusters, the Silhouette (S) value refers to the average contour value of the cluster. It is generally considered that, with an $S > 0.5$, the cluster is reasonable, whereas with an $S > 0.7$, the cluster is convincing. In our study, $S = 0.9316$, indicating that the clustering is efficient and convincing. Around 2013, scholars had just started to contact ^{18}F-FDG PET/CT radiomics research, and they were typically interested in “standardized uptake value”, “computed tomography”, and “image”; by 2016, the keywords involved in the research had changed and were mainly clustered as “reconstruction”, “artificial intelligence”, “risk stratification”, “lung adenocarcinoma”, and “lung cancer”. We can see that lung cancer is the first single disease to be cited in ^{18}F-FDG PET/CT radiomics field; since then, ^{18}F-FDG PET/CT radiomics research topics have become increasingly broad and have begun to focus on different directions, such as “hepatocellular carcinoma”, “cervical cancer”, and “prostate cancer”. The actual frequency numbers of each term of clusters can be seen in Supplementary Table 6.

Literature co-citation analysis

Normally, looking into the co-citation relationship allows us to explore the development and evolutionary dynamics of a

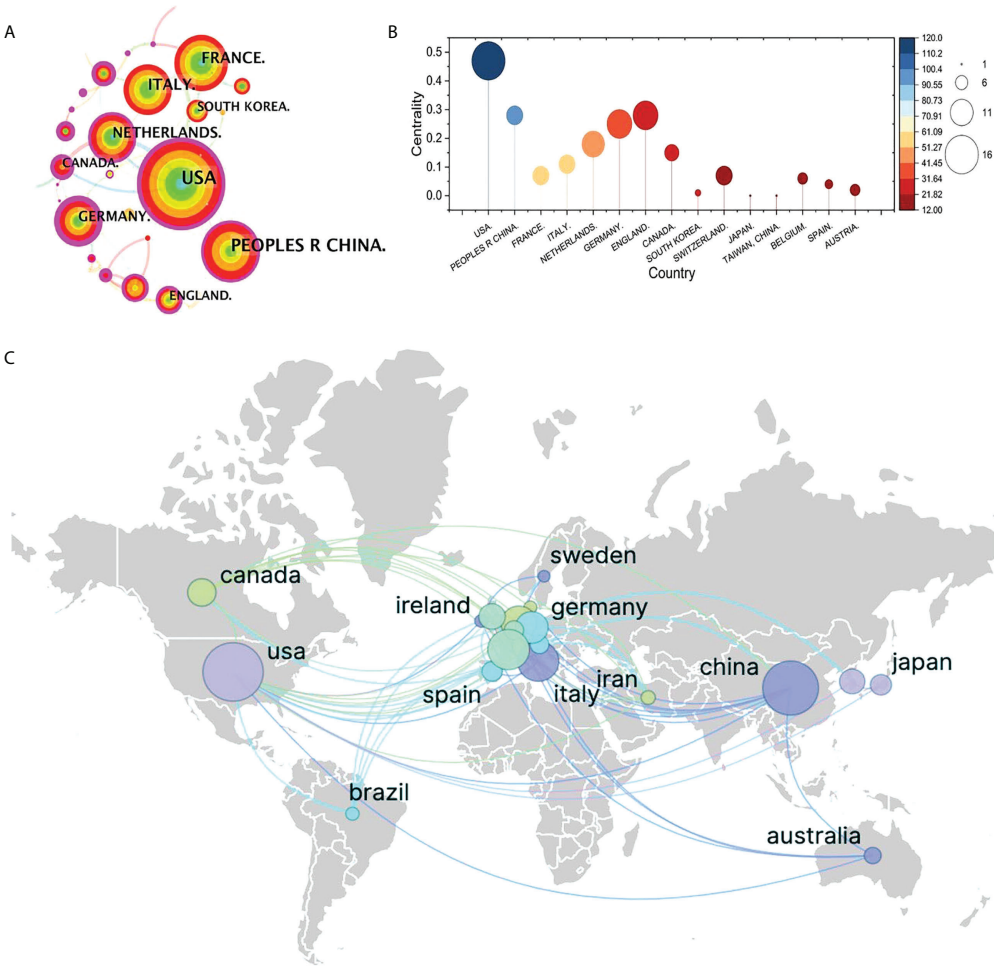


FIGURE 4
Visualization map of cooperation between countries. **(A)** Collaboration visualization map of countries of bibliometric app (CiteSpace), *N* = 641, *E* = 560 (*N* represents the number of visualization map nodes. *E* represents the number of connections). The nodes in the visualization map represent countries/regions/institutions, and the connection line represents the strength of the relationship between the two nodes. The larger the circle of the node, the greater the volume of published documents. The thicker the connection, the closer the cooperative relationship. **(B)** Top 10 most productive countries involved in ^{18}F-FDG PET/CT radiomics research. **(C)** Collaborative research relationships between countries based on VOSvieser and Scimago Graphica.

particular study (24). Each node in the visualization map represents a paper and the first author's information was listed briefly. The connection between the nodes represents the closeness in the relationship. The tighter the relationship is, the thicker the line is. The blue "#" labels show the cluster name, and the blocks of the same color are divided into studies on the same topic. Most of the studies were divided into 11 core clusters, and to be specific, radiation oncology (#3) is the subject most related to ^{18}F-FDG PET/CT radiomics research, indicating that the broad application of artificial intelligence analysis based on machine learning plays a significant role in the evaluation of tumor radiation therapy efficacy, target delineation, prognosis evaluation, etc. CT (#1) and texture analysis (#4) are the basic method or technology of medical examination required for carrying out ^{18}F-FDG PET/CT radiomics research. Imaging biomarkers (#0) are core evaluation methods of radiomics. Hepatocellular carcinoma (#6), lymphoma (#5), lung cancer (#9), and esophageal cancer (#10) are the most researched diseases at this stage. Neoadjuvant chemotherapy (#8) and prognosis (#7) indicate that scholars pay close attention to the application of ^{18}F-FDG PET/CT radiomics on efficacy assessment and prognostic analysis (Figure 8). Generally speaking, in a cluster analysis, $0 \leq Q \leq 1$ and $Q > 0.3$ indicate that there are significant differences in the divided clusters. In this study, the Q-value is 0.8764, indicating that the visualization map has a vital effect of dividing clusters. The top 12 highly cited papers are sorted by the number of

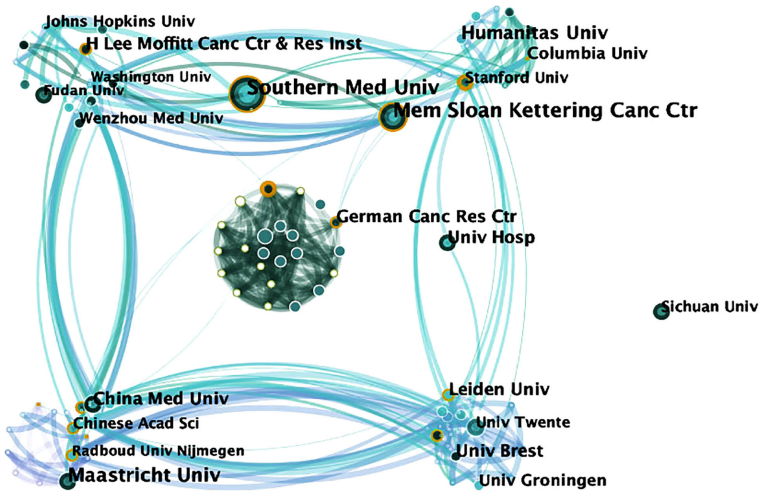


FIGURE 5
Map of active institutions from 2013 to 2021. The size of each circle is proportional to the article counts. The thickness of the curved connecting line represents the collaborative intensity between institutions.

citations, as shown in Table 2. The top 10 co-citation references were shown in Supplementary Table 7.

Citation burst analysis

Keywords with the strong citation burst can show the transfer of research frontiers in different periods and judge potential development trends and frontier research. As shown in Figure 9, *begin* and *end* indicate the start time and end time of the mutation, respectively, and *strength* is the keyword mutation strength. The higher the strength, the greater the influence. The keyword citation burst map contains the 18 most essential keywords in ^{18}F-FDG PET/CT radiomics field. The keyword with the highest strength is "F 18 FDG PET" (strength = 4.12), pointing out the theme of research. The most crucial keyword might be "radiation therapy" and "quantitative assessment" with the intensity of 3.38 and 3.16, respectively. They started to become important from 2016 to 2018. The extended duration of each keyword is no more than 3 years, suggesting that new hot topics burst rapidly, while research on a specific topic is not very persistent. Exploring the timeline view, from 2013 to 2016, the main focus was on the methodologies of ^{18}F-FDG PET/CT radiomics in the identification of tumor heterogeneity and tumor volume measurement; many studies focus on the determination of the efficacy of radiotherapy by ^{18}F-FDG PET/CT radiomics in 2016. During this period, computer-aided tools gradually became one of the research hotspots. Since 2017, many studies on a specific type of tumor have begun to emerge. On a constant basis, the ^{18}F-FDG PET/CT radiomics research of head and neck tumors, gynecological tumors, and adenocarcinoma has continued growth. Observing the citation burst of references, we find that Lambin et al. (2012) represented that literature has the highest intensity value of 18.91. The articles by Zwanenburg A (2016) and Sollini M (2017) have seen a surge in citations starting in 2019.

TABLE 1 Institutions with the highest frequency related to ^{18}F-FDG PET/CT radiomics.

Rank	Institution	Country	Frequency	Degree	Centrality
1	Memorial Sloan-Kettering Cancer Center	USA	17	11	0.11
2	University of Groningen	Netherlands	11	12	0.13
3	Leiden University	Netherlands	11	12	0.19
4	German Cancer Research Center	Germany	11	17	0.13
5	Stanford University	USA	9	9	0.19
6	Chinese Academy of Sciences	China	7	8	0.15
7	Helmholtz-Zentrum Dresden Rossendorf	Germany	7	21	0.21
8	Vrije University Amsterdam Medical Center	Germany	5	14	0.13
9	University of Michigan	USA	2	12	0.16

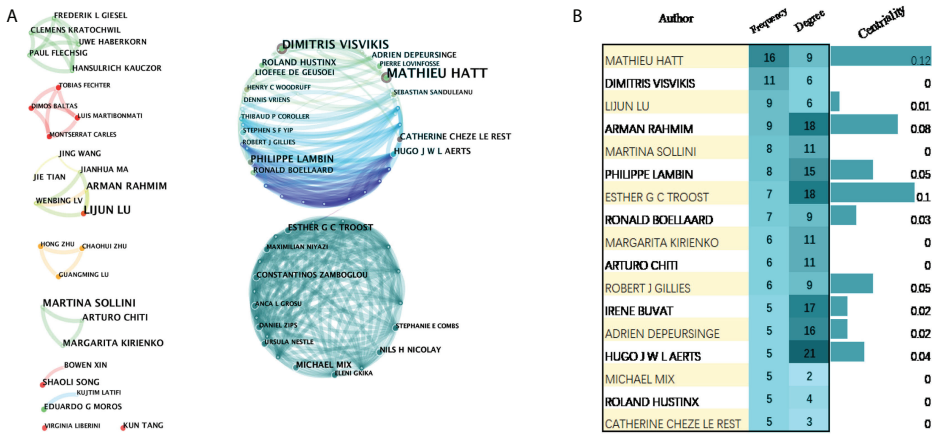


FIGURE 6
Visualization map of active authors in the field of ^{18}F-FDG PET/CT radiomics. **(A)** Cooperation of authors. **(B)** Authors with more than five publications.

Discussion

PET/CT radiomics can be defined as the transformation of PET/CT images into massive quantitative imaging features with the help of computer software and quantitative analysis using statistical and/or machine learning methods to screen out the most valuable radiomics features, so as to serve the clinic. At present, a large number of studies have focused on the application and value of various texture features in PET images, and satisfactory results have been achieved for tumors such as lymphoma, lung cancer, liver cancer, breast cancer, nasopharyngeal cancer, and cervical cancer (18, 25, 26). When the number of patients is insufficient, some studies use technical means, such as external cohort verification or even deep learning technology to obtain constructive results (27, 28). In addition, the PET radiomics signature model also showed good performance in gene prediction, assessment of tumor patient prognosis, and risk stratification (29, 30).

General data

In recent years, ^{18}F-FDG PET/CT radiomics research is getting active, and the temporal trends show continuing

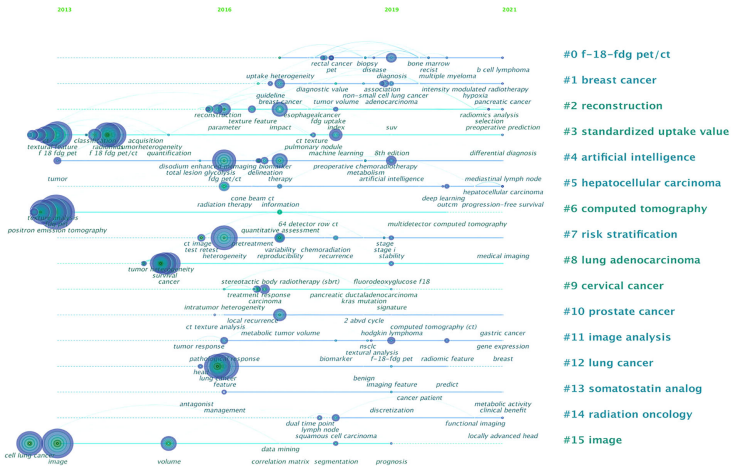


FIGURE 7
Analysis of all keywords in studies related to ^{18}F-FDG PET/CT radiomics research. In the timeline view of keyword cluster analysis, there are N = 298 and E = 598.

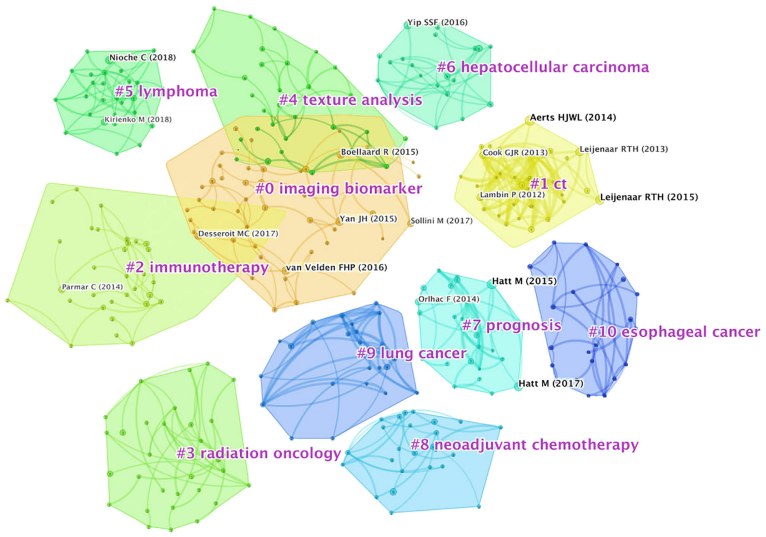


FIGURE 8
Analysis of author co-citation related to ^{18}F-FDG PET/CT radiomics research. *N* = 482, *E* = 837. The connection between the nodes represents the closeness in the relationship. The tighter the relationship is, the thicker the line is. The blue "#" labels show the cluster name, and the blocks of the same color are divided into studies on the same topic.

TABLE 2 Top 12 highly cited papers in ^{18}F-FDG PET/CT radiomics research from WoSCC.

Rank	First Authors	Article Titles	Source	Hot Paper Status	Times Cited	Document Type
1	Zwanenburg A	The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping	Radiology	Yes	613	Original Article
2	Yip SS	Applications and Limitations of Radiomics	Physics in medicine and biology	No	496	Review
3	Coroller TP	CT-Based Radiomic Signature Predicts Distant Metastasis in Lung Adenocarcinoma	Radiotherapy and oncology	No	427	Original Article
4	Nioche C	LIFEx: A Freeware for Radiomic Feature Calculation in Multimodality Imaging to Accelerate Advances in the Characterization of Tumor Heterogeneity	Cancer research	No	344	Original Article
5	Chicklore S	Quantifying Tumor Heterogeneity in ^{18}F-FDG PET/CT Imaging by Texture Analysis	European journal of nuclear medicine and molecular imaging	No	313	Review
6	Leijenaar RT	Stability of FDG-PET Radiomics Features: An Integrated Analysis of Test-Retest and Inter-Observer Variability	Acta oncologica	No	279	Original Article
7	Vallieres M	Radiomics Strategies for Risk Assessment of Tumor Failure in Head-and-Neck Cancer	Scientific reports	No	191	Original Article
8	Larue RT	Quantitative Radiomics Studies for Tissue Characterization: A Review of Technology and Methodological Procedures	British journal of radiology	No	180	Review
9	Lee G	Radiomics and Its Emerging Role in Lung Cancer Research, Imaging Biomarkers and Clinical Management: State of the Art	European journal of radiology	No	161	Original Article
10	Orlhac F	A Postreconstruction Harmonization Method for Multicenter Radiomic Studies in PET	Journal of nuclear medicine	No	142	Original Article
11	Lucia F	Prediction of Outcome Using Pretreatment ^{18}F-FDG PET/CT and MRI Radiomics in Locally Advanced Cervical Cancer Treated With Chemoradiotherapy	European journal of nuclear medicine and molecular imaging	No	117	Original Article
12	Valdora F	Rapid Review: Radiomics and Breast Cancer	Breast cancer research and treatment	No	112	Review

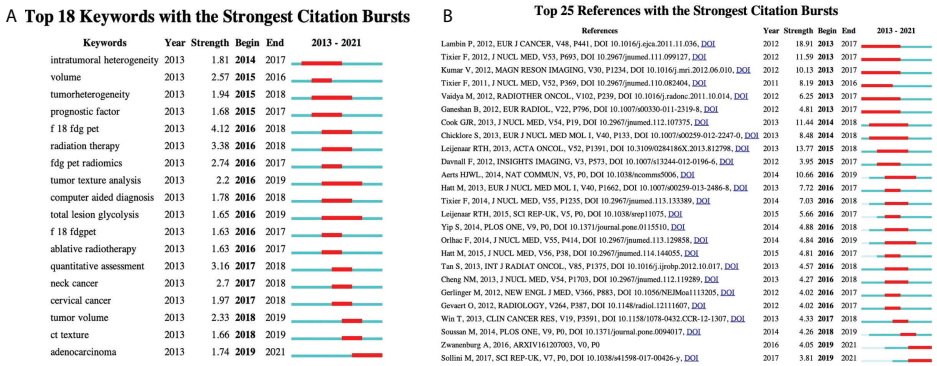


FIGURE 9
Detection of citation bursts. **(A)** Top 18 keywords with the strongest citation bursts. **(B)** Top 25 reference with the strongest citation bursts. *Begin* and *end* indicate the start time and end time of the mutation, respectively, and *strength* is the keyword mutation strength. The higher the strength, the greater the influence.

growth in the number of publications and citations, with the number reaching 146 and 302 in 2021. The reason may be that the application of PET/CT in a tumor is more and more affirmed by clinicians. As a prominent media for academic communication, 134 journals have published research on ^{18}F-FDG PET/CT radiomics. *European Journal of Nuclear Medicine and Molecular Imaging* (IF9.236), *Frontiers in Oncology* (IF6.244), and *Cancers* (IF6.639) were the most productive ones, and *Journal of Nuclear Medicine* (IF10.057) had the highest IF. We found that most of the productive journals were classified as Q1 or Q2, which suggests that the aforementioned journals published high-quality research with convincing and mature results.

As the leading force, the United States is the most prolific country with the highest centrality. We identified the United States, China, Germany, Netherlands as double-high nodes of the visualization map, which means that the authors of these countries have completed high-level research. According to the cooperation relationship, scholars from the United States, Japan, and South Korea have close cooperation in ^{18}F-FDG PET/CT radiomics, and, at the same time, American researchers also maintain close ties with researchers from European countries. Brazilian scholars, from South America, often work with scholars from Western Europe and East Asia. Scientists from China, Australia, and European countries collaborate the most. It is worth noting that the research in the two major countries in North America has a cooperation relationship with the research in the Middle East country (Iran). Another interesting point is that, even with the countries' large volume of publications and despite some countries being geographically adjacent, collaboration in East Asia was still rare and limited. The research capabilities of most developing countries in the field of ^{18}F-FDG PET/CT radiomics should be improved. Throughout the visualization map, research was primarily done by authors from European countries, North American countries, and East Asian countries. In South America, only Brazilian scientists are now paying attention to the ^{18}F-FDG PET/CT radiomics-related topics; in the Middle Eastern and African countries, only Iran was included in our analysis based on WoSCC status.

For the research institutions, Memorial Sloan-Kettering Cancer Center and Southern Medical University, Maastricht University, H. Lee Moffitt Cancer Center and Research Institute, and Humanitas University have led numerous high-quality studies. Among the top 5 institutions, the Memorial Sloan-Kettering Cancer Center was the foremost, with a frequency of 120. Among the high-impact institutions, University of Groningen and Leiden University were from Netherlands. One possible reason might be attributed to the well-developed medical device industry in Netherlands (31). Companies, such as Royal Dutch Philips Electronics Ltd., have long maintained a global leading edge in the field of medical imaging, allowing Dutch scientists to continuously access cutting-edge technologies. As is acknowledged, the concept of radiomics was just proposed by Dutch scientists Lambin et al. (32) from the Maastricht University Medical Center (Netherlands) in 2012. University Hospital (USA), the Memorial Sloan-Kettering Cancer Center (USA), and Southern Medical University (China) served as an outstanding bridge of communication in our visualization map, suggesting that multi-central innovative research in the future might be conducted under the cooperation of these institutions.

We found that nine authors published more than 10 papers, who have high academic reputations in ^{18}F-FDG PET/CT radiomics research and analog research and have contributed significantly to developments and advances. Mathieu Hatt at the University of Brest in France was the most productive author, followed by Dimitris Visvikis, one of the group members of

Mathieu, indicating that a net of collaborations centered on Mathieu and his team has been formed in this field. Mathieu was dedicated to the application of machine learning in PET/CT in oncology, and one of his studies was placed in the top 0.1% of papers in the academic field of Clinical Medicine by WoSCC (33). This hot paper standardized a set of 174 radiomic features, finding that 169 of the features were able to standardize widely, for example, mean, skewness, excess kurtosis, and a minimum of the intensity-based statistics family. Another study (34), cited 117 times, found that radiomics features of PET have higher prognostic power than usual clinical parameters and can be regarded as independent predictors of recurrence and loco-regional control of locally advanced cervical cancer.

Research hotspots and Frontiers

The core content of radiomics is extracting features from quantitative imaging through AI-based high-throughput analysis and evaluating the metabolic biological behavior of particular tissues non-invasively. The comprehensive workflow of radiomics follows these steps: 1) acquisition and reconstruction of image, 2) establishment of data sets, 3) segmentation of image, 4) extraction of feature, 5) dimension reduction, 6) construction of a predicting model, and 7) validation of models using internal and/or external data sets. What is particular is that ^{18}F-FDG PET/CT radiomics has a special attribute, distinct from other medical imaging modalities, e.g., CT radiomics, MRI radiomics, or radiomics in endoscopic examinations. Above all, when dealing with the quantitative imaging segment, PET functional imaging should be integrated with CT anatomical imaging. Then, during image acquisition, routine calibration of PET/CT equipment needs attention to avoid systematic errors caused by equipment status. We can clearly summarize the overall research contents of ^{18}F-FDG PET/CT radiomics into two aspects. One is studying the technical know-how, i.e., accurate tumor segmentation and annotation, massive feature extraction and screening, and artificial intelligence model construction. The other is a discussion on typical clinical applications in oncology, i.e., intelligent diagnosis, efficacy evaluation, prognosis, and survival prediction. For example, Vallières et al. (35) used machine learning to develop a model to predict treatment effects based on ^{18}F-FDG PET/CT images of 300 nasopharyngeal carcinoma patients and found that the developed model could predict local recurrence and distant metastasis with AUC of 0.69 and 0.86, respectively. Li et al. (36) analyzed the overall survival of 127 glioma patients using the Kaplan-Meier curve and log-rank test and concluded that there was a significant difference in predicting overall survival between high-risk and low-risk groups in their study.

Keywords that appear frequently can represent areas of focus in a given field (37), and the most cited papers are usually a concentrated expression of hot topics. Generally speaking, keywords with higher frequency of co-occurrence were radiomics, quantitative imaging, prediction, features, survival, and so on. One of the major concerns in ^{18}F-FDG PET/CT radiomics is what extract features are used over time. There is no unified standard in extracting features of radiomics used in clinical practice. To match different research aims, scholars focus on various kinds of features (18). Commonly used ones include intensity feature, shape feature, texture feature, wavelet feature, clinical feature, etc. The intensity feature is a description of the concentration of PET/CT metabolic imaging agents, and the shape feature offers an insight into morphological change. These features are usually used in conjunction with textures (commonly including first-order gray histogram features and second-order high order). Almost all of the highly cited documents in the WOS database involve texture feature analysis. Among texture indicators, the gray-level co-occurrence matrix (GLCM) is the most commonly used feature calculation matrix, followed by the gray-level run-length matrix (GLRLM) and gray-level zone matrix (GLSZM). Due to the large amount of data, one of the most important tasks before modeling is dimensionality reduction. Wavelet features enhance certain characteristics of the image based on its frequency domain information. Studies use the wavelet feature in pretreatment and dimensionality reduction steps to increase the number of data input. Lue et al. (38) pointed out that features from high-frequency wavelet components were useful for the prediction of response to therapy. In their prognostic stratification model, the high-intensity run emphasis of PET in GLRLM wavelets serves as an independent predictive factor for treatment response. Other dimensionality reduction methods mainly include principal components analysis (PCA), linear discriminant analysis (LDA), and Laplace feature mapping.

We analyzed the clustering maps, finding that the hot research disease in ^{18}F-FDG PET/CT radiomics studies is a malignant tumor and mainly contains lung cancer (45 articles), lymphoma (29 articles), hepatocellular carcinoma (26 articles), esophageal cancer (22 articles), breast cancer (19 articles), etc. suggesting that ^{18}F-FDG PET/CT radiomics is likely to be a better tool for oncology research. According to the primary site of the tumor, the application of PET/CT radiomics in different diseases was categorized as follows: (i) Head and Neck Oncology: Applications in the field of head and neck oncology focus on the precise diagnosis of recurrence of carcinoma by ^{18}F-FDG PET radiomics. A trial (39), covering 76 patients with nasopharyngeal carcinoma after treatment, showed that the AUC of PET radiomics features selection method and device cross-validation was 0.867~0.892 for distinguishing tumor recurrence from inflammatory response, which was excellent better than traditional metabolic parameters (AUC = 0.817). Wang et al. (16) performed an individualized diagnosis of tumor recurrence from radiation necrosis in glioma patients using an integrated ^{18}F-FDG PET radiomics–based model, proposing

that the AI model demonstrated good discrimination (AUC = 0.988, 95% CI: 0.975~1.000), and predictors contained in the individualized diagnosis model included the radiomics signature, the mean tumor-background ratio (TBR) of ^{18}F-FDG, the maximum TBR of ^{11}C-MET PET, and patient age. (ii) Thoracic and Chest Oncology: The research on thoracic cancer started early in the field of lung cancer and the related technologies are becoming relatively mature. With clustering keywords, i.e., survival and tumor heterogeneity, our analysis indicates that ^{18}F-FDG PET/CT radiomics is commonly used to improve the ability to distinguish benign and malignant lung lesions, and the technology is also helpful in the prognostic analysis of lung cancer patients. Radiomics features derived from ^{18}F-FDG PET/CT were associated with local control in patients with non–small cell lung cancer (NSCLC) undergoing stereotactic body radiation therapy (SBRT), and radiomics data can be used as predictors of overall survival (OS), disease-specific survival, and regional control (40). Quantitative imaging features of lung cancer, such as volume, density, and metabolic activity, have been employed to enhance interpretation and improve the prognostic value (41). As to tumor heterogeneity, the deep learning score of EGFR mutation provides a non-invasive method for identifying NSCLC patients sensitive to EGFR tyrosine kinase inhibitors or immune checkpoint inhibitors treatments (19). Another significant application of ^{18}F-FDG PET/CT radiomics in thoracic oncology is in esophageal cancer studies. Radiomics-guided ^{18}F-FDG PET/CT scans can accurately predict the response of neoadjuvant chemoradiotherapy in esophageal cancer patients (42). Combining the radiomic features and traditional parameters, ^{18}F-FDG PET/CT radiomics may also enable better stratification of patients with esophageal squamous cell carcinoma treated with neoadjuvant chemoradiotherapy and surgery into subgroups with various survival rates (43). ^{18}F-FDG PET/CT radiomics is helpful in pathological classification, differential diagnosis, and prognosis prediction of breast cancer. A significant correlation exists between imaging features and the histological type of breast cancer. These features include standard parameters such as mean standard uptake value (SUVmean), total glycolysis of lesions (TLG), metabolic tumor volume (MTV), and advanced imaging features (histogram-based and shape and size features). As for chest tumors, Cheng et al. (44) retrospectively analyzed the PET-based radiomics machine learning model to predict axillary lymph node status in early-stage breast cancer, finding that the PET-based robust machine learning model integrating the clinical characteristics can predict the pathological node status and improve the true positive and true negative rate of pathological classification. The key to differential diagnosis using PET/CT images is to examine changes in tissue metabolism and uptake, and radiomics can extract quantitative variables that cannot be visually assessed in medical images (45). A preliminary study (46) of differential diagnosis used machine learning to differentiate breast carcinoma from breast lymphoma, finding that models based on clinical, SUV, and radiomics features of ^{18}F-FDG PET/CT images have promising discriminative abilities. Lymphadenopathy is commonly found after the injection of the COVID-19 mRNA vaccine. Based on K-nearest neighbors and random forest models, Eifer et al. (47) considered that ^{18}F-FDG PET/CT radiomics may have a role in differentiating benign nodes from malignant ones. Studies also keep on trying to identify radiomic prognosis predictors from PET/CT in breast cancer patient therapeutic efficacy. As for some prognostic factors, e.g., human epidermal growth factor receptor 2 (HER2)–positive, P53 mutation, Ki-67 proliferation index, and ^{18}F-FDG PET/CT radiomics, data have the predictive and prognostic ability for personalized management. PET parameters showed stronger correlations with immunohistochemical factors and immunohistochemical subtype of breast cancer. Texture analysis indicates that HER2-positive tumors had significantly higher uptake of FDG, whereas luminal B–like/HER2$^+$ and HER2-positive non-luminal tumors also showed more regional heterogeneity than Luminal A–like tumors on breast PET image (48). In the situation of the presence of p53 mutation, the use of ^{18}F-FDG PET/CT radiomics data is believed to contribute to breast cancer management (49). Ha et al. used unsupervised clustering to figure out imaging biomarkers for estimation of intratumoral heterogeneity in locally advanced breast cancer, proposing that metabolic radiomics patterns are associated with Ki-67 expression (50). (iii) Abdominopelvic Oncology: The highly cited papers of abdominal tumor studies focused on cervical cancer, verifying that the radiomics model has higher prognostic power than usual clinical parameters (51), and Lucia et al. (26) put forward that entropy GLCM and GLRLM from functional imaging PET might be independent predictors of recurrence and loco-regional control in cervical cancer patients. Liu et al. (52) developed a predictive model by including 351 patients with stages IB to IIA squamous cell carcinoma of the uterine cervix and found that squamous cell carcinoma antigen level and pelvic lymph node SUVmax were independent predictors of pelvic lymph node metastasis, and the resulting line graph showed high sensitivity (70.5%), specificity (94.4%), and positive predictive value (93.9%). Other research on abdominopelvic cavity demonstrates that ^{18}F-FDG PET/CT texture analysis can effectively differentiate renal cell carcinoma from renal lymphoma, and differential response after the first-line treatment of colorectal cancer patients (53, 54). On the basis of citation bursts analysis, we list some recent research status of different clinical applications of ^{18}F-FDG PET/CT radiomics studies below: (i). The most crucial keyword "^{18}F-FDG PET" started to become important in 2016, indicating that medical workers have started to place emphasis on clinical applications of ^{18}F-FDG PET/CT radiomics ever since 2016. (ii) Intelligent diagnosis (keyword: computer-aided diagnosis, burst: 2016~2018), response evaluation (keyword: quantitative

assessment, burst: 2017~2018), and survival prediction for cancer prognosis (keyword: prognostic factor, burst: 2015~2017) are regarded as typical applications of radiomics. One of the most cited studies concerned about whether the composite textures from the combination of FDG-PET and MR imaging information could quantitatively identify aggressive tumors at diagnosis (55). It verified that assessing lung metastasis risk of soft-tissue sarcomas through ^{18}F-FDG PET/CT radiomics could improve patient outcomes. In some respects, response assessment after treatment can provide a reliable basis for predicting expected survival. The most cited study (56) on efficacy assessment and survival prediction reported that baseline ^{18}F-FDG PET scan uptake in NSCLC, showing abnormal texture as measured by coarseness, contrast, and busyness, is associated with nonresponse to chemoradiotherapy by response evaluation criteria in solid tumors and with poorer prognosis. (iii) The ^{18}F-FDG PET/CT radiomics analysis in adenocarcinoma can be speculated to become an emerging academic trend in clinical practice. During the last 3 years, the keyword "adenocarcinoma" was the most concerned and frequently cited. Adenocarcinoma research involves a primary focus on identification and image segmentation, identification of heterogeneity, prediction of recurrence and metastasis, evaluation of treatment efficacy, and prediction of survival (57–59). In clinical practice, the research puts out that the accuracy of the PET/CT imaging analyzed by radiomics in pathological pattern and staging is close to or even equal to the "gold standard" of pathological biopsy (60). We can speculate that ^{18}F-FDG PET/CT radiomic is providing promising performance to move the post-operative pathological characterization analysis of adenocarcinoma forward to pre-operation.

In the co-citation map, it is worth mentioning that many scholars have a co-citation relationship in studies related to imaging biomarkers, texture analysis, and immunotherapy simultaneously. Texture analysis has long been used in the classification of correlations between imaging parameters of glucose metabolism and the expression levels of genomic biomarkers from cancers (55, 61). Furthermore, there are pros and cons in a variety of ^{18}F-FDG PET/CT radiomic research. By convention, traditional machine learning methods, represented by the random forest, decision tree, and regression algorithms, are most commonly used to segment images. Such kind of manual segmentation is easy to operate, and we can obtain intuitive images for radiomics analysis. However, Ding et al. (37) pointed out that these algorithms mostly require a huge amount of matrix manipulations. It takes too much time and the results are unavailable to repeat. To note, the landmark literature of our citation burst especially put emphasis on the reproducibility of FDG-PET radiomics features, which indicates that it is of great necessity to control the variability between test-retest and inter-observer to make sure that all features are robustly measured. Learning-based algorithms are developed to solve this problem. Tixier et al. (62) studied the reproducibility of tumor uptake heterogeneity characterization through the fuzzy locally adaptive Bayesian (FLAB) algorithm; meanwhile, they used parameters on local and regional scales to replace simple SUV measurements, attaining similar or better reproducibility than simple SUV measurements. Learning-based algorithms have made substantive progress. Notwithstanding, it is important to pay attention that the results of these algorithms remain to be interpreted carefully, for most of the machine learning evaluations included insufficient numbers of patients and few studies have external validation on those developed AI models. Accessibility of data and software should be improved for researchers to expand the sample to verify the previous research results (27).

Strengths and limitations

Although there has been a surge in bibliometric studies, knowledge visualization maps of PET/CT imaging are still very rare. We often find limitations such as unprofessional literature retrieval, lack of quality control in map interpretation, and incomplete reporting of original data, which affected the credibility of literature results and conclusions. Note that this study carried out quality control through interviews with librarians, and we listened to the literature retrieval strategies and suggestions provided by professionals. During the literature retrieval process, we included literature as comprehensive and complete as possible. When interpreting the co-occurrence analysis results, we invited three nuclear medicine physicians to offer suggestions to improve the reliability of the study. After a series of quality control procedures, it is reasonable to believe that the results of this study truthfully reflect the current research hotspots and frontiers of ^{18}F-FDG PET/CT radiomics, and the research conclusions are true and reliable. This study is the first comprehensive literature analysis focusing on PET/CT radiomics research with a specific imaging technique. We hope that this study can provide a practical reference for scholars to use ^{18}F-FDG imaging tracer to carry out PET/CT radiomics practice.

At the same time, we also recognize that there are some shortcomings and deficiencies in this research. For example, WoSCC was the only database to collect literature, and PubMed, Embase, CNKI, or other databases would be further needed in a comprehensive analysis. It is worth noting that WoSCC is the most commonly used database in bibliometrics research. Due to the limited number of documents available from the WoSCC, our study is likely not to contain an exhaustive number of records. Additionally, our research is conducted on an annual basis, so studies published in the first half of 2022 were not included, which may lead to missing some real-time hot issues.

Conclusions

In conclusion, the current research status revealed that ^{18}F-FDG PET/CT radiomics research has great development potential, with an increasing number publications and citations. So far, it has received significant attention in intelligent diagnosis, response evaluation, and survival prediction for cancer prognosis. *European Journal of Nuclear Medicine and Molecular Imaging*, *Frontiers in Oncology*, and *Cancers* were considered the top journals. United States, China, Germany, and Netherlands contribute a lot and have played vital roles in developing and expanding this field. The hot research topics in ^{18}F-FDG PET/CT radiomics studies are malignant tumors. It is also necessary to focus on ensuring the reproducibility of ^{18}F-FDG-PET radiomics features. Adenocarcinoma was the most concerned and frequently cited research direction. It is expected to become the frontier and development trend in the future.

For the first time, this study comprehensively screened literature associated with ^{18}F-FDG PET/CT radiomics and provided a theoretical basis for the field of it, which may benefit researchers to deeply understand the current status of the field and encourage them to engage in the hotspots and frontiers. ^{18}F-FDG PET/CT radiomics provides clinicians with a more accurate evaluation of treatment outcomes and more meaningful prognostic risk scores. It is foreseeable that using ^{18}F-FDG PET/CT radiomics is expected to rapidly advance the development of molecular imaging on the basis of existing mature technology and further integrate nuclear medicine findings into clinical diagnosis and treatment aid decision-making.

Data availability statement

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

Author contributions

XL conceived and designed the structure of this manuscript. XL, XH, XY, PL, CG, and YW wrote the paper. JC, PW, and DL revised the paper. All authors contributed to the article and approved the submitted version.

Funding

This study was funded by the National Natural Science Foundation of the People's Republic of China, NSFC (grant number: 81571712), Zunyi Medical College Research Start Fund 2018ZYFY03, and QianKeHe platform talents [2017] (grant number: 5733-035).

Acknowledgments

The authors would like to express their appreciation to Prof. Chaomei Chen, who invented CiteSpace and made it open for use. We'd also like to thanks Lecturer Jimei Cai (the information department of Zunyi Medical University Library, Guizhou Province, China) for her kindly receiving our interview on the strategy of data searching.

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.965773/full#supplementary-material

References

1. Shangguan C, Gan G, Zhang J, Wu J, Miao Y, Zhang M, et al. Cancer-associated fibroblasts enhance tumor ^{18}F-FDG uptake and contribute to the intratumor heterogeneity of PET-CT. *Theranostics* (2018) 8(5):1376–88. doi: 10.7150/thno.22717

2. Aide N, Hicks RJ, Le Tourneau C, Lheureux S, Fanti S, Lopci E. FDG PET/CT for assessing tumour response to immunotherapy: Report on the EANM symposium on immune modulation and recent review of the literature. *Eur J Nucl Med Mol Imaging* (2019) 46(1):238–50. doi: 10.1007/s00259-018-4171-4

3. Cho SY, Huff DT, Jeraj R, Albertini MR. FDG PET/CT for assessment of immune therapy: Opportunities and understanding pitfalls. *Semin Nucl Med* (2020) 50(6):518–31. doi: 10.1053/j.semnuclmed.2020.06.001

4. Holman BL, Tumeh SS. Single-photon emission computed tomography (SPECT). *Appl potential JAMA* (1990) 263(4):561–4. doi: 10.1001/jama.1990.03440040100036

5. Berger A. How does it work? Positron emission tomography. *BMJ* (2003) 326 (7404):1449. doi: 10.1136/bmj.326.7404.1449

6. DeJohn CR, Grant SR, Seshadri M. Application of machine learning methods to improve the performance of ultrasound in head and neck oncology: A literature review. *Cancers (Basel)* (2022) 14:665. doi: 10.3390/cancers14030665

7. van Dijk LV, Fuller CD. Artificial intelligence and radiomics in head and neck cancer care: Opportunities, mechanics, and challenges. *Am Soc Clin Oncol Educ Book* (2021) 41:1–11. doi: 10.1200/EDBK_320951

8. Zhai TT, van Dijk LV, Huang BT, Lin ZX, Ribeiro CO, Brouwer CL, et al. Improving the prediction of overall survival for head and neck cancer patients using image biomarkers in combination with clinical parameters. *Radiother Oncol* (2017) 124:256–62. doi: 10.1016/j.radonc.2017.07.013

9. Wei G, Jiang P, Tang Z, Qu A, Deng X, Guo F, et al. MRI Radiomics in overall survival prediction of local advanced cervical cancer patients tread by adjuvant chemotherapy following concurrent chemoradiotherapy or concurrent chemoradiotherapy alone. *Magn Reson Imaging* (2022) 91:81–90. doi: 10.1016/j.mri.2022.05.019

10. Liang X, Yu X, Gao T. Machine learning with magnetic resonance imaging for prediction of response to neoadjuvant chemotherapy in breast cancer: A systematic review and meta-analysis. *Eur J Radiol* (2022) 150:110247. doi: 10.1016/j.ejrad.2022.110247

11. Lee S, Choi Y, Seo MK, Jang J, Shin NY, Ahn KJ, et al. Magnetic resonance imaging-based radiomics for the prediction of progression-free survival in patients with nasopharyngeal carcinoma: A systematic review and meta-analysis. *Cancers (Basel)* (2022) 14(3):653. doi: 10.3390/cancers14030653

12. Bhandari A, Ibrahim M, Sharma C, Liong R, Gustafson S, Prior M. CT-based radiomics for differentiating renal tumours: A systematic review. *Abdom Radiol (NY)* (2021) 46(5):2052–63. doi: 10.1007/s00261-020-02832-9

13. Li X, Jiang F, Guo Y, Jin Z, Wang Y. Computer-aided diagnosis of gastrointestinal stromal tumors: A radiomics method on endoscopic ultrasound image. *Int J Comput Assist Radiol Surg* (2019) 14(10):1635–45. doi: 10.1007/s11548-019-01993-3

14. Hu Z, Yang Z, Lafata KJ, Yin FF, Wang C. A radiomics-boosted deep-learning model for COVID-19 and non-COVID-19 pneumonia classification using chest x-ray images. *Med Phys* (2022) 49(5):3213–22. doi: 10.1002/mp.15582

15. Kang F, Mu W, Gong J, Wang S, Li G, Li G, et al. Integrating manual diagnosis into radiomics for reducing the false positive rate of ^{18}F-FDG PET/CT diagnosis in patients with suspected lung cancer. *Eur J Nucl Med Mol Imaging* (2019) 46(13):2770–9. doi: 10.1007/s00259-019-04418-0

16. Wang K, Qiao Z, Zhao X, Li X, Wang X, Wu T, et al. Individualized discrimination of tumor recurrence from radiation necrosis in glioma patients using an integrated radiomics-based model. *Eur J Nucl Med Mol Imaging* (2020) 47 (6):1400–11. doi: 10.1007/s00259-019-04604-0

17. Zhang J, Zhao X, Zhao Y, Zhang J, Zhang Z, Wang J, et al. Value of pre-therapy ^{18}F-FDG PET/CT radiomics in predicting EGFR mutation status in patients with non-small cell lung cancer. *Eur J Nucl Med Mol Imaging* (2020) 47 (5):1137–46. doi: 10.1007/s00259-019-04592-1

18. Li P, Wang X, Xu C, Liu C, Zheng C, Fulham MJ, et al. ^{18}F-FDG PET/CT radiomic predictors of pathologic complete response (pCR) to neoadjuvant chemotherapy in breast cancer patients. *Eur J Nucl Med Mol Imaging* (2020) 47 (5):1116–26. doi: 10.1007/s00259-020-04684-3

19. Mu W, Jiang L, Zhang J, Shi Y, Gray JE, Tunali I, et al. Non-invasive decision support for NSCLC treatment using PET/CT radiomics. *Nat Commun* (2020) 11 (1):5228. doi: 10.1038/s41467-020-19116-x

20. Zaucha JM, Chauvie S, Zaucha R, Biggii A, Gallamini A. The role of PET/CT in the modern treatment of Hodgkin lymphoma. *Cancer Treat Rev* (2019) 77:44–56. doi: 10.1016/j.ctrv.2019.06.002

21. Lovinfosse P, Polus M, Van Daele D, Martinive P, Daenen F, Hatt M, et al. FDG PET/CT radiomics for predicting the outcome of locally advanced rectal cancer. *Eur J Nucl Med Mol Imaging* (2018) 45(3):365–75. doi: 10.1007/s00259-017-3855-5

22. Mattonen SA, Davidzon GA, Benson J, Leung ANC, Vasanawala M, Horng G, et al. Bone marrow and tumor radiomics at ^{18}F-FDG PET/CT: Impact on outcome prediction in non-small cell lung cancer. *Radiology* (2019) 293(2):451–9. doi: 10.1148/radiol.2019190357

23. Khan AS, Ur Rehman S, Ahmad S, AlMaimouni YK, Alzamil MAS, Dummer PMH. Five decades of the international endodontic journal: Bibliometric overview 1967-2020. *Int Endod J* (2021) 54(10):1819–39. doi: 10.1111/iej.13595

24. Huang L, Xu G, He J, Tian H, Zhou Z, Huang F, et al. Bibliometric analysis of functional magnetic resonance imaging studies on acupuncture analgesia over the past 20 years. *J Pain Res* (2021) 14:3773–89. doi: 10.2147/JPR.S340961

25. Hatt M, Tixier F, Pierce L, Kinahan PE, Le Rest CC, Visvikis D. Characterization of PET/CT images using texture analysis: The past, the present… any future? *Eur J Nucl Med Mol Imaging* (2017) 44(1):151–65. doi: 10.1007/s00259-016-3427-0

26. Lucia F, Visvikis D, Vallières M, Desseroit MC, Miranda O, Robin P, et al. External validation of a combined PET and MRI radiomics model for prediction of recurrence in cervical cancer patients treated with chemoradiotherapy. *Eur J Nucl Med Mol Imaging* (2019) 46(4):864–77. doi: 10.1007/s00259-018-4231-9

27. Welch ML, McIntosh C, Haibe-Kains B, Milosevic MF, Wee L, Dekker A, et al. Vulnerabilities of radiomic signature development: The need for safeguards. *Radiother Oncol* (2019) 130:2–9. doi: 10.1016/j.radonc.2018.10.027

28. Orlhac F, Boughdad S, Philippe C, Stalla-Bourdillon H, Nioche C, Champion L, et al. A postreconstruction harmonization method for multicenter radiomic studies in PET. *J Nucl Med* (2018) 59(8):1321–8. doi: 10.2967/jnumed.117.199935

29. Li X, Yin G, Zhang Y, Dai D, Liu J, Chen P, et al. Predictive power of a radiomic signature based on ^{18}F-FDG PET/CT images for EGFR mutational status in NSCLC. *Front Oncol* (2019) 9:1062. doi: 10.3389/fonc.2019.01062

30. Lim CH, Cho YS, Choi JY, Lee KH, Lee JK, Min JH, et al. Imaging phenotype using ^{18}F-fluorodeoxyglucose positron emission tomography-based radiomics and genetic alterations of pancreatic ductal adenocarcinoma. *Eur J Nucl Med Mol Imaging* (2020) 47(9):2113–22. doi: 10.1007/s00259-020-04698-x

31. Zheng J, Zhou R, Meng B, Li F, Liu H, Wu X. Knowledge framework and emerging trends in intracranial aneurysm magnetic resonance angiography: A scientometric analysis from 2004 to 2020. *Quant Imaging Med Surg* (2021) 11 (5):1854–69. doi: 10.21037/qims-20-729

32. 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

33. Zwanenburg A, Vallières M, Abdalah MA, Aerts HJWL, Andrearczyk V, Apte A, et al. The image biomarker standardization initiative: Standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology* (2020) 295(2):328–38. doi: 10.1148/radiol.2020191145

34. Lucia F, Visvikis D, Desseroit MC, Miranda O, Malhaire JP, Robin P, et al. Prediction of outcome using pretreatment ^{18}F-FDG PET/CT and MRI radiomics in locally advanced cervical cancer treated with chemoradiotherapy. *Eur J Nucl Med Mol Imaging* (2018) 45(5):768–86. doi: 10.1007/s00259-017-3898-7

35. Vallières M, Kay-Rivest E, Perrin LJ, Liem X, Furstoss C, Aerts HJWL, et al. Radiomics strategies for risk assessment of tumour failure in head-and-neck cancer. *Sci Rep* (2017) 7(1):10117. doi: 10.1038/s41598-017-10371-5

36. Li L, Mu W, Wang Y, Liu Z, Liu Z, Wang Y, et al. A non-invasive radiomic method using ^{18}F-FDG PET predicts isocitrate dehydrogenase genotype and prognosis in patients with glioma. *Front Oncol* (2019) 9:1183. doi: 10.3389/fonc.2019.01183

37. Ding H, Wu Cs, Liao N, Zhan Q, Sun W, Huang Y, et al. Radiomics in oncology: A 10-year bibliometric analysis. *Front Oncol* (2021) 11:689802. doi: 10.3389/fonc.2021.689802

38. Lue KH, Wu YF, Liu SH, Hsieh TC, Chuang KS, Lin HH, et al. Intratumor heterogeneity assessed by ^{18}F-FDG PET/CT predicts treatment response and survival outcomes in patients with Hodgkin lymphoma. *Acad Radiol* (2020) 27 (8):e183–92. doi: 10.1016/j.acra.2019.10.015

39. Du D, Feng H, Lv W, Ashrafinia S, Yuan Q, Wang Q, et al. Machine learning methods for optimal radiomics-based differentiation between recurrence and inflammation: Application to nasopharyngeal carcinoma post-therapy PET/CT images. *Mol Imaging Biol* (2020) 22(3):730–8. doi: 10.1007/s11307-019-01411-9

40. Oikonomou A, Khalvati F, Tyrrell PN, Haider MA, Tarique U, Jimenez-Juan L, et al. Radiomics analysis at PET/CT contributes to prognosis of recurrence and survival in lung cancer treated with stereotactic body radiotherapy. *Sci Rep* (2018) 8(1):4003. doi: 10.1038/s41598-018-22357-y

41. Manafi-Farid R, Karamzade-Ziarati N, Vali R, Mottaghy FM, Beheshti M. 2-[^{18}F] FDG PET/CT radiomics in lung cancer: An overview of the technical aspect and its emerging role in management of the disease. *Methods* (2021) 188:84–97. doi: 10.1016/j.ymeth.2020.05.023

42. Beukinga RJ, Poelmann FB, Kats-Ugurlu G, Viddeleer AR, Boellaard R, De Haas RJ, et al. Prediction of non-response to neoadjuvant chemoradiotherapy in esophageal cancer patients with ^{18}F-FDG PET radiomics based machine learning classification. *Diagnostics (Basel)* (2022) 12(5):1070. doi: 10.3390/diagnostics12051070

43. Chen YH, Lue KH, Chu SC, Chang BS, Wang LY, Liu DW, et al. Combining the radiomic features and traditional parameters of ^{18}F-FDG PET with clinical profiles to improve prognostic stratification in patients with esophageal squamous

cell carcinoma treated with neoadjuvant chemoradiotherapy and surgery. *Ann Nucl Med* (2019) 33(9):657–70. doi: 10.1007/s12149-019-01380-7

44. Cheng J, Ren C, Liu G, Shui R, Zhang Y, Li J, et al. Development of high-resolution dedicated PET-based radiomics machine learning model to predict axillary lymph node status in early-stage breast cancer. *Cancers (Basel)* (2022) 14 (4):950. doi: 10.3390/cancers14040950

45. Mayerhoefer ME, Materka A, Langs G, Häggström I, Szczypiński P, Gibbs P, et al. Introduction to radiomics. *J Nucl Med* (2020) 61(4):488–95. doi: 10.2967/jnumed.118.222893

46. Ou X, Zhang J, Wang J, Pang F, Wang Y, Wei X, et al. Radiomics based on ^{18} f-FDG PET/CT could differentiate breast carcinoma from breast lymphoma using machine-learning approach: A preliminary study. *Cancer Med* (2020) 9(2):496–506. doi: 10.1002/cam4.2711

47. Eifer M, Pinian H, Klang E, Alhoubani Y, Kanana N, Tau N, et al. FDG PET/CT radiomics as a tool to differentiate between reactive axillary lymphadenopathy following COVID-19 vaccination and metastatic breast cancer axillary lymphadenopathy: A pilot study. *Eur Radiol* (2022) 32(9):5921–29. doi: 10.1007/s00330-022-08725-3

48. Moscoso A, Ruibal Á, Domínguez-Prado I, Fernández-Ferreiro A, Herranz M, Albaina L, et al. Texture analysis of high-resolution dedicated breast ^{18}F-FDG PET images correlates with immunohistochemical factors and subtype of breast cancer. *Eur J Nucl Med Mol Imaging* (2018) 45(2):196–206. doi: 10.1007/s00259-017-3830-1

49. Acar E, Turgut B, Yiğit S, Kaya G. Comparison of the volumetric and radiomics findings of ^{18}F-FDG PET/CT images with immunohistochemical prognostic factors in local/locally advanced breast cancer. *Nucl Med Commun* (2019) 40(7):764–72. doi: 10.1097/MNM.0000000000001019

50. Ha S, Park S, Bang JI, Kim EK, Lee HY. Metabolic radiomics for pretreatment ^{18}F-FDG PET/CT to characterize locally advanced breast cancer: Histopathologic characteristics, response to neoadjuvant chemotherapy, and prognosis. *Sci Rep* (2017) 7(1):1556. doi: 10.1038/s41598-017-01524-7

51. Kim BS, Kang J, Jun S, Kim H, Pak K, Kim GH, et al. Association between immunotherapy biomarkers and glucose metabolism from f-18 FDG PET. *Eur Rev Med Pharmacol Sci* (2020) 24(16):8288–95. doi: 10.26355/eurrev_202008_22625

52. Liu S, Feng Z, Zhang J, Ge H, Wu X, Song S. A novel 2-deoxy-2-fluorodeoxyglucose (^{18}F-FDG) positron emission tomography/computed tomography (PET/CT)-based nomogram to predict lymph node metastasis in early stage uterine cervical squamous cell cancer. *Quant Imaging Med Surg* (2021) 11(1):240–8. doi: 10.21037/qims-20-348

53. Zhu S, Xu H, Shen C, Wang Y, Xu W, Duan S, et al. Differential diagnostic ability of ^{18}F-FDG PET/CT radiomics features between renal cell carcinoma and renal lymphoma. *Q J Nucl Med Mol Imaging* (2021) 65(1):72–8. doi: 10.23736/S1824-4785.19.03137-6

54. Van Helden EJ, Vacher YJL, van Wieringen WN, van Velden FHP, Verheul HMW, Hoekstra OS, et al. Radiomics analysis of pre-treatment [^{18}F]FDG PET/CT for patients with metastatic colorectal cancer undergoing palliative systemic treatment. *Eur J Nucl Med Mol Imaging* (2018) 45(13):2307–17. doi: 10.1007/s00259-018-4100-6

55. Vallières M, Freeman CR, Skamene SR, El Naqa I. A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities. *Phys Med Biol* (2015) 60(14):5471–96. doi: 10.1088/0031-9155/60/14/5471

56. Leijenaar RT, Carvalho S, Velazquez ER, van Elmpt WJ, Parmar C, Hoekstra OS, et al. Stability of FDG-PET radiomics features: An integrated analysis of test-retest and inter-observer variability. *Acta Oncol* (2013) 52(7):1391–7. doi: 10.3109/0284186X.2013.812798

57. Zhang Y, Cheng C, Liu Z, Wang L, Pan G, Sun G, et al. Radiomics analysis for the differentiation of autoimmune pancreatitis and pancreatic ductal adenocarcinoma in ^{18} f-FDG PET/CT. *Med Phys* (2019) 46(10):4520–30. doi: 10.1002/mp.13733

58. Hu Y, Zhao X, Zhang J, Han J, Dai M. Value of ^{18}F-FDG PET/CT radiomic features to distinguish solitary lung adenocarcinoma from tuberculosis. *Eur J Nucl Med Mol Imaging* (2021) 48(1):231–40. doi: 10.1007/s00259-020-04924-6

59. Xing H, Hao Z, Zhu W, Sun D, Ding J, Zhang H, et al. Preoperative prediction of pathological grade in pancreatic ductal adenocarcinoma based on ^{18}F-FDG PET/CT radiomics. *EJNMMI Res* (2021) 11(1):19. doi: 10.1186/s13550-021-00760-3

60. Wang K, Lu X, Zhou H, Gao Y, Zheng J, Tong M, et al. Deep learning radiomics of shear wave elastography significantly improved diagnostic performance for assessing liver fibrosis in chronic hepatitis b: A prospective multicentre study. *Gut* (2019) 68(4):729–41. doi: 10.1136/gutjnl-2018-316204

61. Mu W, Jiang L, Shi Y, Tunali I, Gray JE, Katsoulakis E, et al. Non-invasive measurement of PD-L1 status and prediction of immunotherapy response using deep learning of PET/CT images. *J Immunother Cancer* (2021) 9(6):e002118. doi: 10.1136/jitc-2020-002118

62. Tixier F, Hatt M, Le Rest CC, Le Pogam A, Corcos L, Visvikis D. Reproducibility of tumor uptake heterogeneity characterization through textural feature analysis in ^{18}F-FDG PET. *J Nucl Med* (2012) 53(5):693–700. doi: 10.2967/jnumed.111.099127

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 26 October 2022
DOI 10.3389/fonc.2022.1017737

OPEN ACCESS

EDITED BY
Bahram Mohajer,
Johns Hopkins Medicine, United States

REVIEWED BY
Jianguo Lin,
Jiangsu Institute of Nuclear Medicine, China
Hua Zhu,
Beijing Cancer Hospital, China

*CORRESPONDENCE
Pan Wang
1298178828@qq.com
Jiong Cai
jiong_cai@163.com

†These authors have contributed equally to this work

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

RECEIVED 12 August 2022
ACCEPTED 10 October 2022
PUBLISHED 26 October 2022

CITATION
Liang Z, Hu X, Hu H, Wang P and Cai J (2022) Novel small ^{99m}Tc-labeled affibody molecular probe for PD-L1 receptor imaging.
Front. Oncol. 12:1017737.
doi: 10.3389/fonc.2022.1017737

Novel small ^{99m}Tc-labeled affibody molecular probe for PD-L1 receptor imaging

Zhigang Liang†, Xianwen Hu†, Hongyu Hu, Pan Wang* and Jiong Cai*

Department of Nuclear Medicine, Affiliated Hospital of Zunyi Medical University, Zunyi, Guizhou, China

Objective: The *in vivo* imaging of programmed death ligand 1 (PD-L1) can monitor changes in PD-L1 expression and guide programmed death 1 (PD-1) or PD-L1-targeted immune checkpoint therapy. A ^{99m}Tc-labeled affibody molecular probe targeting the PD-L1 receptor was prepared and evaluated its tracing effect in PD-L1-overexpressing colon cancer.

Methods: The PD-L1 affibody was prepared by genetic recombineering. The ^{99m}Tc labeling of the affibody was achieved by sodium glucoheptonate and an $SnCl_2$ labeling system. The labeling rate, radiochemical purity, and stability *in vitro* were determined by instant thin-layer chromatography; MC38-B7H1 (PD-L1-positive) and MC38 (PD-L1-negative) colon cancer cells were used to evaluate its affinity to PD-L1 by cell-binding experiments. The biodistribution of the ^{99m}Tc-labeled affibody molecular probe was then determined in C57BL/6J mice bearing MC38-B7H1 tumors, and tumor targeting was assessed in C57BL/6J mice with MC38-B7H1, MC38 double xenografts.

Result: The nondecayed corrected yield of the ^{99m}Tc-PD-L1 affibody molecular probe was 95.95% ± 1.26%, and showed good stability both in phosphate-buffered saline (PBS) and fetal bovine serum within 6 h. The affinity of the ^{99m}Tc-PD-L1 affibody molecular probe for cell-binding assays was 10.02 nmol/L. Single photon emission–computed tomography imaging showed a rapid uptake of the tracer in PD-L1-positive tumors and very little tracer retention in PD-L1-negative control tumors. The tracer was significantly retained in the kidneys and bladder, suggesting that it is mainly excreted through the urinary system. Heart, liver, lung, and muscle tissue showed no significant radioactive retention. The biodistribution *in vitro* also showed significant renal retention, a small amount of uptake in the thyroid and gastrointestinal tract, and rapid blood clearance, and the tumor-to-blood radioactivity uptake ratio peaked 120 min after drug injection.

Conclusion: The ^{99m}Tc-PD-L1 affibody molecular probe that we prepared can effectively target to PD-L1-positive tumors imaging *in vivo*, and clear in blood

quickly, with no obvious toxic side effects, which is expected to become a new type of tracer for detecting PD-L1 expression in tumors.

KEYWORDS

programmed death receptor-ligand 1, affibody, SPECT imaging, colon cancer, ^{99m}Tc

Introduction

A blockade targeting PD-1/programmed death ligand 1 (PD-L1) is one of the most promising cancer treatments in cancer therapy, but not all cases respond to these drugs (1–3). The molecular imaging of the immune checkpoint receptor PD-1 and its ligand PD-L1 is receiving increasing attention as a strategy to guide and monitor PD-1/PD-L1-targeted immune checkpoint therapy (4). Compared to the long imaging times of radionuclide-labeled monoclonal antibody (mAb) tracers due to slow blood clearance, low-molecular-weight ligands, such as adherends, nanobodies, or peptides, tend to show faster blood clearance and can be labeled with radioisotopes with shorter half-life, allowing same-day imaging (5). Affibody probes targeting the human epidermal growth factor receptor 2 (HER2) and epidermal growth factor receptor (EGFR) have demonstrated rapid blood clearance and good imaging *in vivo* targeting properties in clinical and preclinical studies (6–12). González Trotter et al. (13) reported the first affibody ligand ^{18}F-AIF-NOTA-$Z_{PD\text{-}L1\_1}$ for detecting PD-L1 expression, and PET imaging showed a good targeting of LOX tumors with high PD-L1 expression.

^{99m}Tc is produced by a molybdenum technetium generator, has a suitable gamma ray energy (140.5 KeV) with a short physical half-life (approximately 6 h), and is the most commonly used and ideal radionuclide tracer for single-photon emission computed tomography (SPECT) imaging. In the present study, we developed a new small ^{99m}Tc-labeled PD-L1 affibody molecular probe (^{99m}Tc-PDA) for SPECT imaging and evaluated its binding properties *in vitro* and *in vivo*, biodistribution, and targeting properties in PD-L1 receptor-positive tumor models. The results demonstrated that the molecular probe has rapid blood clearance and high *in vivo* targeting specificity in PD-L1-overexpressing xenograft tumors.

Materials and methods

Materials

Sodium glucoheptonate dihydrate was obtained from TCI (Shanghai, China), dithiothreitol (DTT) from Beijing Xinjingke Biotechnology Co., Ltd. (Beijing, China), tin (II)-chloride dehydrate (SnCl2) and hydroxyethyl piperazine ethanesulfonic acid (HEPES) were purchased from J&K Scientific (Beijing, China), and Ni Sepharose 6FF and Q-Sepharose FF from Beijing Solarbio Science & Technology Co., Ltd. (Beijing, China). MC38 and MC38-B7H1 mouse colon cancer cells were acquired from BMCR (Beijing, China), the instant thin-layer chromatography silica gel (iTLC-SG) chromatography paper was purchased from Agilent Technologies (Palo Alto, CA, USA) and NAP-5 size exclusion columns from GE Healthcare (Uppsala, Sweden), C57BL/6J Mice were purchased from Hunan SJA Laboratory Animal Co (Changsha, China). The GC-2010 gamma radiation counter was purchased from USTC ZONKIA (Hefei, China), the Infinia V Hawkeye 4 SPECT/CT imaging system was purchased from GE Healthcare (Chicago, IL, USA), and ^{99m}Tc was obtained as pertechnetate from an ^{99}Mo/^{99m}Tc Generator (HTA, Beijing, China) eluted with sterile 0.9% sodium chloride.

Preparation of programmed death ligand 1 affibody

Escherichia coli BL21 cells were transformed with plasmids pET26b (+) containing a gene fragment encoding PD-L1 targeting affibody (shorted for PDA) with a histidine–glutamate–histidine–glutamate–histidine–glutamate (HEHEHE)-tag at the amino terminus and a glycine–glycine–glycine–cysteine (GGGC) chelator at the carboxyl terminus, with the amino acid sequence of MAHEHEHEAEAKYAKE RNKAAYEILYLPNLTNAQKWAFIWKLDDDPSQSSELLS EAKKLNDSQAPKGGGSGGGC. Cells were cultivated in an LB medium containing 50 μg/ml kanamycin at 37°C, and protein expression was induced by 1 mmol/L isopropyl-β-D-thiogalactoside (IPTG). After harvesting, cells were disrupted by sonication followed by centrifugation to remove cell debris, and the clarified cell lysate was heat-treated at 60°C for 10 min to precipitate a portion of endogenous *E. coli* proteins. The heat-treated cell lysate was then centrifuged and filtered through a 0.22 μm filter. Affibodies were recovered by immobilized metal affinity chromatography (IMAC, Ni Sepharose 6FF) and further purified by anion exchange (Q-Sepharose FF).

The purified affibodies were identified by sodium dodecyl sulfate–polyacrylamide gel (SDS-PAGE) and matrix-assisted laser desorption ionization–tandem time-of-flight mass spectrometry (Maldi-TOF/TOF).

Radiolabeling and *in vitro* stability of ^{99m}Tc-PDA

The purified PDA was added with DTT at a final concentration of 30 mmol/L, reduced at 37°C for 2 h to destroy the spontaneously formed intermolecular disulfide bonds between cysteines, desalted on a NAP-5 column, and stored at 4°C.

A sterile vial was added with 200 µl of an argon-degassed labeling buffer (10 mM HEPES, 20 mM sodium glucoheptonate, pH 6.6), 100 µl of a reducing affibody (approximately 80 µg), 100 µl of $^{99m}TcO_4$ (typically, 37 MBq), and 0.5 µl of a freshly prepared $SnCl_2$ solution (0.4 mg/ml, in 0.5 mmol/L hydrochloric acid). The reaction solution was incubated at 90°C for 10 min and then cooled at room temperature for 15 min, followed by filtered through a 0.22-micron filter and diluting with PBS.

TLC was performed as described in the reference (14). A small aliquot of samples (~ 0.5 µl) was taken for the TLC analysis of the labeling yield using iTLC-SG with the PBS mobile phase and of reduced hydrolyzed technetium colloid levels using iTLC-SG with a 10:6:3 ratio (pyridine:acetic acid:water) as the mobile phase.

To test the labeling stability *in vitro*, three 100-µl aliquots of ^{99m}Tc-PDA were mixed with 900 µl of either PBS or fetal bovine serum. Each 0.5 µl sample was taken for TLC analysis, respectively, after incubation for 0.5, 1.0, 2.0, 4.0, and 6.0 h at 37°C.

Cell lines

Colon cancer MC38 cells and human PD-L1 gene-transfected MC38 cells (MC38-B7H1) were cultured in Roswell Park Memorial Institute (RPMI) 1640 medium containing 10% fetal bovine serum and 1% penicillin–streptomycin at 37°C in a 5% CO_2 incubator. While they reached confluence, cells were passaged using a trypsin containing 0.25% ethylene diamine tetraacetic acid (Trypsin-EDTA) solution.

Binding specificity and cellular uptake

The binding affinity and specificity of ^{99m}Tc-PDA to human PD-L1 were determined in MC38-B7H1 cells using MC38 cells as a negative control. MC38-B7H1 and MC38 cells were seeded into 24-well plates at a density of 2×10^5 cells/well and cultured overnight at 37°C. At confluency, cells were treated with increasing concentrations (0.02, 0.10, 0.52, 2.62, 13.11, 65.53, 327.63, and 1638.14 nM, three wells for each concentration) of ^{99m}Tc-PDA. After incubation at 4°C for 2 h, the medium was discarded, the cells were washed three times with ice-cold PBS and then lysed with 0.1 M NaOH, and cell lysates were collected. The radioactivity of ^{99m}Tc-PDA that was bound to the cells was measured by a gamma counter. The dissociation constant (K_D) values were calculated by nonlinear fitting (one-site total and nonspecific binding) using GraphPad Prism 6 software.

The specific binding of ^{99m}Tc-PDA to human PD-L1 was also confirmed in a competitive binding experiment using MC38-B7H1 cells with high PD-L1 expression. A set of six dishes containing a cell monolayer (10^6 cells/dish) were incubated with 1.5 nM ^{99m}Tc-PDA for 1 h; cells in three dishes were added a 100-fold excess of an unlabeled affibody 5 min before the addition of ^{99m}Tc-PDA. After incubation, the medium was collected and the cells were washed three times with ice-cold PBS followed by treatment with 0.1 M NaOH to collect cellular bound radioactivity, and the percentage of cell-bound radioactivity was calculated.

Cellular uptake and the internalization of ^{99m}Tc-PDA were studied using PD-L1-expressing colon cancer MC38-B7H1 cells. Briefly, cells (10^6 cells/dish) were incubated with labeled conjugates (1.5 nM) at 37°C. At predetermined time points (1, 2, 4, 8, 12, and 24 h after incubation started), the supernatant from a set of three dishes was collected and the cells were washed twice with ice-cold PBS; the combined fractions represent the unbound radioligand. The cells then were treated with a buffer containing 4 M urea and 0.2 M glycine, pH 2.5 for 5 min on ice, and membrane-bound radio conjugates were collected. The internalized affibodies were collected after the cells were lysed with 1 M NaOH finally. The percentage of membrane-bound and internalized radioactivity was calculated for each time point.

Animal models

All animal experiments were approved by the principles of the Ethics Committee of Affiliated Hospital of Zunyi Medical University (grant number, KLLYA-2021-019). Female C57BL/6J mice, 7 weeks old, were housed in ventilated filter-topped cages with free access to a standard diet and water. Approximately 10^6 of MC38-B7H1 cells (in 100 µl PBS) were implanted subcutaneously in the right armpit. The mice were used for biodistribution studies approximately 2 weeks after injection, when the tumor reached a volume of approximately 1 cm^3.

Dual-tumor xenograft mice were generated by the implantation of 10^6 of both MC38-B7H1 cells and MC38 cells in each armpit, which allows ^{99m}Tc-PDA to assess radioactive uptake in tumors with low and high PD-L1 receptor expression in the same animal. SPECT imaging is performed when the tumor

volume reaches approximately 1 cm^3, The expression of PD-L1 in these tumors was determined by immunohistochemistry with a rabbit anti-human PD-L1 monoclonal antibody (ZR3).

Ex vivo biodistribution studies

MC38-B7H1 tumor-bearing mice were injected with approximately 1.85 MBq of ^{99m}Tc-PDA *via* the lateral tail vein, which corresponds to approximately 4 μg of peptide. Mice were sacrificed immediately (approximately 10 s), 30, 60, 120, 180, and 360 min after injection. The tumor, heart, liver, spleen, lung, kidney, brain, thyroid, muscle, bone (femur), stomach, duodenum, and eyeball blood were collected and weighed, and radioactivity was measured using a gamma counter. Biodistribution measurements were expressed as percent injected dose rate per gram tissue (%ID/g).

SPECT imaging

Dual-tumor xenograft mice were anesthetized with isoflurane (4%–5% induction, 1%–3% maintenance), and approximately 3.7 MBq of ^{99m}Tc-PDA (approximately 8 μg PDA) was injected *via* the lateral tail vein. SPECT imaging was performed at 30, 60, and 120 min after injection, and 200 K counts were collected. Mice in the blocking group were injected with 400 μg of unlabeled affibodies 5 min before the tail vein injection of ^{99m}Tc-PDA, and SPECT imaging was performed 60 min later.

Statistical analysis

Statistical analysis was performed using Statistical Product and Service Solutions (SPSS) software version 25.0, and the mean ± standard deviation ($\bar{x} \pm s$) was used to represent the measurement data conforming to the normal distribution. The comparison between two groups was performed by an independent sample *t*-test, and the comparison among multiple groups was performed by one-way analysis of variance (one-way ANOVA). If $P<0.05$, the pairwise mean difference test (i.e., least significant difference method) was further used for pairwise comparison when the variances were equal, Tamhane's T2 test or Dunnett's T3 test was used when variances were unequal, and $P<0.05$ was considered statistically significant for differences.

Results

Production, purification, and characterization of PDA

PDA was expressed in *E. coli* and recovered by IMAC. The imidazole gradient was further divided, and it was found that elution with 60 mM imidazole yields a purity more than 90% (as shown in Figure 1A), which was further purified by anion exchange and was eluted with 200 mM NaCl to a purity of >95% (as shown in Figure 1B). The purified affibodies were confirmed by Maldi-TOF/TOF (as shown in Figure 2), showing that the amino acid coverage was 95%, and the molecular mass was 8,227 Da.

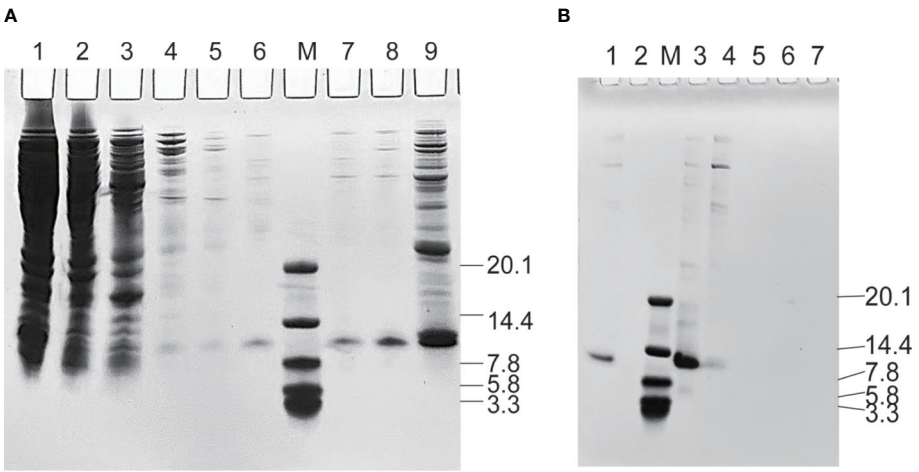


FIGURE 1
Sodium dodecyl sulfate–polyacrylamide gel analysis during the purification of PDA. **(A)** Purification of affibodies by immobilized metal affinity chromatography. Lane M, the protein molecular marker (kDa), Lanes 1~9 are cell lysate after heat treatment, unbound protein, 5, 10, 20, 30, 40, 60, and 300 mM imidazole-eluted product, respectively. **(B)** Purification of affibodies by anion exchange. Lane M, the protein molecular marker (kDa); Lane 1~7 sequentially are sample eluted with 60 mM imidazole and 50, 200, 500, 800, 1,200, and 4,000 mM NaCl-eluted products.

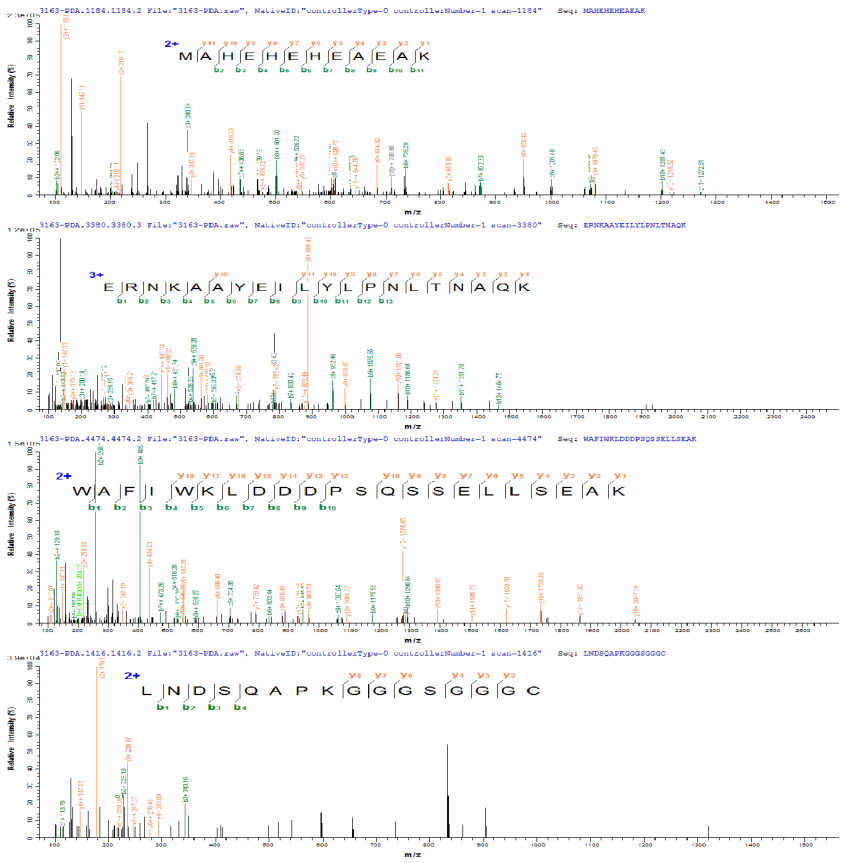


FIGURE 2
Amino acid sequence of PDA determined by matrix-assisted laser desorption ionization–tandem time-of-flight mass spectrometry (Maldi-TOF/TOF).

Radiolabeling and *in vitro* stability of ^{99m}Tc-PDA

^{99m}Tc-PDA was obtained with a high labeling yield (95.95% ± 1.26%); reduced hydrolyzed technetium colloid (3.21% ± 0.37%) (n=10) could be used for biological experiments without additional purification. Bacterial endotoxin assay results show less than 1 EU/ml. TLC analysis showed that ^{99m}Tc-PDA had good stability, and the radiochemical purity was decreased slightly after 4~6 h of incubation in PBS and serum at 37°C (p<0.05) and still greater than 90% (as shown in Figure 3).

Binding specificity and cellular uptake

Radioactivity was significantly higher in MC38-B7H1 than in MC38 cells at all concentration points (P<0.01), and competitive binding assays showed that excess unlabeled affibodies obviously reduced the binding of ^{99m}Tc-PDA to MC38-B7H1 cells,which suggests that the binding of ^{99m}Tc-PDA to living PD-L1-expressing cells was receptor mediated (as shown in Figure 4A). ^{99m}Tc-PDA had a high affinity to MC38-B7H1 cells with a K_D value of approximately 10.02 nM (as shown in Figure 4B).

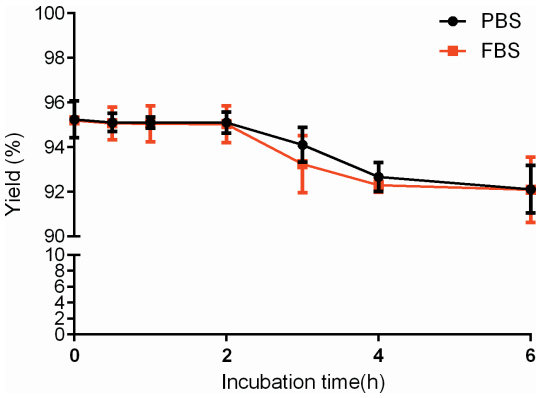


FIGURE 3
Stability of ^{99m}Tc-labeled PD-L1 affibody molecular probe (^{99m}Tc-PDA) in phosphate-buffered saline and serum.

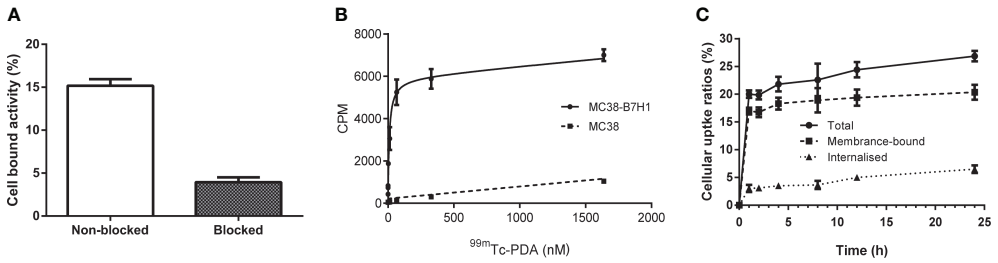


FIGURE 4
Binding specificity and cellular uptake. **(A)** Binding specificity of ^{99m}Tc-PDA to PD-L1 in competitive binding assays. **(B)** Affinity analysis of ^{99m}Tc-PDA to PD-L1-expressing MC38-B7H1 cells. **(C)** Uptake and internalization of ^{99m}Tc-PDA at 37°C by MC38-B7H1 cells.

Binding of ^{99m}Tc-PDA to MC38-B7H1 cells increased rapidly during the first 1 h of culture; only a minor increase was observed after this time. The internalization of ^{99m}Tc-PDA by MC38-B7H1 cells increased with time; about 24.25% ± 2.99% of the total cell-associated radioactivity was internalized after 24 h of incubation (as shown in Figure 4C).

Identification of MC38/MC38-B7H1 xenograft model

Immunohistochemistry (IHC) confirmed the strong positive expression of PD-L1 in MC38-B7H1 xenograft tumors and the negative expression of PD-L1 in MC38 tumors (as shown in Figure 5).

Biodistribution in MC38-B7H1 xenograft-bearing mice

The data concerning the biodistribution of ^{99m}Tc-PDA in MC38-B7H1 xenograft mice are presented in Table 1. The molecular probe ^{99m}Tc-PDA showed rapid clearance from the blood (14.13 ± 1.59%ID/g at 10 s after injection, whereas 0.50 ± 0.12%ID/g at 60 min after injection). The heart, spleen, and lung also showed the highest ^{99m}Tc-PDA uptake immediately after injection and then gradually decreased along with time. Renal retention was obvious; the uptake was the highest at 30 min after injection (%ID/g=87.53 ± 15.09) and gradually decreased with time and decreased to at 360 min (%ID/g=5.63 ± 1.61).

Tumors exhibited rapid uptake, with a significant increase in tracer tumor uptake after 30 min followed by a slow increase and a gradual decrease after peaking at 120 min. The %ID/g ratio of tumors compared with blood, the liver, and muscle reached a peak at 120 min after injection, and the ratios were (32.40 ± 10.18), (7.79 ± 1.54), and (42.72 ± 12.44), respectively (as shown in Figure 6).

SPECT imaging of dual-flank MC38-B7H1/MC38 xenograft-bearing mice

The SPECT imaging of tumor-bearing mice used an imaging system for the human body; the spatial resolution of the images

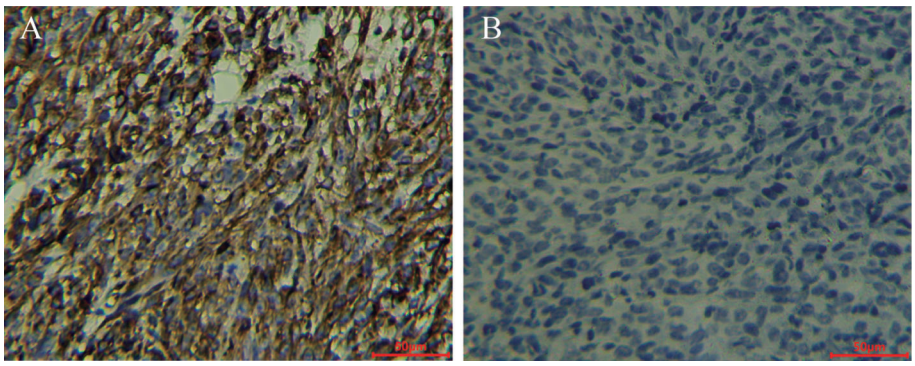


FIGURE 5
ZR3 immunohistochemical staining of MC38-B7H1 and MC38 xenograft tumors Scale bar: 50 μm; **(A)** Membranous staining was observed in MC38-B7H1 samples. **(B)** There was no signal in MC38 tumors.

TABLE 1 Biodistribution of ^{99m}Tc-labeled PD-L1 affibody molecular probe in MC38-B7H1 xenograft-bearing mice (%ID/g, $\bar{x} \pm s$).

Organ or tissue	10 s (n = 3)	30 min (n = 7)	60 min (n = 6)	120 min (n = 7)	180 min (n= 8)	360 min (n = 3)
Blood	14.13 ± 1.59	2.11 ± 0.58	1.04 ± 0.25	0.50 ± 0.12	0.33 ± 0.08	0.11 ± 0.003
Heart	3.01 ± 0.13	1.98 ± 0.36	0.85 ± 0.12	0.57 ± 0.05	0.30 ± 0.07	0.09 ± 0.01
Liver	2.71 ± 0.46	3.50 ± 0.53	2.51 ± 0.24	2.00 ± 0.31	2.03 ± 0.32	2.01 ± 0.04
Spleen	2.32 ± 0.44	1.70 ± 0.35	0.81 ± 0.14	0.77 ± 0.17	0.71 ± 0.11	0.58 ± 0.02
Lung	5.55 ± 1.47	4.10 ± 0.68	2.03 ± 0.27	1.66 ± 0.41	0.87 ± 0.26	0.43 ± 0.08
Kidney	8.37 ± 1.54	87.53 ± 15.09	35.27 ± 3.08	17.86 ± 2.13	12.96 ± 1.94	5.6 3± 1.61
Brain	0.55 ± 0.07	0.14 ± 0.03	0.11 ± 0.03	0.11 ± 0.02	0.04 ± 0.01	0.02 ± 0.005
Thyroid	2.45 ± 0.28	3.63 ± 0.56	3.69 ± 0.26	4.50 ± 0.89	2.75 ± 0.74	0.41 ± 0.09
Gastrointestinal	1.59 ± 0.09	4.19 ± 0.94	4.24 ± 1.03	4.82 ± 1.09	2.71 ± 0.73	0.17 ± 0.04
Tumor	1.03 ± 0.15	12.06 ± 1.61	14.97 ± 3.47	15.50 ± 3.68	8.77 ± 2.70	0.98 ± 0.07
Muscle	0.57 ± 0.09	1.34 ± 0.22	0.77 ± 0.15	0.37 ± 0.07	0.29 ± 0.07	0.09 ± 0.02
Bone	1.80 ± 0.28	2.09 ± 0.47	1.62 ± 0.31	0.38 ± 0.07	0.88 ± 0.24	0.18 ± 0.04

was therefore limited. However, the radiosonde accumulation in the MC38-B7H1 tumor was clearly observed 30 min after injection, while the MC38 tumor was never visible. The highest radioactive accumulation was demonstrated in bilateral kidneys and the bladder, indicating that ^{99m}Tc-PDA was mainly excreted through the urinary system, as previously observed in the biodistribution studies. The thyroid and stomach were not visualized, conjecturing that ^{99m}Tc-PDA had good stability *in vivo* within 120 min. A blockade with 400 μg of PDA caused a significant reduction of ^{99m}Tc-PDA uptake in MC38-B7H1 tumors (as shown in Figure 7). We manually demarcated the tumor and lower extremity (representing the radioactivity of the muscle) and obtained the mean of the radioactivity counts per unit volume, and the ratio of radioactivity counts of the tumor to muscle was 17.84 ± 2.80 at 120 min after injection, which is lower than in biodistribution.

Discussion

High PD-L1 expression is associated with a poor prognosis in various malignant tumors (15, 16); blockades targeting the PD-1/PD-L1 pathway have shown controllable safety and durable remissions in lung cancer, melanoma, bladder cancer, and other tumors (17–28). However, 30%–60% of patients do not respond to the PD-1/PD-L1 blockade (2). In addition, considering the immune-related adverse events of the drugs (29, 30) and the high cost of treatment, it is of great significance to screen patients who can benefit from these drugs before treatment.

The efficacy of PD-1/PD-L1 blockade therapy differs significantly between PD-L1-positive and PD-L1-negative patients (31). The expression of PD-L1 is currently one of the most important predictive biomarkers, and the detection of PD-L1 expression in the tumor environment by IHC is the

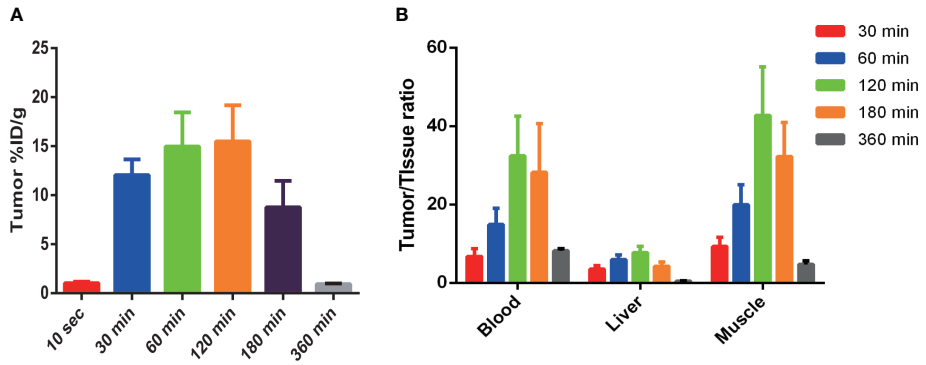


FIGURE 6
^{99m}Tc-PDA uptake in MC38-B7H1 tumors **(A)** and comparison of ^{99m}Tc-PDA in tumor to tissue **(B)**. **(A)** Tumor uptake was significantly higher at all time points after 30 min than 10 s post-^{99m}Tc-PDA injection ($P < 0.01$). **(B)** Significant difference in tumor-to-blood ratios were observed at 30 min vs. 120 min and 60 min vs. 120 min after injection ($P < 0.01$); the tumor-to-liver ratios were significantly higher at 120 min compared to 30 and 180 min ($P < 0.01$), and the tumor-to-muscle ratios were significantly higher at 120 min compared to 30 and 60 min.

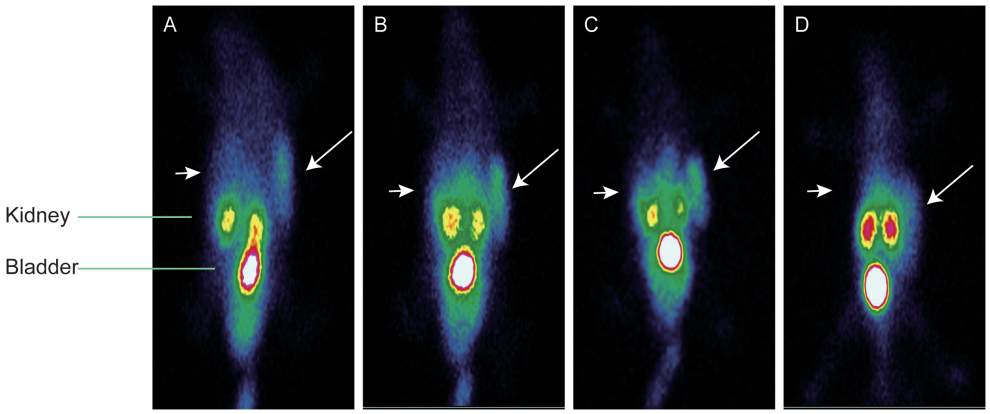


FIGURE 7
Representative single photon emission–computed tomography maximum-intensity projection images of ^{99m}Tc-PDA in dual-flank MC38-B7H1/MC38 xenograft-bearing mie; long arrows indicate MC38-B7H1 tumors, and short arrows indicate MC38 tumors. **(A)** Images at 30 min. **(B)** Images at 60 min. **(C)** Images at 120 min. **(D)** Blocking images at 60 min.

most widely used patient screening method for immunotherapy at present. However, IHC results are adversely affected by biopsy sampling errors and the consistent ratio of PD-L1 expression in primary tumors and metastatic lesions (32, 33), and static biopsies cannot dynamically monitor changes in PD-L1 expression during treatment. Molecular imaging using radiolabeled enables a noninvasive, comprehensive, and dynamic assessment of PD-L1 expression *in vivo*.

Currently, commonly used molecular imaging probes targeting PD-L1 include antibodies, small peptides, or proteins that specifically bind to the receptor. Monoclonal antibodies have high affinity and specificity; however, due to their large molecular weight and weak tissue penetration, which has long circulation retention time, the needed labeling with isotopes with a longer half-life and high-contrast imaging can often only be performed after multiple days and with a risk of false-positive results due to the remaining blood pool activity. Furthermore, the use of long-lived isotopes risks increasing the exposure of subjects to radiation (34–36). Proteins that bind specifically to receptors such as affibodies is an artificial non-immunoglobulin molecule derived from the B domain of staphylococcal A protein with a molecular weight of approximately 6.5 kDa, consisting of 58 amino acid residues and lacking cysteine (37). The introduction of cysteine can provide a binding site for the specific site binding of thiol-reactive radionuclides or chelates. In our present study, the GGGC sequence was introduced into the carboxyl terminus of the PD-L1 affibody, and the thiol group of cysteine together with the amide nitrogen of the adjacent amino acid formed an N_3S chelate structure, thereby realizing the stable labeling of ^{99m}Tc (14, 38, 39). Moreover, a hydrophilic HEHEHE-tag was introduced at the N-terminus of the PD-L1 affibody, on one hand, to facilitate the recovery of protein by IMAC; on the other hand, the hydrophobicity of the HEHEHE-tag reduces the hepatic retention of the tracer (40–42).

^{99m}Tc-PDA exhibits very rapid blood clearance, which is associated with its rapid excretion through the urinary system. The uptake of the molecule probe in the heart and lung was highest immediately after injection, which was related to the abundant blood supply of these organs. Both biodistribution studies and *in vivo* imaging showed low uptake in the spleen, indicating that PDA does not cross-react with murine PD-L1. In addition, increased radioactive uptake in the thyroid and gastrointestinal tract was observed within 30–120 min postinjection; this was presumably associated with a small amount of unconjugated ^{99m}TcO in the labeled compound. ^{99m}TcO can be absorbed by a normal thyroid and gastric mucosa and secreted by the gastric mucosa and then enter the intestinal tract. Sustained decrease was observed over the subsequent time, indicating that ^{99m}Tc-PDA was stable *in vivo*.

^{99m}Tc-PDA has a high affinity for the PD-L1 receptor, and the K_D value reaches the nM level, similar to the previously reported PD-L1 affibody molecular probes (13, 43). Compared with ^{18}F and ^{68}Ga labeled affibodies, ^{99m}Tc-PDA had lower renal retention, which may be explained by differences in radioisotopes and the use of bifunctional chelators.

We had tested the detection ability of ^{99m}Tc-PDA for PD-L1 expression at different times after injection, and the results showed that ^{99m}Tc-PDA could quickly penetrate into the tumor tissue after injection; high-contrast tumor imaging could be obtained within 30 min. Combined with the characteristics of biodistribution, it is speculated that the optimal imaging time is 1~2 h after injection, comparable to the imaging time of peptide- or nanobody-based imaging agents (44–46), which is associated with the rapid blood clearance and tumor penetration of small-molecular-weight ligands. Compared with mAb-based tracers (35, 47), imaging time was significantly shortened, and liver retention was also reduced due

to the different metabolic pathways between affibodies and antibodies.

There were still some shortcomings in this study such as unsatisfactory image quality due to SPECT imaging using an imaging system for the human body and a higher uptake of the tracer in the kidneys, which would limit the injected activity, resulting in reduced sensitivity to low-expressing lesions. The alteration of the affibody structure is needed to reduce renal uptake in future research. Furthermore, the uptake of tracers in the thyroid and gastrointestinal tract was relatively high although significantly lower than in tumors, optimizing the formulation of the labeling system, so improving the labeling rate is an effective method to solve such problems.

In a nutshell, our developed molecular probe ^{99m}Tc-PDA showed rapid blood clearance and good targeting *in vivo* and is expected to be a candidate drug for the SPECT/CT imaging of PD-L1 expression in cancer patients. Further clinical studies are needed to clarify its metabolic characteristics and imaging potential in humans.

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 animal study was reviewed and approved by Institutional Animal Ethics and Use Committee of Zunyi Medical University, Zunyi Medical University.

Author contributions

JC and ZL designed the experiments; ZL and XH performed the experiments and wrote the original manuscript; HH analyzed the data; JC and PW revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This study was funded by ZunShiKeHe HZ2014214, Zunyi Medical College Research Start Fund 2018ZYFY03.

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.

References

1. Hamanishi J, Mandai M, Matsumura N, Abiko K, Baba T, Konishi I. PD-1/PD-L1 blockade in cancer treatment: perspectives and issues. *Int J Clin Oncol* (2016) 21(3):462–73. doi: 10.1007/s10147-016-0959-z

2. Song M, Chen X, Wang L, Zhang Y. Future of anti-PD-1/PD-L1 applications: Combinations with other therapeutic regimens. *Chin J Cancer Res* (2018) 30 (2):157–72. doi: 10.21147/j.issn.1000-9604.2018.02.01

3. Makuku R, Khalili N, Razi S, Keshavarz-Fathi M, Rezaei N. Current and future perspectives of PD-1/PDL-1 blockade in cancer immunotherapy. *J Immunol Res* (2021) 2021:6661406. doi: 10.1155/2021/6661406

4. Liberini V, Laudicella R, Capozza M, Huellner MW, Burger IA, Baldari S, et al. The future of cancer diagnosis, treatment and surveillance: A systemic review on immunotherapy and immuno-PET radiotracers. *Molecules* (2021) 26:2201. doi: 10.3390/molecules26082201

5. Broos K, Lecocq Q, Raes G, Devoogdt N, Keyaerts M, Breckpot K. Noninvasive imaging of the PD-1:PD-L1 immune checkpoint: Embracing nuclear medicine for the benefit of personalized immunotherapy. *Theranostics* (2018) 8(13):3559–70. doi: 10.7150/thno.24762

6. Sandström M, Lindskog K, Velikyan I, Wennborg A, Feldwisch J, Sandberg D, et al. Biodistribution and radiation dosimetry of the anti-HER2 affibody molecule 68Ga-ABY-025 in breast cancer patients. *J Nucl Med* (2016) 57(6):867–71. doi: 10.2967/jnumed.115.169342

7. Sörensen J, Velikyan I, Sandberg D, Wennborg A, Feldwisch J, Tolmachev V, et al. Measuring HER2-receptor expression in metastatic breast cancer using [68Ga]ABY-025 affibody PET/CT. *Theranostics* (2016) 6(2):262–71. doi: 10.7150/thno.13502

8. Burley TA, Da Pieve C, Martins CD, Ciobota DM, Allott L, Oyen W, et al. Affibody-based PET imaging to guide EGFR-targeted cancer therapy in head and neck squamous cell cancer models. *J Nucl Med* (2019) 60(3):353–61. doi: 10.2967/jnumed.118.216069

9. Jussing E, Lu L, Grafström J, Tegnebratt T, Arnberg F, Rosik HW, et al. [(68)Ga]ABY-028: an albumin-binding domain (ABD) protein-based imaging tracer for positron emission tomography (PET) studies of altered vascular permeability and predictions of albumin-drug conjugate transport. *EJNMMI Res* (2020) 10(1):106. doi: 10.1038/s41598-021-97694-6

10. Persson J, Puuvuori E, Zhang B, Velikyan I, Åberg O, Müller M, et al. Discovery, optimization and biodistribution of an affibody molecule for imaging of CD69. *Sci Rep* (2021) 11(1):19151. doi: 10.1038/s41598-021-97694-6

11. Rinne SS, Leitao CD, Abouzayed A, Vorobyeva A, Tolmachev V, Ståhl S, et al. HER3 PET imaging: (68)Ga-labeled affibody molecules provide superior HER3 contrast to (89)Zr-labeled antibody and antibody-Fragment-Based tracers. *Cancers (Basel)* (2021) 13(19):4791. doi: 10.3390/cancers13194791

12. Hu X, Li D, Fu Y, Zheng J, Feng Z, Cai J, et al. Advances in the application of radionuclide-labeled HER2 affibody for the diagnosis and treatment of ovarian cancer. *Front Oncol* (2022) 12:917439. doi: 10.3389/fonc.2022.917439

13. González Trotter DE, Meng X, McQuade P, Rubins D, Klimas M, Zeng Z, et al. *In vivo* imaging of the programmed death ligand 1 by (18)F PET. *J Nucl Med* (2017) 58(11):1852–7. doi: 10.2967/jnumed.117.191718

14. Oroujeni M, Andersson KG, Steinhardt X, Altai M, Orlova A, Mitran B, et al. Influence of composition of cysteine-containing peptide-based chelators on

biodistribution of 99mTc-labeled anti-EGFR affibody molecules. *Amino Acids* (2018) 50(8):981–94. doi: 10.1007/s00726-018-2571-1

15. Dave K, Ali A, Magalhaes M. Increased expression of PD-1 and PD-L1 in oral lesions progressing to oral squamous cell carcinoma: a pilot study. *Sci Rep* (2020) 10(1):9705. doi: 10.1038/s41598-020-66257-6

16. Wang S, Yuan B, Wang Y, Li M, Liu X, Cao J, et al. Clinicopathological and prognostic significance of PD-L1 expression in colorectal cancer: a meta-analysis. *Int J Colorectal Dis* (2021) 36(1):117–30. doi: 10.1007/s00384-020-03734-4

17. Plimack ER, Bellmunt J, Gupta S, Berger R, Chow LQ, Juco J, et al. Safety and activity of pembrolizumab in patients with locally advanced or metastatic urothelial cancer (KEYNOTE-012): a non-randomised, open-label, phase 1b study. *Lancet Oncol* (2017) 18(2):212–20. doi: 10.1016/S1470-2045(17)30007-4

18. Gandhi L, Rodriguez-Abreu D, Gadgeel S, Esteban E, Felip E, De Angelis F, et al. Pembrolizumab plus chemotherapy in metastatic non-Small-Cell lung cancer. *N Engl J Med* (2018) 378(22):2078–92. doi: 10.1056/NEJMoa1801005

19. Socinski MA, Jotte RM, Cappuzzo F, Orlandi F, Stroyakovskiy D, Nogami N, et al. Atezolizumab for first-line treatment of metastatic nonsquamous NSCLC. *N Engl J Med* (2018) 378(24):2288–301. doi: 10.1056/NEJMoa1716948

20. Armand P, Rodig S, Melnichenko V, Thieblemont C, Bouabdallah K, Tumyan G, et al. Pembrolizumab in relapsed or refractory primary mediastinal Large b-cell lymphoma. *J Clin Oncol* (2019) 37(34):3291–9. doi: 10.1200/JCO.19.01389

21. Kato K, Cho BC, Takahashi M, Okada M, Lin CY, Chin K, et al. Nivolumab versus chemotherapy in patients with advanced oesophageal squamous cell carcinoma refractory or intolerant to previous chemotherapy (ATTRACTION-3): a multicentre, randomised, open-label, phase 3 trial. *Lancet Oncol* (2019) 20 (11):1506–17. doi: 10.1016/S1470-2045(19)30626-6

22. Shi Y, Su H, Song Y, Jiang W, Sun X, Qian W, et al. Safety and activity of sintilimab in patients with relapsed or refractory classical Hodgkin lymphoma (ORIENT-1): a multicentre, single-arm, phase 2 trial. *Lancet Haematol* (2019) 6(1): e12–9. doi: 10.1016/S2352-3026(18)30192-3

23. Cortes J, Cescon DW, Rugo HS, Nowecki Z, Im SA, Yusof MM, et al. Pembrolizumab plus chemotherapy versus placebo plus chemotherapy for previously untreated locally recurrent inoperable or metastatic triple-negative breast cancer (KEYNOTE-355): a randomised, placebo-controlled, double-blind, phase 3 clinical trial. *Lancet* (2020) 396(10265):1817–28. doi: 10.1016/S0140-6736(20)32531-9

24. Finn RS, Ikeda M, Zhu AX, Sung MW, Baron AD, Kudo M, et al. Phase ib study of lenvatinib plus pembrolizumab in patients with unresectable hepatocellular carcinoma. *J Clin Oncol* (2020) 38(26):2960–70. doi: 10.1200/JCO.20.00808

25. Geoerger B, Kang HJ, Yalon-Oren M, Marshall LV, Vezina C, Pappo A, et al. Pembrolizumab in paediatric patients with advanced melanoma or a PD-L1-positive, advanced, relapsed, or refractory solid tumour or lymphoma (KEYNOTE-051): interim analysis of an open-label, single-arm, phase 1-2 trial. *Lancet Oncol* (2020) 21(1):121–33. doi: 10.1016/S1470-2045(19)30671-0

26. Hughes B, Munoz-Couselo E, Mortier L, Bratland Å, Gutzmer R, Roshdy O, et al. Pembrolizumab for locally advanced and recurrent/metastatic cutaneous squamous cell carcinoma (KEYNOTE-629 study): an open-label, nonrandomized, multicenter, phase II trial. *Ann Oncol* (2021) 32(10):1276–85. doi: 10.1016/j.annonc.2021.07.008

27. Sun JM, Shen L, Shah MA, Enzinger P, Adenis A, Doi T, et al. Pembrolizumab plus chemotherapy versus chemotherapy alone for first-line treatment of advanced oesophageal cancer (KEYNOTE-590): a randomised, placebo-controlled, phase 3 study. *Lancet* (2021) 398(10302):759–71. doi: 10.1016/S0140-6736(21)01234-4

28. Zsiros E, Lynam S, Attwood KM, Wang C, Chilakapati S, Gomez EC, et al. Efficacy and safety of pembrolizumab in combination with bevacizumab and oral metronomic cyclophosphamide in the treatment of recurrent ovarian cancer: A phase 2 nonrandomized clinical trial. *JAMA Oncol* (2021) 7(1):78–85. doi: 10.1001/jamaoncol.2020.5945

29. Baxi S, Yang A, Gennarelli RL, Khan N, Wang Z, Boyce L, et al. Immune-related adverse events for anti-PD-1 and anti-PD-L1 drugs: systematic review and meta-analysis. *BMJ* (2018) 360:k793. doi: 10.1136/bmj.k793

30. Wang Y, Zhou S, Yang F, Qi X, Wang X, Guan X, et al. Treatment-related adverse events of PD-1 and PD-L1 inhibitors in clinical trials: A systematic review and meta-analysis. *JAMA Oncol* (2019) 5(7):1008–19. doi: 10.1001/jamaoncol.2019.0393

31. Liu W, Huo G, Chen P. Efficacy of atezolizumab for advanced non-small cell lung cancer based on clinical and molecular features: A meta-analysis. *Front Immunol* (2022) 13:909027. doi: 10.3389/fimmu.2022.909027

32. Patel SP, Kurzrock R. PD-L1 expression as a predictive biomarker in cancer immunotherapy. *Mol Cancer Ther* (2015) 14(4):847–56. doi: 10.1158/1535-7163.MCT-14-0983

33. Doroshow DB, Bhalla S, Beasley MB, Sholl LM, Kerr KM, Gnjatic S, et al. PD-L1 as a biomarker of response to immune-checkpoint inhibitors. *Nat Rev Clin Oncol* (2021) 18(6):345–62. doi: 10.1038/s41571-021-00473-5

34. Jagoda EM, Vasalatiy O, Basuli F, Opina A, Williams MR, Wong K, et al. Immuno-PET imaging of the programmed cell death-1 ligand (PD-L1) using a zirconium-89 labeled therapeutic antibody, avelumab. *Mol Imaging* (2019) 18:1536012119829986. doi: 10.1177/1536012119829986

35. Jung KH, Park JW, Lee JH, Moon SH, Cho YS, Lee KH. (89)Zr-labeled anti-PD-L1 antibody PET monitors gemcitabine therapy-induced modulation of tumor PD-L1 expression. *J Nucl Med* (2021) 62:656–64. doi: 10.2967/jnumed.120.250720

36. Smit J, Borm FJ, Niemeijer AN, Huisman MC, Hoekstra OS, Boellaard R, et al. PD-L1 PET/CT imaging with radiolabeled durvalumab in patients with advanced-stage non-small cell lung cancer. *J Nucl Med* (2022) 63(5):686–93. doi: 10.2967/jnumed.121.262473

37. Braisted AC, Wells JA. Minimizing a binding domain from protein a. *Proc Natl Acad Sci USA* (1996) 93:5688–92. doi: 10.1073/pnas.93.12.5688

38. Mitran B, Altai M, Hofström C, Honarvar H, Sandström M, Orlova A, et al. Evaluation of 99mTc-z IGF1R:4551-GGGC affibody molecule, a new probe for imaging of insulin-like growth factor type 1 receptor expression. *Amino Acids* (2015) 47(2):303–15. doi: 10.1007/s00726-014-1859-z

39. Yang Y, Zhao X, Xing Y, Yu T, Zhang J, Wang J. Preclinical evaluation of (99m)Tc direct labeling Z(HER2:V2) for HER2 positive tumors imaging. *Oncol Lett* (2018) 16(4):5361–6. doi: 10.3892/ol.2018.9279

40. Lindberg H, Hofström C, Altai M, Honorvar H, Wållberg H, Orlova A, et al. Evaluation of a HER2-targeting affibody molecule combining an n-terminal HEHEHE-tag with a GGGC chelator for 99mTc-labelling at the c terminus. *Tumour Biol* (2012) 33(3):641–51. doi: 10.1007/s13277-011-0305-z

41. Orlova A, Hofström C, Strand J, Varasteh Z, Sandstrom M, Andersson K, et al. [99mTc(CO)3]+-(HE)3-ZIGF1R:4551, a new affibody conjugate for visualization of insulin-like growth factor-1 receptor expression in malignant tumours. *Eur J Nucl Med Mol Imaging* (2013) 40(3):439–49. doi: 10.1007/s00259-012-2284-8

42. Dahlsson Leitao C, Rinne SS, Mitran B, Vorobyeva A, Andersson KG, Tolmachev V, et al. Molecular design of HER3-targeting affibody molecules: Influence of chelator and presence of HEHEHE-tag on biodistribution of 68Ga-labeled tracers. *Int J Mol Sci* (2019) 20(5):1080. doi: 10.3390/ijms20051080

43. Rubins DJ, Meng X, McQuade P, Klimas M, Getty K, Lin SA, et al. *In vivo* evaluation and dosimetry estimate for a high affinity affibody PET tracer targeting PD-L1. *Mol Imaging Biol* (2021) 23(2):241–9. doi: 10.1007/s11307-020-01544-2

44. Xing Y, Chand G, Liu C, Cook G, O'Doherty J, Zhao L, et al. Early phase I study of a (99m)Tc-labeled anti-programmed death ligand-1 (PD-L1) single-domain antibody in SPECT/CT assessment of PD-L1 expression in non-small cell lung cancer. *J Nucl Med* (2019) 60:1213–20. doi: 10.2967/jnumed.118.224170

45. Bridoux J, Broos K, Lecocq Q, Debie P, Martin C, Ballet S, et al. Anti-human PD-L1 nanobody for immuno-PET imaging: Validation of a conjugation strategy for clinical translation. *Biomolecules* (2020) 10:1388. doi: 10.3390/biom10101388

46. Robu S, Richter A, Gosmann D, Seidl C, Leung D, Hayes W, et al. Synthesis and preclinical evaluation of a (68)Ga-labeled adnectin, (68)Ga-BMS-986192, as a PET agent for imaging PD-L1 expression. *J Nucl Med* (2021) 62:1228–34. doi: 10.2967/jnumed.120.258384

47. Bensch F, van der Veen EL, Lub-de Hooge MN, Jorritsma-Smit A, Boellaard R, Kok IC, et al. (89)Zr-atezolizumab imaging as a non-invasive approach to assess clinical response to PD-L1 blockade in cancer. *Nat Med* (2018) 24:1852–8. doi: 10.1038/s41591-018-0255-8

frontiers | Frontiers in Oncology

TYPE Systematic Review
PUBLISHED 24 November 2022
DOI 10.3389/fonc.2022.1037896

OPEN ACCESS

EDITED BY
Bahram Mohajer,
Johns Hopkins Medicine, United States

REVIEWED BY
Chunhao Wang,
Duke University Medical Center, United States
Santiago Cepeda,
Hospital Universitario Río Hortega, Spain

*CORRESPONDENCE
Mueez Waqar
Mueez.waqar@manchester.ac.uk

†These authors share senior authorship

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

RECEIVED 06 September 2022
ACCEPTED 31 October 2022
PUBLISHED 24 November 2022

CITATION
Waqar M, Van Houdt PJ, Hessen E, Li K-L, Zhu X, Jackson A, Iqbal M, O'Connor J, Djoukhadar I, van der Heide UA, Coope DJ and Borst GR (2022) Visualising spatial heterogeneity in glioblastoma using imaging habitats. *Front. Oncol.* 12:1037896. doi: 10.3389/fonc.2022.1037896

Visualising spatial heterogeneity in glioblastoma using imaging habitats

Mueez Waqar[1,2*], Petra J. Van Houdt[3], Eline Hessen[3], Ka-Loh Li[2], Xiaoping Zhu[2], Alan Jackson[2,4], Mudassar Iqbal[5], James O'Connor[2,6], Ibrahim Djoukhadar[4], Uulke A. van der Heide[3], David J. Coope[1,2†] and Gerben R. Borst[2,7†]

[1]Department of Neurosurgery, Geoffrey Jefferson Brain Research Centre, Manchester Centre for Clinical Neurosciences, Northern Care Alliance NHS Foundation Trust, Manchester Academic Health Sciences Centre, Manchester, United Kingdom, [2]Division of Cancer Sciences, School of Medical Sciences, Faculty of Biology, Medicine and Health and Manchester Cancer Research Centre, University of Manchester, Manchester, United Kingdom, [3]Department of Radiation Oncology, the Netherlands Cancer Institute, Amsterdam, Netherlands, [4]Department of Neuroradiology, Geoffrey Jefferson Brain Research Centre, Manchester Centre for Clinical Neurosciences, Northern Care Alliance NHS Foundation Trust, Manchester Academic Health Sciences Centre, Manchester, United Kingdom, [5]Division of Informatics, Imaging and Data Sciences, Faculty of Biology, Medicine and Health and Manchester Cancer Research Centre, University of Manchester, Manchester, United Kingdom, [6]Department of Radiology, The Christie NHS Foundation Trust, Manchester, United Kingdom, [7]Department of Clinical Oncology, The Christie NHS Foundation Trust, Manchester, United Kingdom

Glioblastoma is a high-grade aggressive neoplasm characterised by significant intra-tumoral spatial heterogeneity. Personalising therapy for this tumour requires non-invasive tools to visualise its heterogeneity to monitor treatment response on a regional level. To date, efforts to characterise glioblastoma's imaging features and heterogeneity have focussed on individual imaging biomarkers, or high-throughput radiomic approaches that consider a vast number of imaging variables across the tumour as a whole. Habitat imaging is a novel approach to cancer imaging that identifies tumour regions or 'habitats' based on shared imaging characteristics, usually defined using multiple imaging biomarkers. Habitat imaging reflects the evolution of imaging biomarkers and offers spatially preserved assessment of tumour physiological processes such perfusion and cellularity. This allows for regional assessment of treatment response to facilitate personalised therapy. In this review, we explore different methodologies to derive imaging habitats in glioblastoma, strategies to overcome its technical challenges, contrast experiences to other cancers, and describe potential clinical applications.

KEYWORDS

glioblastoma, imaging, biomarker, habitats, MRI, preoperative, heterogeneity

Introduction

Glioblastoma is the most common form of primary brain cancer with a median survival of just 15 months (1). The treatment outcome of this tumour has not changed in decades and there are increasing efforts to personalize care for glioblastoma patients. This includes novel strategies that deliver intensified upfront treatment around the time of diagnosis, such as preoperatively, to prevent the phenomenon of rapid early progression, a strongly negative prognostic factor (2, 3). These approaches could improve the treatment outcome and require robust non-invasive tools to monitor treatment response. For glioblastoma, this should be on a regional basis given its significant spatial heterogeneity (4).

Intra-tumoral spatial heterogeneity is a well-recognised phenomena in glioblastoma, especially at the genomic and transcriptomic levels (5, 6). Studies utilising multiple regional sampling have described spatially distinct expression of key driver mutations including Epithelial Growth Factor Receptor (EGFR), TP53 and neurofibromatosis type 1 (NF1), and also the presence of at least two transcriptomic Verhaak classes within the same tumour in up to 60% of cases (5). At the microscopic level, spatial heterogeneity can also be appreciated by the presence of distinct tumour niches, which are groups of cells localising to particular regions within the tumour microenvironment. The perivascular niche for example, includes endothelial cells in close proximity to glioblastoma cancer stem cells (7). Tumour niches are characterised by distinct gene expression patterns that could influence response to treatment (8). On a macroscopic level, there is currently no robust method to detect glioblastoma's spatial heterogeneity, which could otherwise aid patient stratification for early time-point clinical trials for example.

Magnetic resonance imaging (MRI) is used to guide glioblastoma treatment including surgery and radiotherapy, and could be used to monitor treatment response on a regional basis. To date, most efforts utilising MRI data in glioblastoma have focussed on radiomic approaches to extract innumerable quantitative imaging metrics with less emphasis on spatially relating these to the tumour microenvironment. Habitat imaging is an emerging imaging technique to delineate the tumour into distinct spatial regions with shared imaging characteristics. These regions can be visualised and interrogated longitudinally to characterise tumour regions and monitor their treatment response (Figure 1).

Traditional approaches to delineating imaging habitats in glioblastoma have considered regions based on their location on structural MRI sequences (optimised for visualising brain anatomy). For example, at least five habitats could be defined by considering just two structural imaging sequences – T1 with gadolinium and Fluid Attenuated Inversion Recovery (FLAIR): the necrotic core, peri-necrotic enhancing rim, enhancing core, enhancing rim and the 'infiltrative zone' defined by FLAIR hyperintensity in the absence of contrast enhancement (9, 10). However, these regions are not always easy to segment and their size is defined by arbitrary and subjective thresholds. The regions themselves are also inherently heterogeneous – for example, the non-enhancing FLAIR hyperintensity is a mixture of oedema and infiltrative tumour with no clear delineation between them. There is therefore a need for alternative methods of deriving imaging habitats in glioblastoma.

In this review, we will provide an overview of the current status of habitat imaging in glioblastoma, highlighting its potential use as a non-invasive tool for more personalised treatment. We will explore different methodologies to derive imaging habitats, strategies to overcome its technical challenges, contrast experiences to other cancers, and describe potential clinical applications.

Imaging biomarkers

A biomarker is defined as a characteristic that is measured as an indicator of normal biological processes, pathogenic processes or responses to an exposure or intervention, including therapeutic interventions (11). Imaging biomarkers are biomarkers that are derived from clinical imaging sequences such as MRI. Examples of conventional imaging biomarkers used in glioblastoma derived from diffusion and perfusion MRI are listed in Table 1.

In addition to conventional imaging biomarkers, it is possible to apply data-mining approaches to imaging data to yield quantifiable data, under the theme of radiomics. Radiomics typically produces a vast set of imaging features that are derived from the tumour as a whole. This feature set is a distinct imaging biomarker in its own right that is useful for aiding in diagnosis, prognostication and predicting treatment response (13). Although this may have advantages to histopathological analysis by decreasing the likelihood of intraoperative under-sampling by considering the tumour as a whole (14), it does not typically relate imaging metrics to individual tumour regions. Tumour subregion radiomic analyses have also focused on relating radiomic features to patient related outcomes, relaying little about the underlying tumour microenvironment limiting its use in guiding novel treatment strategies (15). This limitation of assessing regional response may be overcome by enhanced use of conventional imaging biomarkers used in isolation/together.

Imaging biomarkers provide information about tumour biological characteristics with varying specificity. In current practice, imaging biomarkers are largely used in isolation, which is advantageous given the simplicity of this approach. However, there may be benefit in combining different biomarkers using the additional and differential information provided by considering their overlapping areas. In one study for example, the positive predictive value (PPV) of relative cerebral blood volume (rCBV; defined in Table 1), apparent diffusion coefficient (ADC; defined in Table 1) and the FLAIR signal, to

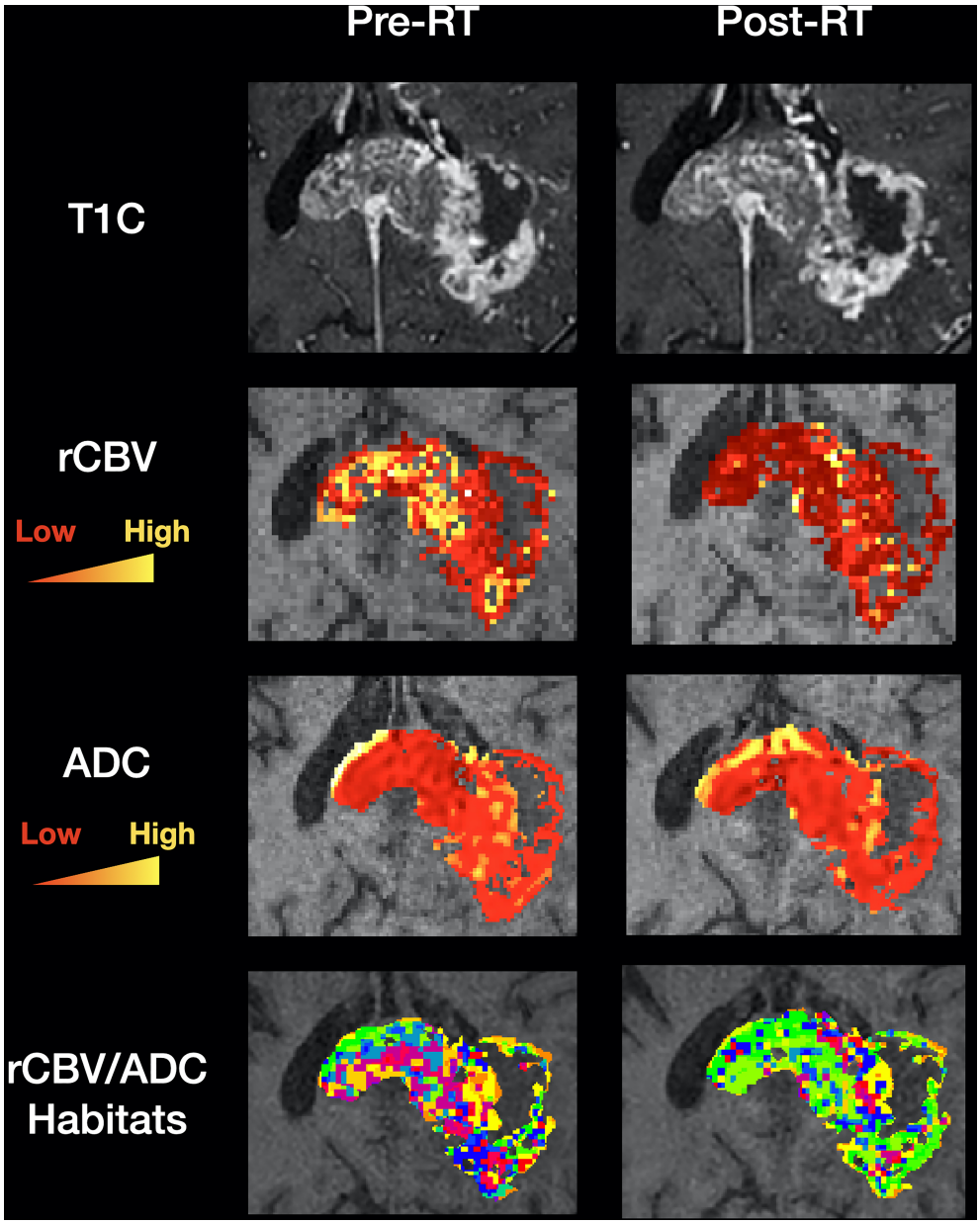


FIGURE 1
Clinical utility of habitat imaging in glioblastoma: assessment of changes pre and post-radiotherapy. This figure demonstrates the clinical utility of habitat imaging in glioblastoma pre and post-radiotherapy. Top row – structural imaging (T1 with contrast) demonstrates no significant changes in tumour anatomy. Middle two rows – diffusion and perfusion MRI scans demonstrate changes in tumour physiology with treatment with a decrease in rCBV for example (red to yellow represents low to high values for each biomarker). Bottom row – imaging habitats map where each voxel is labelled according to both rCBV and ADC values. This method produced 16 different habitats for this patient. After radiotherapy, the biggest increase was in a habitat defined by low rCBV and low ADC (10.5% increase). The biggest decrease was in a habitat defined by high rCBV and medium ADC (5.7% decrease). Habitats that are more resistant to treatment can be spatially visualised and offered targeted therapy. RT, radiotherapy; T1C, T1 with contrast; rCBV, relative cerebral blood volume normalised to contralateral white matter; ADC, apparent diffusion coefficient.

predict disease recurrence in glioblastoma was evaluated. The PPV for recurrence was improved by considering the overlap of high FLAIR, rCBV and low ADC (PPV = 31.9%), versus individual biomarkers alone (PPV for rCBV = 21.6%) (16).

Habitat imaging utilises imaging biomarkers to delineate distinct spatial regions with homogenous biological and physical characteristics within an individual tumour (17). This has specific applications in glioblastoma as identification of more aggressive/

TABLE 1 Conventional imaging biomarkers in glioblastoma (12).

Imaging biomarker	Details
Apparent diffusion coefficient (ADC)	Measurement of inferred ('apparent' rather than actual) water diffusion with DWI. It is a measure of the relative decrease in the transverse magnetization induced by additional dephasing and rephasing magnetic field gradients. Net dephasing and therefore signal loss is greater in freely diffusive tissue. Quantitatively, the ADC is the slop of a line plotting the natural logarithm of the MRI signal (y-axis) per unit of applied magnetic field strength (b-value plotted on x-axis; units mm^2/s).
Mean diffusivity (MD)	This is the magnitude of mean diffusion in a given voxel obtained with diffusion tensor imaging (DTI). ADC may not be uniform at all orientations. MD is the average diffusivity from the three eigenvalues of the diffusion tensor. It is often regarded as an approximation of the overall ADC (units mm^2/s).
Fractional anisotropy (FA)	DTI provides FA values which indicate the overall directionality of water diffusion within a voxel. FA is a scalar value between 0-1 that describes the degree of anisotropy of the diffusion process. A value of zero means that diffusion is isotropic (i.e. equal in all directions, and the diffusion ellipsoid is a sphere). A value of one means that diffusion is totally anisotropic (i.e. diffusion occurs only along one axis and is fully restricted along all other directions).
Cerebral blood volume (CBV)	CBV is the volume of blood in a given amount of brain tissue, most commonly millilitres of blood per 100 g of brain tissue. CBV can be calculated by assessing the area under the concentration-time curve, which in turn can be generated from signal intensity-time curves generated using Dynamic Contrast Enhanced (DCE) MRI (measuring T1 signal recovery) or Dynamic Susceptibility Contrast (DSC) MRI (measuring T2 signal loss), respectively (units ml/100g).
Cerebral blood flow (CBF)	Cerebral blood flow is the volume of blood passing through a given amount of brain tissue per unit of time, most commonly millilitres of blood per minute per 100 g of brain tissue. Alternatively, one may express CBF in terms of flow per unit volume of brain tissue, thus in ml blood/min/100 ml tissue.
Mean transit time (MTT)	Mean transit time is the average period of time that blood spends within the blood vessels in a particular part of the brain (units seconds).
Volume transfer constant (K^{trans})	K^{trans} is the volume transfer constant for contrast agent between blood plasma and the tissue extravascular extracellular space (EES). K^{trans} is derived from a pharmacokinetic model and represents a mix of flow and permeability. It most commonly serves as a measure of permeability/vascular leak under permeability-limited conditions (units min^{-1}).
Rate constant (k_{ep})	k_{ep} determines the washout rate of contrast agent from the extravascular extracellular space back into the blood plasma ($k_{ep} = K^{trans}/v_e$; units min^{-1}).
Extravascular extracellular space fractional volume (v_e)	v_e is defined as the volume of the extravascular extracellular space (EES) per unit volume of tissue, and thus is a dimensionless number between 0 and 1. The parameter v_e reflects the amount of "room" available within the tissue interstitium for accumulating contrast agent. Note that v_e is different from V_e, which represents the total volume of extravascular extracellular space in ml.
Fractional plasma volume (v_p)	Represents the volume of blood plasma per unit volume of tissue (therefore unitless). It is derived from a pharmacokinetic model.
Native longitudinal relaxation rate ($R1_N$)	R1 is the longitudinal relaxation rate of the protons of tissue water (R1 = 1/T1). $R1_N$ is the baseline tissue R1 in the absence of the contrast agent. The $R1_N$ measurement inversely reflects the free water content of tissue (units s^{-1}).

This table provides an overview of the most commonly cited imaging biomarkers used in glioblastoma patients. Note that the prefix of 'r' before these imaging biomarkers represents comparison to a reference region, that is usually the contralateral normal appearing brain parenchyma, but defined differently from study to study.

treatment resistant habitats could enable locally targeted treatment, such as targeted resection for hypoxic areas for example, that correlate with a shorter survival (18). Habitat imaging could also overcome limitations in the sensitivity of individual imaging biomarkers in assessing and monitoring multiple physiological processes, and provide a more accurate representation of the tumour molecular profile non-invasively (19, 20).

Habitat imaging definition

Cancer exhibits marked spatial heterogeneity at the anatomical, physiological and molecular levels (21). Imaging can be interrogated to visualize this spatial heterogeneity and identify imaging habitats (17). Imaging habitats are tumour regions with distinct imaging characteristics that arise from their unique intrinsic cell populations and/or local environmental conditions. Although individual imaging biomarkers could be used in theory to define habitats, based on analysis of voxel signal intensity distributions for example, it is more conventional to use the term when tumour regions are defined using multiple imaging biomarkers. Thus, for the purposes of this review, imaging habitats will refer to tumour regions defined using multiple imaging biomarkers. As each imaging biomarker assesses a different aspect of tumour biology, a multiple biomarker approach also increases the degree of tissue heterogeneity that can be assessed.

Habitat imaging in glioblastoma: Status and potential

Current experience

Several studies have investigated the potential of habitat imaging for predicting relevant clinical endpoints in glioblastoma (Table 2). Supplementary Figure 1 outlines the search strategy and methodology used for this section. In general, there are two main approaches to habitat imaging (Figure 2). The first ('one step') involves using bioinformatics to cluster multi-dimensional imaging biomarker datasets. In this approach, data from multiple imaging biomarkers is combined into a common data table and clustering methods such as hierarchical clustering are used to identify groups (27, 30). The second approach involves two steps, in which data from each individual biomarker is firstly split into data clusters and multiple combinations of those clusters can be used to define habitats (26).

The one step approach to deriving habitats is akin to clustering across a genomic dataset. You et al. derived habitats using hierarchical clustering in 21 glioblastoma patients and were able to relate these to survival and tumour biology descriptively. They found three main biomarker clusters that were named based on the most clinically relevant biomarker of the group: 'FLAIR' cluster – FLAIR, quantitative T1 and T2 signal, and ADC; MET cluster - methionine positron emission tomography (MET PET; methionine is an amino acid PET tracer that localises to metabolically active tumour), CBV and K^{Trans} (marker of permeability; defined in Table 1); and diffusion-weighted imaging (DWI) cluster – DWI and fractional anisotropy (FA; defined in Table 1). The necrotic core was represented by the FLAIR cluster, surrounded by the core cellular component represented by both the FLAIR and DWI clusters, in turn encircled by a metabolically active rim represented by the methionine cluster. This method was therefore able to capture a degree of glioblastoma's heterogeneity. The clinical utility of their method was limited in detecting tumour recurrence, localised to the methionine cluster in only 5 out of 10 cases (27).

Other studies have also found utility in applying the one-step approach. Juan-Albarracín et al. developed an automated method of habitat generation using gaussian mixture modelling applied to rCBV and relative cerebral blood flow (rCBF; defined in Table 1) to produce four habitats – two in the enhancing core (high/low angiogenic) and two in the oedema (infiltrated and vasogenic). They demonstrated that the median rCBVmax or rCBFmax values in the high and low angiogenic habitats were predictive of survival (35). These findings were subsequently validated in a multi-centre study (30). This group has made their technique to generate habitats into a standardised and adaptable pipeline for other centres (36).

The two step approach to generating habitats is more commonly used in the literature (16, 22–26, 28, 29, 34). The first step of biomarker clustering itself can be done using simple methods, such as by dividing intensity values based on average values/quartiles, or using machine learning methods. Habitats are then visualised as the overlaps of individual biomarker clusters. For example, Lee et al. used this approach in 74 glioblastoma patients from The Cancer Genome Atlas (TCGA), using Gaussian mixture modelling to cluster biomarkers (enhancement on T1 with contrast and FLAIR), which were then used to produce four habitats. Spatial features of these habitats were predictive of survival and had a high sensitivity for predicting glioblastoma transcriptomic subtype – highest for the proneural subtype with an area under curve value of 93% (22).

Only few studies have attempted to relate habitats to specific histological or molecular signatures (Table 2) (25, 26, 28, 29). Dextraze et al. analysed 85 glioblastoma patients from TCGA, and reported that the volume of a habitat localised to necrotic regions was positively correlated with an upregulation in Nuclear factor kappa B signaling, for example (26). Bailo et al. is the only study that attempted to directly sample characterised habitats (34). They studied 17 high grade glioma patients and used the two step approach with three biomarkers: v_p - plasma volume (a DCE-MRI biomarker; defined in Table 1), mean diffusivity (defined in Table 1) and uptake of a PET tracer that localizes to hypoxic regions. They undertook multi-regional tumour sampling and related habitats to histopathological features. Although conclusions were difficult to draw in view of the sample size, they identified habitats correlated with more aggressive histological features such as high cellularity and neovascularization (34).

In summary, current experience with habitat imaging in glioblastoma has mostly focused on the prognostic value of the technique applied to preoperative imaging and, in general, correlating habitat frequencies with global tumour biological features/molecular pathways. Existing studies have not explored technical considerations that are of critical importance to habitat imaging and its validation, including derivation method, biomarker selection, imaging acquisition parameters and tissue sampling. These will be explored in the following sections to provide a framework for future studies.

Technical considerations

There are several technical considerations of relevance to habitat imaging.

One versus two step approach. Habitat generation has been described using two main methods (Figure 2). The one step approach clusters data from multiple imaging biomarkers directly, whereas the two step approach has an intermediate clustering step for each imaging biomarker selected. Both techniques are dependent on accurate image registration, which refers to the process of aligning different MRI scans. Quantitative maps represent imaging biomarker values on each

TABLE 2 Habitat imaging in glioblastoma: the current evidence.

Paper	Patients	Imaging habitats					Survival/progression	Histopathology	Molecular
		Stage of imaging	MRI sequences/ biomarkers	Intensity normalisation	Biomarker clustering	Habitat generation			
Lee 2015 Texas, USA (22)	Glioblastoma N = 74 from The Cancer Genome Atlas	Preoperative	T1+C FLAIR	Yes	Gaussian mixture modelling - 2 clusters	Overlap of each cluster to produce 4 habitats	Top 5 spatial features had good area under curve = 0.76 for predicting survival	Top 5 spatial features had high accuracy for predicting subtypes - pro-neural 0.93, classical 0.88, neural 0.85, mesenchymal 0.70	
Lee 2015 Texas, USA (23)	Glioblastoma N = 65 from The Cancer Genome Atlas	Preoperative	T1+C FLAIR	Yes	As above				Top 5 spatial features had high accuracy for predicting EGFR status (AUC 0.845)
McGarry 2016 Single-centre Wisconsin, USA (24)	Glioblastoma N = 81	Preoperative	T1 T1+C FLAIR ADC	Yes	Automated tissue segmentation – 3 clusters	Overlap of each cluster to produce 83 habitats	Identified 5 habitats associated with shorter overall survival High intensity of both FLAIR and contrast enhancement present in 4 out of 5 habitats associated with overall survival	One habitat could be hyper-cellular	
Zhou 2016 Florida (25)	Glioblastoma N = 32 from The Cancer Genome Atlas N = 22 internal cohort	Preoperative	T1+C T2 FLAIR	Yes	Otsu thresholding – 2 clusters	Interested in two habitats formed by overlap of high/high or low/low clusters	Spatial mapping between habitats were better at predicting survival than the presence of habitats themselves In particular, spatial mapping between high T2/high FLAIR region achieved >80% accuracy		
Khalifa 2016 Toulouse, France (16)	Glioblastoma N = 15 Primary glioblastoma enrolled in a trial, ≤5cm diameter	Postoperative pre-radiotherapy	FLAIR rCBV ADC	NR	Gaussian mixture modelling to produce 2 (ADC) and 3 (rCBV) clusters	Interested in 4 habitats formed by various overlap combinations	Positive predictive value for recurrence was highest at 31.9% in overlap region of high FLAIR, hyper-perfusion and restricted diffusion		
Dextraze 2017 Texas, USA (26)	Glioblastoma N = 85 from The Cancer Genome Atlas	Preoperative	T1 T1+C T2 FLAIR	Yes	K-means – 2 clusters	Overlap of each cluster to produce 16 habitats	All three habitats associated with survival present in high enhancing segment One habitat predictive of survival occurred in high FLAIR/high T1+C intensity One habitat in very periphery, two in enhancing core	One habitat predictive of survival correlated with necrosis quantification	Habitats associated with various pathways including NFkB, DNA damage response/ transduction and STAT1/NK activation

(Continued)

10.3389/fonc.2022.1037896

TABLE 2 Continued

Paper	Patients	Imaging habitats					Survival/progression	Histopathology	Molecular
		Stage of imaging	MRI sequences/ biomarkers	Intensity normalisation	Biomarker clustering	Habitat generation			
You 2018 Michigan, USA (27)	Glioblastoma N = 21	Post-resection pre-radiotherapy	T1 T1+C T2 FLAIR DWI ADC FA CBV kTrans MET PET	Yes	Hierarchical clustering to derive habitats		5/10 patients had recurrence related to MET only area (metabolically active rim)	1. High T2/FLAIR habitat mostly in necrotic core 2. High DWI habitat surrounded FLAIR region – core cellular component 3. High methionine habitat in periphery – metabolically active rim	
Stringfield 2019 Multi-centre (28)	Glioblastoma N = 74 (37 in each cohort of long and short term survivors)	Preoperative	T1 T1+C FLAIR	Yes	Otsu thresholding – 2 (FLAIR) or 3 (T1+C) clusters	Overlap to produce 6 habitats	High FLAIR/high T1c habitat present in significantly higher volume in long term survivors		
Li 2019 Cambridge, UK (29)	Glioblastoma Maximal resection Performance status 0-1 N = 112	Preoperative	ADC rCBV	Yes	Quartiles – 4 clusters each	Interested in two habitats – lowest quartile rCBV/lowest quartile ADC, lowest quartile rCBV/highest quartile ADC	Higher volume of these habitats associated with better PFS Lactate/Creatine ratio in these regions associated with shorter PFS and OS	Minimally invasive phenotype defined on DTI had lower proportion of the low rCBV/low ADC habitat	
Alvarez-Torres 2019 Multi-centre (30)	Glioblastoma (N = 184)	Preoperative	rCBV rCBF	NR	Gaussian mixture modelling to yield 4 habitats per patient		Several habitats predictive of survival		
Park & Kim 2020-2021 Seoul, South Korea (31, 32)	Glioblastoma (IDH wildtype; various N)	Post chemo-radiotherapy	T1+C ADC rCBV +/- EPT	NR	K-means clustering to define various habitats using combinations of ADC, rCBV and EPT		Hypovascular cellular habitat (low rCBV and ADC) and hypovascular low conductivity (low rCBV and EPT) habitats strongly correlated with site of future progression.		
Xu 2021 New-York USA (33)	Glioblastoma N = 263 from BraTS 2020 training dataset	Preoperative	T1+C T2 FLAIR	NR	Used simple linear interactive clustering (SLIC) – method that depends on intensity of pixels and their location		Graph features of habitats improved overall survival cox regression model		

(Continued)

10.3389/fonc.2022.1037896

TABLE 2 Continued

Paper	Patients	Imaging habitats					Survival/progression	Histopathology	Molecular
		Stage of imaging	MRI sequences/ biomarkers	Intensity normalisation	Biomarker clustering	Habitat generation			
Bailo 2022 Milan, Italy (34)	High grade gliomas (including 12 glioblastomas)	Preoperative	Vp MD FAZA PET	NR	Otsu thresholding to identify high/ low regions of each biomarker in enhancing tumour or oedema	Multiple overlaps between clusters – 8 habitats possible		Habitats with high Vp/high FAZA uptake (regardless of MD) correlated with hyperplastic vessels and cellularity with low rate of necrosis. Largest volumetric representation was by 'less aggressive' habitats comprising low Vp/low FAZA uptake. These correlated with low cellularity and no signs of necrosis/angiogenesis.	
Yang 2022 Xi'an, China (9)	Glioblastoma (test cohort of 122, validation cohort of 65 patients)	Preoperative	T1 T1+C T2 FLAIR	Yes	Investigated oedema region only. Determined optimal number of K-means clusters using elbow plot method. Performed K-means clustering and produced 4 habitats.		Defined high risk habitat in oedema region based on radiomic features. This habitat improved performance of cox regression model of overall survival.		High risk habitat not correlated with MGMT methylation status

This table summarises data from 15 studies that have performed habitat imaging in glioblastoma. T1 + C, T1 with contrast; FLAIR, Fluid Attenuated Inversion Recovery; ADC, Apparent Diffusion Coefficient; DWI, Diffusion Weighted Imaging; FA, Fractional Anisotropy; CBV, cerebral blood volume; MET PET, methionine positron emission tomography; FAZA PET, 8F-labeled fluoroazomycinarabinoside PET, localises to hypoxic regions; EPT, Electrical Properties Tomography imaging.

10.3389/fonc.2022.1037896

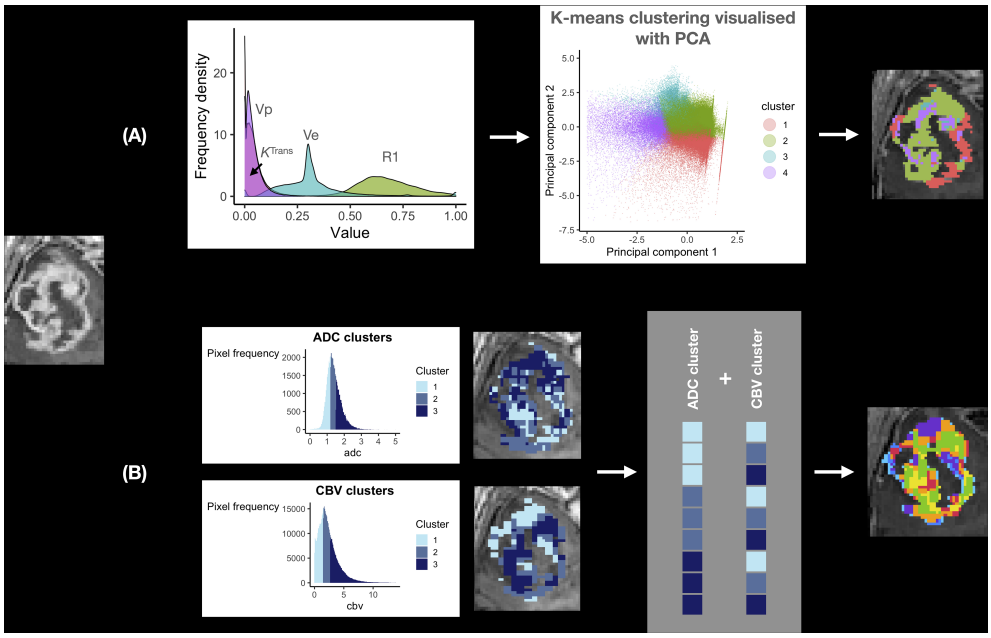


FIGURE 2
Habitat imaging methods in glioblastoma. This figure provides an overview of the two main approaches to deriving imaging habitats utilising local preoperative data from 12 patients with glioblastoma undergoing surgery. **(A)** one step approach: a multi-dimensional dataset can be produced utilising multiple imaging biomarkers from the same MRI acquisition (to avoid interpolation/registration errors), in this case Dynamic Contrast Enhanced (DCE) MRI. Data from $R1_N$ – defined in Table 1 and three DCE-MRI imaging biomarkers (K^{trans}, v_p and v_e) were input into a machine learning K-means clustering algorithm to produce four distinct imaging habitats, that were distinct on Principal Component Analysis (PCA; right). A disadvantage of this approach is its 'black-box' nature, such that it is not straightforward to define each habitat for prospective validation. **(B)** Two step approach: this step first requires clustering of individual imaging biomarkers, in this case ADC and rCBV (left). Each pixel is then assigned to a habitat based on its ADC/rCBV cluster, with multiple cluster combinations defining each habitat (grey box). The advantage of this approach is that imaging biomarkers from different MRI acquisitions (e.g. diffusion and perfusion MRI) can be utilised. It is also easier to define each habitat as the definition of each is derived from its individual ADC/rCBV cluster composition. This approach therefore allows for prospective validation with pre-defined cluster thresholds.

pixel (instead of signal intensity) and are produced from MRI sequences to which they are inherently aligned (e.g. an ADC/rCBV image is aligned to the DWI/dynamic susceptibility contrast MRI from which it was derived, respectively). When biomarker values are extracted from a three-dimensional image to a two-dimensional table of data for clustering, they are done so in pixel-order (i.e. from one edge of the image to the other), which will differ from sequence to sequence due to differences in resolution and therefore the number of potential 'rows of data'. In order to correct for this, image registration is performed to spatially align and transform two images, but this results in distortion of individual values. The newly aligned images will include pixel values that were not present in the original data-set but derived from neighbouring values through interpolation (37). It is not ideal to register the whole quantitative map therefore as it creates artificial data values and can amplify artifacts, but this has been universally done in studies utilising the one step approach to habitat imaging (34). A workaround is to use multiple imaging biomarkers from the same MRI sequence – such as with dynamic contrast enhanced (DCE) MRI (demonstrated in Figure 2 – top panel). DCE-MRI offers a multitude of imaging biomarkers that can assess several aspects of brain tumour physiology including vessel permeability (K^{Trans}), vascularity (v_p - plasma volume), blood flow and cell density (v_e - extravascular extracellular space) (38). These have been validated for use in other brain tumours such as vestibular schwannomas (38), though DCE-MRI is under-utilised in glioblastoma literature (39). Another solution is to use the two step approach, which clusters each biomarker individually prior to habitat generation. Registration is done on clustered data which minimises the effect of extreme/artifactual values (which would otherwise be present in the up-scaled, registered data at a higher frequency). It is also predominantly the edge voxels between different clusters that are affected by spatial transformation steps. A comparison between the one and two step techniques is required in future studies.

Biomarker choice. A significant limitation of most existing studies is their reliance on imaging biomarkers derived from non-quantitative, structural MRI sequences (e.g. T1- and T2-weighted MRI). These sequences were developed for visualisation of gross anatomy and for this purpose, there is a high degree of consistency in brain structural morphology (40).

However, their signal intensity values are affected by hardware factors, such as magnetic field strength inhomogeneity, head placement within the receiver coil, image intensity scaling factors and image acquisition parameters (41). It is difficult to completely negate these effects or correct them using a normalisation step (41). Habitats derived from functional MRI sequences (e.g. diffusion/perfusion) have demonstrated the greatest external validity and this approach should therefore be favoured (30). The functional imaging biomarkers to consider for habitat generation depend on the purpose of the exercise. If this is a clinical aim, such as the identification of treatment-resistant habitats, then robust biomarkers of cellularity and perfusion are important. The imaging biomarkers should also be readily available across centres to allow external validation/adoption. In this case, we hypothesise that ADC, rCBV and K^{Trans} are good candidates to further explore, given their sensitivity to treatment-related change and widespread use (42).

Biomarker calculation. The calculation method is an important consideration for functional imaging biomarkers. For example, the numerical value of DCE-MRI biomarkers such as K^{Trans} can vary in the same dataset depending on the pharmacokinetic model used, due to different underlying physiological assumptions (43). For DCE-MRI analysis in glioblastoma, an extended Toft's model is usually employed that models contrast leakage between intra/extravascular tissue compartments, modified ('extended') for appropriate contribution of the intravascular compartment (44). This can be combined with new processing techniques such as the Legatos method, described and validated by our group, which combines high temporal and high spatial resolution DCE-MRI data, to facilitate habitat imaging (used for Figure 2 top panel) (38, 45). Different model assumptions also also apply to diffusion-derived biomarkers such as ADC, which can be defined using a mono-exponential model (fits a straight line through a graph of signal intensity versus b-values - usually 0 and 1000 s/mm^2; b-values denote the strength of the magnetic field gradient applied in diffusion MRI studies) and more complex exponential models (fits a more complex function involving multiple b-values), with the latter typically producing more accurate results (46). However, to date, studies using ADC for habitat imaging have used monoexponential models (16, 27). The Quantitative Imaging Biomarkers Alliance (QIBA) is an ongoing effort that aims to produce standards for use of specialist imaging such as diffusion/perfusion MRI in clinical and research environments (47). This work could help to standardise biomarker calculation methods, which are currently diverse, poorly understood and not robustly validated (47). As an example, although rCBV is frequently cited in glioblastoma literature, the variation in how it is derived is often not acknowledged. Indeed, it can be derived from dynamic susceptibility contrast enhanced MRI (DSC-MRI) using almost any major imaging analysis software package (including FMRIB Software Library, 3D slicer©, Matlab© and Osirix©) and each uses a different calculation method (48). In general, we recommend utilisation of robust biomarker calculation methods that are amenable to external uptake and therefore validation.

Image acquisition. Habitat generation requires a relatively high spatial resolution. This should be small enough with respect to the size of the tumour to avoid partial volume effects, which occur with larger voxel sizes (i.e. thicker MRI slices) that average MR signal from multiple tissue components included in each voxel (49). However, there is a trade-off between spatial resolution and the signal-to-noise ratio (SNR) that is needed for accurate estimation of imaging biomarker values on a voxel-level, like ADC and K^{trans} (47, 50). The SNR is proportional to voxel volume as larger voxels contain a higher number of protons that subsequently produce a greater MR signal (49). The imaging time must also be considered, as longer durations can result in motion artifacts. Initiatives like QIBA provide guidance for the acquisition of quantitative imaging biomarkers and the use of such MR acquisition parameters would allow a more robust comparison of habitats across centres, especially in the case of multi-centre tissue sampling (47). Scan angulation is another important consideration for the two step approach to remain consistent between functional and anatomical sequences (50). Where data from multiple acquisitions is being utilised, if scan angulations are not aligned, then potentially all data will be resampled and interpolated during image registration. In summary, habitat imaging requires a relatively high spatial resolution (2-3.5mm slice thickness in our experience) that preserves the SNR, and utilises sequences with relatively consistent angulation to structural sequences.

Individual versus group level data. This is of particular relevance to glioblastoma given its significant inter-patient heterogeneity. The techniques described above consider imaging biomarker data on an individual patient level. Biomarker clustering is performed using threshold values defined per patient, rather than the larger group. This is largely because they utilize structural MRI sequences alone, which are not validated for scaled comparisons between patients. However, for quantitative imaging biomarkers, this is an important consideration for glioblastoma given its significant inter-patient heterogeneity. In our previous meta-analysis for example, the mean tumoral blood flow relative to normal appearing white matter across glioblastoma patients in the literature varied from 1.6-7.9 (39). To demonstrate the importance of group-level data, Figure 3 demonstrates differences in ADC thresholds when clustering is performed at the individual patient versus group level. The advantage of group-level definitions is that they allow for reproducibility across both retrospective and prospective datasets. Group-level definitions should therefore be utilised in future studies.

Machine learning. Unsupervised machine learning techniques can be used for clustering purposes. Studies have used two main approaches:

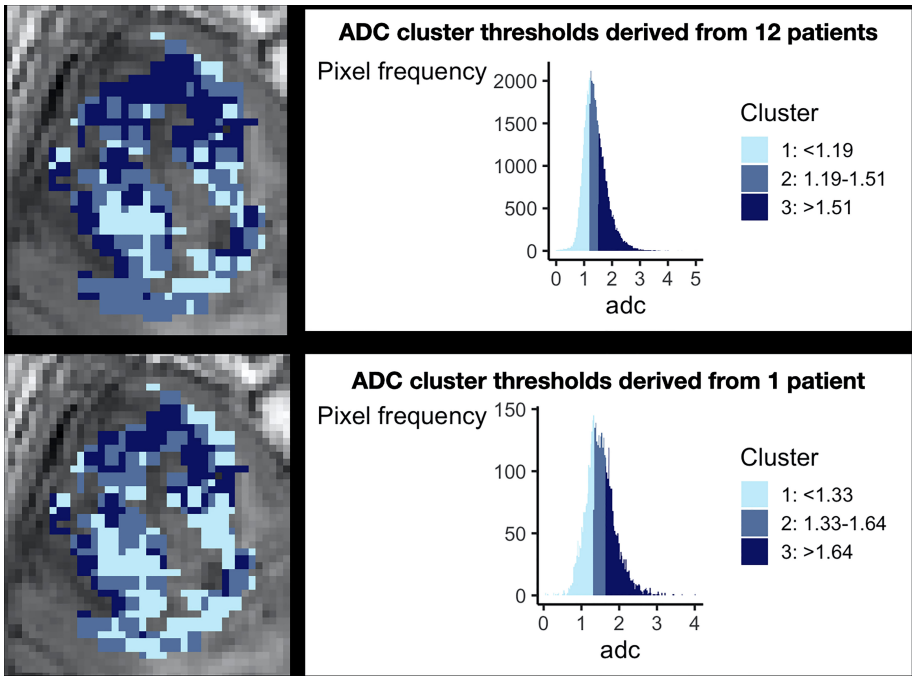


FIGURE 3
The importance of considering group level data during clustering. This figure demonstrates the necessity of combining patient data for clustering. The top panel shows preoperative ADC data from 12 glioblastoma patients after clustering, demonstrating a histogram with a smooth gaussian shape. The bottom panel shows the results of clustering when data from only one individual patient is considered, revealing a more irregular histogram and different cut off values for each cluster. Corresponding cluster regions are displayed visually on the left of each panel. This technical consideration is of particular importance as it has implications for prospective habitat generation in validation cohorts, which is dependent on robust predefined cut offs.

- Imaging sciences approaches: Otsu thresholding analyses the distribution of pixel intensity values to determine threshold value(s) to maximise discrimination between (usually two) pixel classes (51). This has been utilised in studies using structural as well as quantitative MRI sequences (25, 28, 34). Both commercial and open-source software packages are also capable of applying predetermined thresholds to automatically segment a region of interest into different classes. For example, the open-source FMRIB's Automated Segmentation Tool ('FAST') can segment brain images into white matter, grey matter and cerebrospinal fluid (52). Although this tool was used in one study investigating glioblastoma habitats, it is not designed to segment tumour regions (24). These techniques are usually applied to individual MRI sequences and do not therefore account for inter-patient heterogeneity.
- Classical approaches: k-means clustering separates data into clusters by iteratively allocating data points to cluster 'centroids' (numerical points that represent a group of adjacent data points) and updating centroids to minimise the sum of squared distances between data points and corresponding cluster centroids to which they are assigned. This algorithm is very simple and efficient, but sensitive to extreme values, given its reliance on the mean for centroids. It also requires the user to specify the number of clusters required (53). Most existing studies using k-means clustering to generate habitats have not described how the number of clusters (i.e. habitats) was determined (26). This typically requires additional analysis such as the within cluster sum-of-squared or 'Elbow plot' method (54). This method plots the number of clusters (x-axis) against the sum of squared distance between each point and the centroid (y-axis). The optimal number of clusters is the point of maximal 'bend' or 'elbow' (33, 54). Gaussian mixture modelling clusters data by identifying gaussians (i.e. normal distributions) within the data distribution and it can perform hard or soft

clustering of data to different gaussians (53). The presence of multiple gaussians is therefore an assumption of this technique, which may not be accurate. For example, Figure 3 shows only one smooth ADC gaussian when data from all patients is considered. Hierarchical clustering groups together datapoints based on local proximity (53). This method has only been applied to individual patient data rather than group-level data, for which it was designed (27).

In summary therefore, the use of machine learning techniques to generate imaging habitats in glioblastoma requires further evaluation using robust methodology. In particular, studies should justify the number of clusters selected, rather than basing this figure on an arbitrary value. The role of machine learning techniques should also be clarified through comparison to simpler techniques such as 'binning' of data-values into clusters based on quartile or mean values.

Deep-learning. Deep learning (DL) is a subfield of machine learning that is capable of learning which features are most relevant for classification/clustering problems. It is classically described in three stages (55, 56):

1. Input of labelled training data – this is high-dimensional data, which has been assigned labels manually. For example, for a tumour segmentation task, this may be pixels assigned as tumour or brain.
2. Development of neural network – this comprises an input layer, one or more hidden layers and an output layer. In simple terms, the input is mathematically mapped to the output by a series of functions (contained in hidden layers) that try to model the relationship between the two. In imaging research, a 'convolutional' neural network (CNN) is typically used which applies an additional convolution function (also referred to as a kernel) to the input to provide an estimation of spatial relationships (55).
3. Validation of neural network – this step utilises an additional validation dataset to validate the neural network that has been developed.

DL can be implemented in habitat imaging pipelines in at least three ways:

1. Tumour segmentation: habitat imaging requires accurate three-dimensional delineation of the tumour and/or peritumoral oedema, to allow precise monitoring of longitudinal changes and treatment planning (57). Manual segmentation is time consuming and subjective, even in expert hands, with a high inter-rater variability. This was best illustrated in the Multimodal Brain Tumour Image Segmentation Benchmark (BRATS) challenge, which compared glioma segmentation algorithms against expert labelling. The authors found a high degree of disagreement between human raters (58). Approaches utilising a CNN can achieve/exceed performance of experts. For example, in habitat imaging, the ONCOhabitats algorithm proposed by Juan- Albarracín et al. utilises an initial segmentation step incorporating a CNN. The authors developed this utilising 210 high grade glioma scans from the BRATS dataset, basing segmentation on structural sequences (T1 pre and post gadolinium, T2-weighted and FLAIR-MRI). Their method achieved a high sensitivity of up to 87% for whole tumour, and very high specificity of 99% for all tumour regions (enhancing tumour/oedema/whole tumour) (36).
2. Pharmacokinetic model fitting: this is of relevance to techniques such as DCE-MRI. Traditionally, a non-linear least square (NLLS) method is used to fit pharmacokinetic models to the four-dimensional data obtained from DCE-MRI (i.e. 3D volumes acquired serially with time). DL methods such as CNN can produce more precise parameter estimates with less noise, although they are also prone to systematic errors (59).
3. Habitat generation: a difficulty in using DL for habitat generation in glioblastoma is its reliance on labelled data and as a result, DL has not yet been used for this purpose. A comprehensive reference resource with labelled habitats would facilitate the development of robust DL methods for habitat generation (see Discussion).

In summary, DL is an evolving and exciting field, whose methodology could be incorporated into the first arm of habitats pipeline to allow semi-automated tumour segmentation. However, at present, experience with DL is limited and its role remains to be defined.

Habitat volume and sampling. This factor is of particular clinical relevance to validate habitat methodology. In theory, any number of imaging biomarkers can be clustered and combined to produce habitats. However, an increasing number of biomarkers and biomarker clusters decreases the habitat volume limiting the possibility to cross validate the habitat with tumour sampling. Furthermore, the conceptual meaning of habitats may decrease the more biomarkers are used to define them. Bailo et al. utilised image-guided biopsies to sample habitats derived from three biomarkers clustered into two categories each (low/high). However, only 19/31 biopsies they performed contained a single habitat, whereas others contained multiple habitats (34). This would suggest that even fewer

biomarkers should be utilised to allow a large enough volume to allow accurate histological sampling. In the context of glioblastoma, this should be at least 1 mm^3, which is the minimum volume of a brain biopsy (60). In reality, the sampled area is likely to be even larger than this and therefore, without adjusting habitat size, sampling of tissue will include multiple habitats that will confound results. Relating habitats to autopsy specimens should be avoided as these may be obtained several months after the imaging study and sampled areas can be much larger than habitat (24). As habitat samples are likely to be small, efforts must be taken to preserve tumour cell viability. This includes transporting them on dry ice and fixing/snap freezing samples at the earliest possible convenience (61). An alternative strategy to validating habitats, as utilised in other cancers, relates to correlation with metabolic imaging such as PET, although this less widely available (62). In summary, for heterogenous tumours like glioblastoma, habitat sampling and validation is important to guide the development of personalised therapy.

These technical considerations highlight the need for future studies to evaluate divergent methodologies that are not fully explored, to provide reproducible habitats across centres.

Clinical application

A robust and reproducible method of defining glioblastoma habitats has several clinical applications.

Tumour sampling. Habitat guided tumour sampling is possible as habitat maps can be imported into conventional neuronavigation software used in neurosurgical planning (34). This software is capable of image registration but is optimised for structural and functional MRI sequences. To avoid registration errors in this specialist setting, it is therefore important for the final habitat output map to be registered to a structural MRI sequence (typically T1 post gladolinium) prior to its export into neuronavigation software. It can then be used to direct surgical sampling. The location of intra-operative biopsies can be mapped back to MRI scans using the FMRIB Software Library's upcoming Tensor Imaging Registration Library (TIRL) tool, which can act as a bridge between imaging and histopathology (79). Habitat-guided tumour sampling has the potential to reduce spatial heterogeneity between acquired specimens. Furthermore, in glioblastoma patients undergoing biopsy alone, which comprise around 40% of all cases (63), the tumoral yield could be increased by targeting more cellular habitats - with lower ADC values for example. This is also potentially advantageous for genomic sequencing analyses. Treatment resistant habitats could also be sampled, especially in the case of multi-focal and 'butterfly' glioblastomas (that cross the corpus callosum) where a surgical target for biopsy is not always clear.

Diagnostics. Habitat imaging provides an additional tool for radiologists to define a lesion's imaging signature, which could aid diagnostics. This is of particular relevance at present given the increasing emphasis on early time point interventions for newly-diagnosed glioblastoma, including preoperative therapies, which may require imaging diagnosis alone (2). As an example, PreOperative Brain Irradiation in Glioblastoma (POBIG - NCT03582514) is an ongoing phase I trial (led by the senior author of this review) that will evaluate the safety and feasibility of preoperative radiotherapy in newly-diagnosed glioblastoma patients based on imaging diagnosis alone (64). Confirmation of diagnosis is of critical importance in preoperative treatment studies and some have implemented a first step of a pre-resection tumour biopsy to offset the risk of a misdiagnosis (65).

Targeted treatment. The habitat profile of a tumour may correlate with key molecular changes such as O^6-methylguanine-DNA methyltransferase (MGMT) promotor methylation, which could non-invasively aid the selection of patients for future neoadjuvant trials (2, 66). In addition, interventional approaches would benefit from prior knowledge of habitats that have associated aggressive histopathological tumour signatures such hypoxia. Notably, this cancer hallmark is present both microscopically in tumour niches around palisading necrotic regions, but also macroscopically, in hypoperfused areas such as the peri-necrotic rim (8, 67). Treatment-resistant habitats could be targeted with regional dose-boost radiotherapy and/or surgical resection, such as in the case of butterfly lesions where there is discrepancy in surgical decision making (68). This is an important area to explore given the negative results from dose escalation based on structural imaging and the ongoing attempts to improve the outcome by escalating the dose in tumour areas identified on functional imaging (69–71). Habitat-guided radiotherapy dose boost is already being prospectively evaluated in prostate cancer (72). Dynamic assessment of habitat treatment response offers a more personalised approach that allows intensification of treatment only when required, on a regional basis (Figure 1).

Discussion

Habitat imaging in glioblastoma has several potential clinical benefits and applications but there remain a number of technical challenges. Based on the imaging biomarker roadmap, suitable data does not currently exist to evaluate this strategy towards validation and more robust data is required (11).

Existing studies that have derived imaging habitats in glioblastoma patients and studied their associated histological/molecular characteristics are not comprehensive or sufficiently robust (26, 34). There are multiple technical considerations of relevance to both the process of imaging habitat generation and subsequent histological validation, that require further study.

The methodology employed to generate imaging habitats should offer low variation/high repeatability within the same patient in the absence of clinical change when imaging is performed longitudinally. Such repeatability depends on the underlying imaging biomarkers selected and has been demonstrated for quantitative imaging biomarkers derived from diffusion and perfusion MRI (73, 74).

Habitat volume is a key challenge that should be overcome prior to histological validation. Multiple habitat inclusion in image-guided biopsies can lead to non-specific results. For example, the usual inverse correlation between mean diffusivity and cellularity was not observed in the study by Bailo et al., in which over one third of biopsies contained multiple habitats (34). Better characterisation of the biology of habitats could also pave the way for DL techniques to optimize habitat generation. As reviewed above, DL techniques depend on labelled data points. A comprehensive investigation that spatially links histopathological features (e.g. cellularity, perfusion and necrosis) to multi-modal imaging would allow a CNN to be trained that can provide parameter maps relating to these features. This step is of primary importance towards translation and clinical use of habitat imaging, which is otherwise time consuming and reliant on specialist software/expertise.

Habitat imaging in other cancer types including breast, prostate and sarcoma has reached histological or preclinical validation, and even clinical use (62, 72, 75). Some experiences have utilised additional strategies to those reviewed above that merit discussion. Xing et al. described an initial step of qualitatively defining five habitats based on radiologists' assessment of T2/diffusion weighted MRI in 18 patients with biopsy-proven soft tissue sarcoma. As a second step, they then utilised gaussian mixture modelling to create quantitative definitions that described the probability of a pixel belonging to one of these specific habitats. This approach identified a validated necrotic habitat that correlated well with preoperative fluorodeoxyglucose-(FDG)-PET, which increased after preoperative radiotherapy (62). Another approach utilised in a preclinical sarcoma mouse model involves registering fine cut tissue sections to multiparametric imaging, to identify imaging signatures predictive of specific histologically defined habitats (76). This is challenging in glioblastoma patients as it requires *en-bloc* resection, which is only feasible in limited locations within the brain and in most cases will not capture infiltrative components of the tumour periphery (77), or the availability of temporally correlated post-mortem specimens. However, unlike experiences in glioblastoma patients, both of these strategies limit the number of imaging habitats to those apparent clinically.

Understanding the biological meaning of habitats is of direct clinical importance and experience in prostate cancer has demonstrated its value. Stoyanova et al. defined habitats based on DCE/ADC MRI and correlated them with Gleason scores on finely cut prostate cancer sections. Their prior work had identified thresholds based on DCE/ADC that correlated with higher Gleason scores. They identified a habitat that correlated well with a Gleason score of ≥7, representing increased likelihood of cancerous tissue, with an area under curve of 0.8. This habitat is now being prospectively targeted with regional dose boost radiotherapy in a phase II randomised trial (72).

A limitation of concepts presented in this work is reliance on relatively small studies with largely un-validated methodologies. There is wide scope for refinement and validation of imaging habitat techniques in glioblastoma patients specifically given that firstly, multiparametric MRI is a standard of care, and secondly that MRI-guided surgery is routine in the brain (78). Future studies should therefore focus on histologically validating robustly-derived imaging habitats. A generic limitation of studies in other cancer types is the lack of real-time tissue sampling from habitats and reliance on registration of histological sections with imaging. This is not always reliable, given the gantry angle of MRI machines and potential for tissue distortion during slice extraction and preparation.

Conclusion

Habitat imaging is a relatively novel concept that reflects the evolution of imaging biomarkers, to potentially offer a superior means to assess tumour biology and response to treatment in glioblastoma. At present, literature is limited and further studies are required to both robustly generate and validate this technique. This is an important area of research given the multiple clinical applications of habitat imaging that could facilitate more personalised therapy Glioblastoma. Future studies should investigate clustering techniques (machine learning vs. simpler strategies), choice of imaging biomarkers, habitat reproducibility/external validity and means to histologically validate findings, towards the common goal of identifying strategies to overcome treatment-resistance of habitat defined regions.

Data availability statement

All datasets presented in this study are included in the article/Supplementary Material.

Author contributions

MW: drafted manuscript, performed literature review, aided with figures. PVH: drafted manuscript, aided in literature review, reviewed manuscript. EH: aided in literature review and reviewing manuscript draft. K-LL and XZ: produced tables and data for figures, reviewed manuscript draft. AJ: designed

concept, reviewed manuscript. MI: oversaw machine learning sections, reviewed manuscript draft. JO'C: designed concept, aided in processing figures, reviewed manuscript. ID: designed concept, reviewed manuscript. UVH: designed concept, reviewed manuscript. DC and GB: formulated concept of study, obtained original data for figures, oversaw manuscript editing and finalization. All authors contributed to the article and approved the submitted version.

Acknowledgments

The authors acknowledge Istvan Huszar (contributor to the FMRIB Software Library) from the University of Oxford for his assistance with image registration.

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.1037896/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Search strategy. To review the current evidence on imaging habitats, MEDLINE was queried systematically using the search terms described by two independent authors (MW, PVH) with discrepancies resolved through discussion with a third (EH). A total of 15 articles were included in the final review relating to imaging habitats that are summarised in .

References

1. Stupp R, Mason WP, Van Den Bent MJ, Weller M, Fisher B, Taphoorn MJ, et al. Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *N Engl J Med* (2005) 352(10):987–96. doi: 10.1056/NEJMoa043330

2. Waqar M, Roncaroli F, Lehrer EJ, Palmer JD, Villanueva-Meyer J, Braunstein S, et al. Early therapeutic interventions for newly diagnosed glioblastoma: Rationale and review of the literature. *Curr Oncol Rep* (2022) 24(3):311–24. doi: 10.1007/s11912-021-01157-0

3. Waqar M, Trifiletti DM, Mcbain C, O'connor J, Coope DJ, Akkari L, et al. Rapid early progression (REP) of glioblastoma is an independent negative prognostic factor: Results from a systematic review and meta-analysis. *Neurooncol Adv* (2022) 4(1):vdac075.

4. Barthel FP, Johnson KC, Varn FS, Moskalik AD, Tanner G, Kocakavuk E, et al. Longitudinal molecular trajectories of diffuse glioma in adults. *Nature* (2019) 576(7785):112–20. doi: 10.1038/s41586-019-1775-1

5. Sottoriva A, Spiteri I, Piccirillo SG, Touloumis A, Collins VP, Marioni JC, et al. Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. *Proc Natl Acad Sci U S A* (2013) 110(10):4009–14. doi: 10.1073/pnas.1219747110

6. Ravi VM, Will P, Kueckelhaus J, Sun N, Joseph K, Salie H, et al. Spatially resolved multi-omics deciphers bidirectional tumor-host interdependence in glioblastoma. *Cancer Cell* (2022) 40(6):639–55.e13. doi: 10.1016/j.ccell.2022.05.009

7. Aderetti DA, Hira VVV, Molenaar RJ, and Van Noorden CJF. The hypoxic peri-arteriolar glioma stem cell niche, an integrated concept of five types of niches in human glioblastoma. *Biochim Biophys Acta Rev Cancer* (2018) 1869(2):346–54. doi: 10.1016/j.bbcan.2018.04.008

8. Lam KHB, Leon AJ, Hui W, Lee SC, Batruch I, Faust K, et al. Topographic mapping of the glioblastoma proteome reveals a triple-axis model of intra-tumoral heterogeneity. *Nat Commun* (2022) 13(1):116. doi: 10.1038/s41467-021-27667-w

9. Yang Y, Han Y, Zhao S, Xiao G, Guo L, Zhang X, et al. Spatial heterogeneity of edema region uncovers survival-relevant habitat of glioblastoma. *Eur J Radiol* (2022) 154:110423. doi: 10.1016/j.ejrad.2022.110423

10. Choi SW, Cho HH, Koo H, Cho KR, Nenning KH, Langs G, et al. Multi-habitat radiomics unravels distinct phenotypic subtypes of glioblastoma with clinical and genomic significance. *Cancers (Basel)* (2020) 12(7). doi: 10.3390/cancers12071707

11. O'Connor JP, Aboagye EO, Adams JE, Aerts HJ, Barrington SF, Beer AJ, et al. Imaging biomarker roadmap for cancer studies. *Nat Rev Clin Oncol* (2017) 14(3):169–86. doi: 10.1038/nrclinonc.2016.162

12. Li KL, Djoukhadar I, Zhu X, Zhao S, Lloyd S, Mccabe M, et al. Vascular biomarkers derived from dynamic contrast-enhanced MRI predict response of vestibular schwannoma to antiangiogenic therapy in type 2 neurofibromatosis. *Neuro-oncology* (2016) 18(2):275–82. doi: 10.1093/neuonc/nov168

13. Singh G, Manjila S, Sakla N, True A, Wardeh AH, Beig N, et al. Radiomics and radiogenomics in gliomas: a contemporary update. *Br J Cancer* (2021) 125(5):641–57. doi: 10.1038/s41416-021-01387-w

14. 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

15. Xie C, Yang P, Zhang X, Xu L, Wang X, Li X, et al. Sub-Region based radiomics analysis for survival prediction in oesophageal tumours treated by definitive concurrent chemoradiotherapy. *EBioMedicine* (2019) 44:289–97. doi: 10.1016/j.ebiom.2019.05.023

16. Khalifa J, Tensaouti F, Lotterie JA, Catalaa I, Chaltiel L, Benouaich-Amiel A, et al. Do perfusion and diffusion MRI predict glioblastoma relapse sites following chemoradiation? *J Neurooncol* (2016) 130(1):181–92. doi: 10.1007/s11060-016-2232-8

17. Napel S, Mu W, Jardim-Perassi BV, Aerts H, Gillies RJ. Quantitative imaging of cancer in the postgenomic era: Radio(geno)mics, deep learning, and habitats. *Cancer* (2018) 124(24):4633–49. doi: 10.1002/cncr.31630

18. Spence AM, Muzi M, Swanson KR, O'sullivan F, Rockhill JK, Rajendran JG, et al. Regional hypoxia in glioblastoma multiforme quantified with [18F] fluoromisonidazole positron emission tomography before radiotherapy: Correlation with time to progression and survival. *Clin Cancer Res* (2008) 14(9):2623–30. doi: 10.1158/1078-0432.CCR-07-4995

19. Pruis IJ, Koene SR, Van Der Voort SR, Incekara F, Vincent A, Van Den Bent MJ, et al. Noninvasive differentiation of molecular subtypes of adult nonenhancing glioma using MRI perfusion and diffusion parameters. *Neurooncol Adv* (2022) 4(1):vdac023. doi: 10.1093/noajnl/vdac023

20. Lee DH, Park JE, Kim N, Park SY, Kim YH, Cho YH, et al. Tumor habitat analysis by magnetic resonance imaging distinguishes tumor progression from

radiation necrosis in brain metastases after stereotactic radiosurgery. *Eur Radiol* (2022) 32(1):497–507. doi: 10.1007/s00330-021-08204-1

21. O'Connor JP, Rose CJ, Waterton JC, Carano RA, Parker GJ, Jackson A, et al. Imaging intratumor heterogeneity: role in therapy response, resistance, and clinical outcome. *Clin Cancer Res* (2015) 21(2):249–57. doi: 10.1158/1078-0432.CCR-14-0990

22. Lee J, Narang S, Martinez J, Rao G, Rao A. Spatial habitat features derived from multiparametric magnetic resonance imaging data are associated with molecular subtype and 12-month survival status in glioblastoma multiforme. *PloS One* (2015) 10(9):e0136557. doi: 10.1371/journal.pone.0136557

23. Lee J, Narang S, Martinez JJ, Rao G, Rao A. Associating spatial diversity features of radiologically defined tumor habitats with epidermal growth factor receptor driver status and 12-month survival in glioblastoma: Methods and preliminary investigation. *J Med Imaging (Bellingham)* (2015) 2(4):041006. doi: 10.1117/1.JMI.2.4.041006

24. McGarry SD, Hurrell SL, Kaczmarowski AL, Cochran EJ, Connelly J, Rand SD, et al. Magnetic resonance imaging-based radiomic profiles predict patient prognosis in newly diagnosed glioblastoma before therapy. *Tomography* (2016) 2 (3):223–8. doi: 10.18383/j.tom.2016.00250

25. Zhou M, Chaudhury B, Hall LO, Goldgof DB, Gillies RJ, Gatenby RA, et al. Identifying spatial imaging biomarkers of glioblastoma multiforme for survival group prediction. *J Magn Reson Imaging* (2017) 46(1):115–23. doi: 10.1002/jmri.25497

26. Dextraze K, Saha A, Kim D, Narang S, Lehrer M, Rao A, et al. Spatial habitats from multiparametric MR imaging are associated with signaling pathway activities and survival in glioblastoma. *Oncotarget* (2017) 8(68):112992–3001. doi: 10.18632/oncotarget.22947

27. You D, Kim MM, Aryal MP, Parmar H, Piert M, Lawrence TS, et al. Tumor image signatures and habitats: A processing pipeline of multimodality metabolic and physiological images. *J Med Imaging (Bellingham)* (2018) 5(1):011009.

28. Stringfield O, Arrington JA, Johnston SK, Rognin NG, Peeri NC, Balagurunathan Y, et al. Multiparameter MRI predictors of long-term survival in glioblastoma multiforme. *Tomography* (2019) 5(1):135–44. doi: 10.18383/j.tom.2018.00052

29. Li C, Yan JL, Torheim T, Mclean MA, Boonzaier NR, Zou , et al. Low perfusion compartments in glioblastoma quantified by advanced magnetic resonance imaging and correlated with patient survival. *Radiother Oncol* (2019) 134:17–24. doi: 10.1016/j.radonc.2019.01.008

30. Del Mar Alvarez-Torres M, Juan-Albarracin J, Fuster-Garcia E, Bellvis-Bataller F, Lorente D, Reynes G, et al. Robust association between vascular habitats and patient prognosis in glioblastoma: An international multicenter study. *J Magn Reson Imaging* (2020) 51(5):1478–86. doi: 10.1002/jmri.26958

31. Park JE, Kim HS, Kim N, Kim YH, Kim JH, Kim E, et al. Low conductivity on electrical properties tomography demonstrates unique tumor habitats indicating progression in glioblastoma. *Eur Radiol* (2021) 31(9):6655–65. doi: 10.1007/s00330-021-07976-w

32. Park JE, Kim HS, Kim N, Park SY, Kim YH, Kim JH, et al. Spatiotemporal heterogeneity in multiparametric physiologic MRI is associated with patient outcomes in IDH-wildtype glioblastoma. *Clin Cancer Res* (2021) 27(1):237–45. doi: 10.1158/1078-0432.CCR-20-2156

33. Xu X, Samaras D, Prasanna P. Radiologically defined tumor-habitat adjacency as a prognostic biomarker in glioblastoma. *Annu Int Conf IEEE Eng Med Biol Soc* (2021) 2021:3998–4001. doi: 10.1109/EMBC46164.2021.9629779

34. Bailo M, Pecco N, Callea M, Scifo P, Gagliardi F, Presotto L, et al. Decoding the heterogeneity of malignant gliomas by PET and MRI for spatial habitat analysis of hypoxia, perfusion, and diffusion imaging: A preliminary study. *Front Neurosci* (2022) 16:885291. doi: 10.3389/fnins.2022.885291

35. Juan-Albarracin J, Fuster-Garcia E, Perez-Girbes A, Aparici-Robles F, Alberich-Bayarri A, Revert-Ventura A, et al. Glioblastoma: Vascular habitats detected at preoperative dynamic susceptibility-weighted contrast-enhanced perfusion MR imaging predict survival. *Radiology* (2018) 287(3):944–54. doi: 10.1148/radiol.2017170845

36. Juan-Albarracin J, Fuster-Garcia E, Garcia-Ferrando GA, Garcia-Gomez JM. ONCOhabitats: A system for glioblastoma heterogeneity assessment through MRI. *Int J Med Inform* (2019) 128:53–61. doi: 10.1016/j.ijmedinf.2019.05.002

37. Jenkinson M, Chappell M. *Introduction to neuroimaging analysis. 1st ed.* New York, NY: Oxford University Press (2018) xvii(258).

38. Lewis D, Roncaroli F, Agushi E, Mosses D, Williams R, Li KL, et al. Inflammation and vascular permeability correlate with growth in sporadic vestibular schwannoma. *Neuro-oncology* (2019) 21(3):314–25. doi: 10.1093/neuonc/noy177

39. Waqar M, Lewis D, Agushi E, Gittins M, Jackson A, Coope D. Cerebral and tumoral blood flow in adult gliomas: A systematic review of results from magnetic resonance imaging. *Br J Radiol* (2021) 94(1125):20201450. doi: 10.1259/bjr.20201450

40. McGuire SA, Wijtenburg SA, Sherman PM, Rowland LM, Ryan M, Sladky JH, et al. Reproducibility of quantitative structural and physiological MRI measurements. *Brain Behav* (2017) 7(9):e00759. doi: 10.1002/brb3.759

41. Nerland S, Jorgensen KN, Nordhoy W, Maximov Ii Bugge RaB , Westlye LT, et al. Multisite reproducibility and test-retest reliability of the T1w/T2w-ratio: A comparison of processing methods. *Neuroimage* (2021) 245:118709. doi: 10.1016/j.neuroimage.2021.118709

42. Kong Z, Yan C, Zhu R, Wang J, Wang Y, Wang Y, et al. Imaging biomarkers guided anti-angiogenic therapy for malignant gliomas. *NeuroImage Clin* (2018) 20:51–60. doi: 10.1016/j.nicl.2018.07.001

43. Sourbron SP, Buckley DL. Classic models for dynamic contrast-enhanced MRI. *NMR BioMed* (2013) 26(8):1004–27. doi: 10.1002/nbm.2940

44. Sourbron SP, Buckley DL. On the scope and interpretation of the tofts models for DCE-MRI. *Magn Reson Med* (2011) 66(3):735–45. doi: 10.1002/mrm.22861

45. Li KL, Lewis D, Coope DJ, Roncaroli F, Agushi E, Pathmanaban ON, et al. The LEGATOS technique: A new tissue-validated dynamic contrast-enhanced MRI method for whole-brain, high-spatial resolution parametric mapping. *Magn Reson Med* (2021). doi: 10.1002/mrm.28842

46. Kwee TC, Galban CJ, Tsien C, Junck L, Sundgren PC, Ivancevic MK, et al. Comparison of apparent diffusion coefficients and distributed diffusion coefficients in high-grade gliomas. *J Magn Reson Imaging* (2010) 31(3):531–7. doi: 10.1002/jmri.22070

47. Shukla-Dave A, Obuchowski NA, Chenevert TL, Jambawalikar S, Schwartz LH, Malyarenko D, et al. Quantitative imaging biomarkers alliance (QIBA) recommendations for improved precision of DWI and DCE-MRI derived biomarkers in multicenter oncology trials. *J Magn Reson Imaging* (2019) 49(7):e101–e21.

48. *OSIPI task force 1.2: DCE/DSC software inventory* . Available at: https://osipi.org/task-force-1-2/.

49. Scouten A, Papademetris X, Constable RT. Spatial resolution, signal-to-noise ratio, and smoothing in multi-subject functional MRI studies. *Neuroimage* (2006) 30(3):787–93. doi: 10.1016/j.neuroimage.2005.10.022

50. Winfield JM, Payne GS, Weller A, Desouza NM. DCE-MRI, DW-MRI, and MRS in cancer: Challenges and advantages of implementing qualitative and quantitative multi-parametric imaging in the clinic. *Top Magn Reson Imaging* (2016) 25(5):245–54. doi: 10.1097/RMR.0000000000000103

51. Xue JH, Titterington DM. T-tests, f-tests and otsu's methods for image thresholding. *IEEE Trans Image Process* (2011) 20(8):2392–6.

52. Zhang Y, Brady M, Smith S. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Trans Med Imaging* (2001) 20(1):45–57. doi: 10.1109/42.906424

53. Kassambara A. *Practical guide to cluster analysis in r: unsupervised machine learning*. (2017).

54. Sammouda R, El-Zaart A. An optimized approach for prostate image segmentation using K-means clustering algorithm with elbow method. *Comput Intell Neurosci* (2021) 2021:4553832. doi: 10.1155/2021/4553832

55. Yamashita R, Nishio M, Do RKG, Togashi K. Convolutional neural networks: An overview and application in radiology. *Insights Imaging* (2018) 9 (4):611–29. doi: 10.1007/s13244-018-0639-9

56. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* (2015) 521 (7553):436–44. doi: 10.1038/nature14539

57. Shah GD, Kesari S, Xu R, Batchelor TT, O'neill AM, Hochberg FH, et al. Comparison of linear and volumetric criteria in assessing tumor response in adult high-grade gliomas. *Neuro-oncology* (2006) 8(1):38–46. doi: 10.1215/S1522851705000529

58. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, et al. The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Trans Med Imaging* (2015) 34(10):1993–2024. doi: 10.1109/TMI.2014.2377694

59. Ottens T, Barbieri S, Orton MR, Klaassen R, Van Laarhoven HWM, Crezee H, et al. Deep learning DCE-MRI parameter estimation: Application in pancreatic cancer. *Med Image Anal* (2022) 80:102512. doi: 10.1016/j.media.2022.102512

60. Katzendobler S, Do A, Weller J, Dorostkar MM, Albert NL, Forbrig R, et al. Diagnostic yield and complication rate of stereotactic biopsies in precision medicine of gliomas. *Front Neurol* (2022) 13:822362. doi: 10.3389/fneur.2022.822362

61. Mook S, Bonnefoi H, Pruneri G, Larsimont D, Jaskiewicz J, Sabadell MD, et al. Daily clinical practice of fresh tumour tissue freezing and gene expression profiling; logistics pilot study preceding the MINDACT trial. *Eur J Cancer* (2009) 45(7):1201–8. doi: 10.1016/j.ejca.2009.01.004

62. Xing S, Freeman CR, Jung S, Turcotte R, Levesque IR. Probabilistic classification of tumour habitats in soft tissue sarcoma. *NMR Biomed* (2018) 31 (11):e4000. doi: 10.1002/nbm.4000

63. Bauchet L, Mathieu-Daude H, Fabbro-Peray P, Rigau V, Fabbro M, Chinot O, et al. Oncological patterns of care and outcome for 952 patients with newly

diagnosed glioblastoma in 2004. *Neuro-oncology* (2010) 12(7):725–35. doi: 10.1093/neuonc/noq030

64. *PreOperative brain irradiation in glioblastoma (POBIG)* (2022). Available at: https://clinicaltrials.gov/ct2/show/NCT03582514.

65. *Neoadjuvant chemoradiation for resectable glioblastoma (NeoGlio)* (2022). Available at: https://clinicaltrials.gov/ct2/show/NCT04209790.

66. Jiang H, Zeng W, Ren X, Cui Y, Li M, Yang K, et al. Super-early initiation of temozolomide prolongs the survival of glioblastoma patients without gross-total resection: A retrospective cohort study. *J Neurooncol* (2019) 144(1):127–35. doi: 10.1007/s11060-019-03211-1

67. Collet S, Guillamo JS, Berro DH, Chakhoyan A, Constans JM, Lechapt-Zalcman E, et al. Simultaneous mapping of vasculature, hypoxia, and proliferation using dynamic susceptibility contrast MRI, (18)F-FMISO PET, and (18)F-FLT PET in relation to contrast enhancement in newly diagnosed glioblastoma. *J Nucl Med* (2021) 62(10):1349–56. doi: 10.2967/jnumed.120.249524

68. Chawla S, Kavouridis VK, Boaro A, Korde R, Amaral Medeiros S, Edrees H, et al. Surgery vs. biopsy in the treatment of butterfly glioblastoma: A systematic review and meta-analysis. *Cancers (Basel)* (2022) 14(2). doi: 10.3390/cancers14020314

69. Harat M, Blok M, Miechowicz I, Wiatrowska I, Makarewicz K, Malkowski B, et al. Safety and efficacy of irradiation boost based on 18F-FET-PET in patients with newly diagnosed glioblastoma. *Clin Cancer Res* (2022) 28(14):3011–20. doi: 10.1158/1078-0432.CCR-22-0171

70. Laack NN, Pafundi D, Anderson SK, Kaufmann T, Lowe V, Hunt C, et al. Initial results of a phase 2 trial of (18)F-DOPA PET-guided dose-escalated radiation therapy for glioblastoma. *Int J Radiat Oncol Biol Physics* (2021) 110 (5):1383–95. doi: 10.1016/j.ijrobp.2021.03.032

71. Gondi V, Pugh S, Tsien C, Chenevert T, Gilbert M, Omuro A, et al. Radiotherapy (RT) dose-intensification (DI) using intensity-modulated RT (IMRT) versus standard-dose (SD) RT with temozolomide (TMZ) in newly diagnosed glioblastoma (GBM): Preliminary results of NRG oncology BN001. *Int J Radiat Oncol Biol Physics* (2020) 108(3):S22–S3. doi: 10.1016/j.ijrobp.2020.07.2109

72. Stoyanova R, Chinea F, Kwon D, Reis IM, Tschudi Y, Parra NA, et al. An automated multiparametric MRI quantitative imaging prostate habitat risk scoring system for defining external beam radiation therapy boost volumes. *Int J Radiat Oncol Biol Physics* (2018) 102(4):821–9. doi: 10.1016/j.ijrobp.2018.06.003

73. Paldino MJ, Barboriak D, Desjardins A, Friedman HS, Vredenburgh JJ. Repeatability of quantitative parameters derived from diffusion tensor imaging in patients with glioblastoma multiforme. *J Magn Reson Imaging* (2009) 29(5):1199–205. doi: 10.1002/jmri.21732

74. Jafari-Khouzani K, Emblem KE, Kalpathy-Cramer J, Bjornerud A, Vangel MG, Gerstner ER, et al. Repeatability of cerebral perfusion using dynamic susceptibility contrast MRI in glioblastoma patients. *Transl Oncol* (2015) 8 (3):137–46. doi: 10.1016/j.tranon.2015.03.002

75. Jardim-Perassi BV, Huang S, Dominguez-Viqueira W, Poleszczuk J, Budzevich MM, Abdalah MA, et al. Multiparametric MRI and coregistered histology identify tumor habitats in breast cancer mouse models. *Cancer Res* (2019) 79(15):3952–64. doi: 10.1158/0008-5472.CAN-19-0213

76. Jardim-Perassi BV, Mu W, Huang S, Tomaszewski MR, Poleszczuk J, Abdalah MA, et al. Deep-learning and MR images to target hypoxic habitats with evofosfamide in preclinical models of sarcoma. *Theranostics* (2021) 11 (11):5313–29. doi: 10.7150/thno.56595

77. Burger PC, Heinz ER, Shibata T, Kleihues P. Topographic anatomy and CT correlations in the untreated glioblastoma multiforme. *J Neurosurg* (1988) 68 (5):698–704. doi: 10.3171/jns.1988.68.5.0698

78. Thomas NWD, Sinclair J. Image-guided neurosurgery: History and current clinical applications. *J Med Imaging Radiat Sci* (2015) 46(3):331–42. doi: 10.1016/j.jmir.2015.06.003

79. Huszar IN, Pallebage-Gamarallage M, Foxley S, Tendler BC, Leonte A, Hiemstra M, et al. Tensor Image registration library: Automated non-linear registration of sparsely sampled histological specimens to post-mortem MRI of the Whole Human Brain. *bioRxiv* 849570

frontiers | Frontiers in Oncology

TYPE Systematic Review
PUBLISHED 24 November 2022
DOI 10.3389/fonc.2022.1010643

OPEN ACCESS

EDITED BY
Bahram Mohajer,
Johns Hopkins Medicine, United States

REVIEWED BY
Ali Ghasemi,
Johns Hopkins Medicine, United States
AmirHussein Abdolalizadeh,
Tehran University of Medical Sciences, Iran

*CORRESPONDENCE
Dorota Bartusik-Aebisher
dbartusikaebisher@ur.edu.pl

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

RECEIVED 03 August 2022
ACCEPTED 31 October 2022
PUBLISHED 24 November 2022

CITATION
Micek M, Aebisher D, Surówka J, Bartusik-Aebisher D and Madera M (2022) Applications of T_1 and T_2 relaxation time calculation in tissue differentiation and cancer diagnostics—a systematic literature review. *Front. Oncol.* 12:1010643. doi: 10.3389/fonc.2022.1010643

Applications of T_1 and T_2 relaxation time calculation in tissue differentiation and cancer diagnostics—a systematic literature review

Marta Micek[1], David Aebisher[2], Justyna Surówka[1], Dorota Bartusik-Aebisher[3]* and Michał Madera[1]

[1]SoftSystem Sp. z o.o, Rzeszow, Poland, [2]Department of Photomedicine and Physical Chemistry, Medical College of The University of Rzeszow, Rzeszow, Poland, [3]Department of Biochemistry and General Chemistry, Medical College of The University of Rzeszow, Rzeszow, Poland

Introduction: The purpose of this review was to summarize current applications of non-contrast-enhanced quantitative magnetic resonance imaging (qMRI) in tissue differentiation, considering healthy tissues as well as comparisons of malignant and benign samples. The analysis concentrates mainly on the epithelium and epithelial breast tissue, especially breast cancer.

Methods: A systematic review has been performed based on current recommendations by publishers and foundations. An exhaustive overview of currently used techniques and their potential in medical sciences was obtained by creating a search strategy and explicit inclusion and exclusion criteria.

Results and Discussion: PubMed and Elsevier (Scopus & Science Direct) search was narrowed down to studies reporting T1 or T2 values of human tissues, resulting in 404 initial candidates, out of which roughly 20% were found relevant and fitting the review criteria. The nervous system, especially the brain, and connective tissue such as cartilage were the most frequently analyzed, while the breast remained one of the most uncommon subjects of studies. There was little agreement between published T1 or T2 values, and methodologies and experimental setups differed strongly. Few contemporary (after 2000) resources have been identified that were dedicated to studying the relaxation times of tissues and their diagnostic applications. Most publications concentrate on recommended diagnostic standards, for example, breast acquisition of T1- or T2-weighted images using gadolinium-based contrast agents. Not enough data is available yet to decide how repeatable or reliable analysis of relaxation times is in diagnostics, so it remains mainly a research topic. So far, qMRI might be recommended as a diagnostic help providing general insight into the nature of lesions (benign vs. malignant). However, additional means are generally necessary to differentiate between specific lesion types.

KEYWORDS

MRI, cancer, diagnostics, T1 relaxation time, T2 relaxation time

1 Introduction

Magnetic resonance imaging (MRI) has been widely used since the seventies, and since almost the same time, quantitative magnetic resonance imaging (qMRI) techniques have been developed to assess the relaxatory parameters of tissues (Figure 1). Relaxation time calculation can nowadays be used in various applications, starting with relatively simple cases of cartilage degradation and ending up helping diagnose and contain the most dangerous cancers.

qMRI departs pretty significantly from the original approach, dedicated at best to obtain single T_1- or T_2-weighted images. In opposition to performing a single acquisition with set parameters of echo and repetition time, a series of scans are acquired in time with varying echo times for transverse relaxation or repetition times for longitudinal relaxation. Such an approach allows tracking the relative change of magnetization, beginning with the maximum (T_2) or minimum (T_1) signal strength at the beginning of a study and then calculating 33% (T_2) or 67% (T1) of it, passing through consecutive time points. As a result, a signal change in time is acquired, and it is possible to calculate other parameters, such as the T_1/T_2 ratio or relaxation rates R_1 and R_2. Obtained results can be standardized, making it easy to compare relaxation curves and parameters between samples, devices, and studies.

As much as this description is simplified due to the multiplicity of sequences and practices in use, it should give an impression of one strength of qMRI: mathematics and models can, in certain cases, describe biological phenomena in more detail or maybe even completely different than a human eye. Due to that fact, the review has been written to get an overview of the potential usefulness of qMRI in clinical settings, as it might one day provide support for medical practitioners when it comes to tissue characterization and differentiation.

In this review, breast cancer and breast tissues were of primary interest. Due to limited literature on studies without the use of contrast agents, other tissues and organs were also considered; such an approach might be justified by the heterogeneous structure of the breast, built of epithelium, fat, and being very diverse when it comes to proportions of these components, which might affect results of relaxation time calculation. A few publications report successful differentiation or diagnosis of breast features when using qMRI, contrast-enhanced or not.

Abbreviations: BOLD MRI, blood oxygenation level-dependent MRI; CT, computer tomography; DCE-MRI, dynamic contrast-enhanced MRI; DWI, diffusion-weighted imaging; FSE, fast spin echo; MRI, magnetic resonance imaging; MT, magnetization transfer; PAI, photoacoustic imaging; PET, positron emission tomography; qMRI, quantitative magnetic resonance imaging; ROI, a region of interest; T_1w, (images) T_1-weighted; T_2w, (images) T_2-weighted.

It must be stated that qMRI currently is not a recommended diagnostic method according to different authorities and organizations. The European Society of Breast Cancer Specialists (EUSOMA) 2008 produced a statement regarding, in their opinion, the best practices for diagnosing and treating breast cancer (1). While the publication mentions the acquisition of T_1- and T_2-weighted images using gadolinium-based contrast agents, all other practices are considered experimental and might be used, at best, as support for the basic analysis. Such techniques included diffusion-weighted imaging (DWI) and analysis of apparent diffusion coefficient (ADC), but there was no mention of qMRI, no matter the application.

Apart from these guidelines, the American Society of Breast Surgeons reminds us that MRI is not a modality of the first choice when screening or diagnosing patients unless other methods, such as x-ray mammography or ultrasonography, fail (2).

The European Commission Initiative on Breast Cancer (ECIBC) Guidelines Development Group (GDG) also presents a series of recommendations and suggestions regarding breast cancer screening, diagnosis, and treatment (3). Magnetic resonance imaging is generally described as a method "with very low certainty of evidence" compared with mammography (4).

It should be mentioned that qMRI is not the only modality currently being tested when it comes to screening for or diagnosing breast cancer. This is caused mainly by low mammography performance when applied to dense or extremely dense breast tissue. In such cases, many approaches have been tested, including ultrasound or X-ray-based techniques (5). Dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI), currently a standard for high-sensitivity breast screening, comes with certain disadvantages, such as the need for intravenous contrast agent administration, which might result in rare, although possible, allergic reactions and is considered a more inconvenient protocol than mammography—a golden standard in screening programs.

Diffusion magnetic resonance imaging (dMRI) may be implemented in breast cancer imaging. The European Society of Breast Radiology (EUSOBI) recommends dMRI as a technique complementary to DCE-MRI. The organization opts for including standardized dMRI protocols in the Breast Imaging Reporting and Data System (BI-RADS). The main advantage of using dMRI together with other modalities is the acquisition of additional information on tissue metabolism and processes related to tissue perfusion, whereas the low resolution of dMRI makes it less informative if not used in conjunction with more precise imaging protocols.

Partridge et al. (6) described in detail the procedures and challenges related to dMRI. Fat suppression is essential but difficult due to its abundance in the breast. Despite that, and frequent other visual artifacts populating diffusion-weighted images, it still performs satisfyingly. One meta-analysis (7) reports pooled sensitivity of 84% when using dMRI to

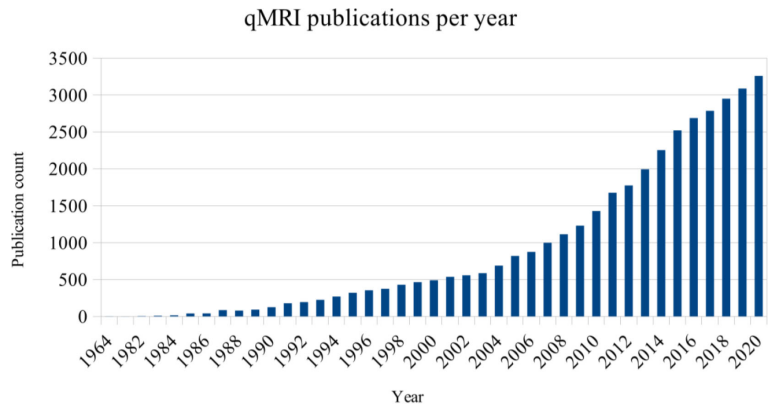


FIGURE 1
Number of entries in PubMed regarding "quantitative MRI", by year of publication.

discriminate benign from malignant lesions. Another meta-analysis confirms good dMRI performance in supporting the analysis of DCE-MRI data (8). Multiple publications report correlations existing between dMRI-derived apparent diffusion coefficient (ADC) and tissue structure, where malignant changes, because of their cellular structure, result in lower ADC values (9, 10). Similar processes take part in non-malignant changes, such as ductal carcinoma *in situ* (DCIS), which can be differentiated from surrounding tissues with the use of dMRI (11).

The ADC has been suggested as a diagnostic biomarker in therapy assessment, allowing to differentiate between responders and non-responders to neoadjuvant treatment before changes in lesion size became visible (12). The ADC values obtained from samples were deemed repeatable, and their changes were significant after a month since the beginning of the therapy.

In one study, dedicated to performance assessment of mammography, DCE-MRI, and diffusion-weighted images, it was discovered that DCE-MRI was the most useful modality in cancer detection by professionals (13). The use of mammography resulted in the least precise predictions by observers, whereas decisions based on mammography together with diffusion-weighted and T2-weighted images were moderately correct. It means that if no DCE-MRI examination is possible, other techniques might still increase the sensitivity of screening.

Another review, by Amornsiripanitch et al. (14), talks specifically about non-enhanced MRI (without the use of contrast agents). Due to restricted water particle movement in cancerous tissues, malignant changes appear hyperintense in dMRI. The use of diffusion-weighted imaging, requiring less time, effort, and financial resources than DCE-MRI, seems to be a promising supporting technique for mammography screening with 89% effectiveness of contrast-enhanced methods (15). It also seemed that dMRI was resistant to factors affecting other modalities and consisting of menopausal status, menstrual cycle, breast density, or lesion size, although the last one might still be important due to the low resolution of dMRI (16, 17).

Other approaches to breast imaging involve coupling MRI together with positron emission tomography (PET) or performing multiparametric MRI (mpMRI), based on a simultaneous analysis of images from complementary modalities (18).

The use of complementary imaging methods is also possible in the screening and diagnosis of prostate and brain cancers and tumors (19, 20). Specific methods have not yet gained similar popularity but seem to be employed in specific clinical trial settings (21). These methods include dynamic susceptibility contrast MRI (DSC-MRI), chemical exchange saturation transfer (CEST), and hyperpolarized MRI.

The scarcity of qMRI publications regarding breast tissues may be a result of factors such as accessibility to equipment and software capable of making high-quality measurements. Performing non-standard procedures requires highly trained staff, and any studies involving patients by engaging them in full or by using samples obtained from biopsies requires additional approvals from relevant ethics committees, not to mention the patients' consent.

By performing this review, we tried to get an overview of possible applications and capabilities of qMRI as well as verify how easy or difficult it is to query scientific databases and how much relevant information can be acquired. By carefully preparing search phrases and specification of acceptable search criteria, such as tissue and modality type, multiple unrelated publications were filtered out not because of their improper quality but often due to unfortunate wording or unspecific writing. Through a systematic review, it is also possible to learn predominant practices when performing qMRI experiments, which might be a suggestion for future researchers, as there are numerous possible combinations of qMRI scan parameters and procedures.

The systematic literature review methodology used in this work is presented in Section 2. Outlined was the justification for the literature review with the research questions and search query statements. The overview of the search process was summarized with the PRISMA diagram shown in Section 3. Next, Section 4 shows previous work related to reviews on the main subject topic, followed by the analysis of the original publications. The conclusions are presented in Section 5, which provides information on good practice and research trends in applications of calculating the relaxation times T1 and T2 in tissue differentiation and cancer diagnosis.

2 Structure of the review

An attempt was made to follow recommendations regarding conducting systematic reviews, although not all were met due to limited time and resources. As guidance, the Cochrane Handbook (22) served as a valuable source of knowledge and suggestions from BioMed Central (23).

2.1 Justification for the review

Application of qMRI might result in numerous benefits when properly applied, such as

- ease of data standardization and comparison due to calculation of relative signal intensity and T_1/T_2 proportions,
- support in ROI selection based on tissue properties, especially in areas where it is difficult to tell apart tissue types or malignant changes on weighted and other MRI scans,
- no need for contrasting agents when assessing patients with allergy to gadolinium or kidney issues; go-to solution for *ex vivo* studies, where DCE-MRI (dynamic contrast-enhanced MRI) cannot be applied.

All the promises and benefits make it necessary to ask questions such as the following: is qMRI really that efficient when differentiating tissues, or could it cause more harm than good when applied, especially incorrectly and without sufficient knowledge? Are T_1 and T_2 constants uniform or diverse enough across tissues, samples, and patients to be indicators of malignancies?

2.2 Questions asked

Before the beginning of the review process, the following questions were formulated:

1. What approaches are in use when applying qMRI to tissue differentiation?
2. Which tissues and organs are the most frequently analyzed?
3. Does practice or efficiency differ for epithelial tissue compared with other tissue types?
4. Are T1 and T2 time calculation results significant and comparable between samples, patients, and studies?
5. Which database or search engine is the complete source of topic-related literature?

The reasoning behind categorizing by tissue was forced due to the diverse structure of the human body. Depending on the tissue, different approaches might be used to compensate for specific phenomena, such as blood flow in the myocardium or oxygen and carbon dioxide exchange in the lungs.

A more detailed description of subjects of interest is found in Table 1, where the research questions were formulated to fit the Population-Intervention-Comparison-Outcome-Context (PICOC) scheme.

2.3 Source selection

Based on personal experience and suggestions from specialists, the largest biomedical databases were chosen for the review:

- PubMed (24)
- Elsevier [Scopus (25)/Science Direct (26)]

TABLE 1 Review questions in light of the PICOC scheme.

Population	Patients and study participants of all genders and ages; any human tissue samples being analyzed—healthy and bearing signs of pathological processes. Exceptionally, human cell cultures.
Intervention	Calculation of T_1 and T_2 constants and/or T_1/T_2 ratio based on qMRI data to identify or differentiate tissues.
Comparison	Between different tissues or in a single tissue before and after treatment. Intra- or inter-patient.
Outcomes	A positive outcome of a study would be a way to efficiently differentiate between samples using qMRI techniques and individually formed criteria.
Context	Quantitative experiments take place mostly in clinical settings, which makes it difficult to obtain a large number of participants and later access acquired data due to their sensitivity. As a result, there is a risk of sample groups being limited or too small to be significant. Relaxation time analysis is also not a diagnostic standard, so it is most often a subject of academic studies, not clinical studies and trials.

2.4 Search strategy

The search was performed in English, using queries formed and refined to retrieve as many relevant studies as possible. Although it would be possible to include other languages, they might not necessarily be known by readers, and thus it would not be easy to follow and validate such references.

Sensitivity was more critical than specificity, so there was a higher tendency to include irrelevant sources in initial lists than unintentionally discard relevant papers.

The following queries were used in all databases:

1. "(t1 relaxation OR t2 relaxation) AND (malignant OR benign)"
2. "(t1/t2 ratio OR t1 relaxation or t2 relaxation) AND differentiation"
3. "(t1/t2 ratio OR t1 relaxation or t2 relaxation) AND differentiation AND (benign OR malignant)"
4. "t1/t2 AND relaxation AND breast"
5. "breast AND (t1 relaxation OR t2 relaxation) AND differentiation
6. "breast AND qMRI"
7. "breast AND (quantitative MRI)"
8. "magnetic resonance imaging AND breast AND quantitative"

The search was narrowed to sources published in or after the year 2000. Sporadically earlier publications or articles not meeting all of the criteria might be mentioned in the literature review part and results due to their overall value as reviews or novelties, but such cases are clearly stated.

Matches were valid only when found in titles, abstracts, and keywords. Additionally, references used in relevant publications were manually analyzed in search of further related sources, and their relatedness was again assessed based on title, abstract, and, if necessary, full-text analysis, as their numbers were lower while it was more likely that some of them would be relevant.

Queries based on elimination, using "not" and terms referring to modalities and technologies, were discarded due to the unverifiable elimination of seemingly too many results.

The final collection of publications was assembled after performing all queries and eliminating duplicate entries.

2.5 Inclusion and exclusion criteria

Inclusion and exclusion criteria were defined to extract a set of similar qMRI techniques with diagnostic potential.

It was expected that every primary publication provided sufficient information about the hardware and software used (scanner brand and basic parameters like operating magnetic field and type of coil used, image acquisition, and data processing software) as well as applied signal sequences.

T_1 and T_2 constants should be calculated based on a series of qMRI scans performed over time without using contrasting agents, at least in part of a study. Using gadolinium, iron or any other substances would make a potential comparison of study results difficult, as contrasting agents work by altering the relaxation times of tissues. In that case, T_1 and T_2 values and their ratio would be different than during normal measurements.

There was no requirement to report exact formulas used to calculate T_1 and T_2 for as long as possible to trace down software used for calculations and methodology applied if there was any freedom to use software modules. This did not apply to computing environments and programs where users needed to provide their code. In such cases, the formula or model should have been referenced optimally with a fitting algorithm and other operations affecting the data.

Approaches targeted at healthy tissue differentiation were included, as well as differentiation between healthy and benign or malignant changes. Statistical analysis of researched differences was not required but considered a disadvantage if missing. Comparisons of T_1/T_2 acquisition methods with other diagnostic techniques were also accepted if all other criteria were met but were not analyzed due to not being the main topic of the review.

No meta-analysis was attempted, and the purpose of the review was not to grade existing works in any way. Efficiency and appropriateness of described techniques were noted in the form of comments in tables based on the precision of reporting, size of the study group, and results of any statistical analyses if present. Diagnostic aspects of studies were also noted whenever available. Some of them, especially human-dependent factors, might significantly affect obtained results. Table 2 contains information regarding the reasoning behind operations such as ROI selection, which can be directed purely by human judgment or computer-assisted.

3 Search strategy summary

Identification of topic-related publications has proven more complicated than was initially assumed. Although many original articles were identified based on a defined search strategy utilizing titles, abstracts, and keywords, further analysis showed that roughly 20% of initially included articles were relevant to the review's topic. That finding aligns somewhat with the review strategy, putting more significance on sensitivity than specificity.

The most frequent feature leading to the elimination of an article was unspecific vocabulary. Authors frequently used T_1 and T_2 relaxation time-related terms, but their works described only the acquisition of T_1- or T_2-weighted images; another case was low precision of titles and abstracts when it came to describing modalities—multiple original publications were

TABLE 2 Methodology and approaches to analysis in the detailed review studies.

Publication	ROI selection	Conclusions
Breast		
Relaxation times of breast tissue at 1.5T and 3T measured using IDEAL (27)	Menstrual cycle considered: patients scanned no more than 2 days apart each; ROI: average of three points in fat or glandular area. Size varied—drawn to maximize the area of homogeneous tissue.	An increase in a magnetic field leads to relaxation time decrease; IDEAL in the same field increased relaxation times. because it removes the water signal from adipose tissue (leaving only signal from actual fat). Significant differences between fields and tissues ONLY for T1, not significant for T2.
Longitudinal and Multi-Echo Transverse Relaxation Times of Normal Breast Tissue at 3 Tesla (28)	Single 3-mm-thick coronal slice midway between nipple and chest wall. Voxel-wise relaxation maps.	Knowledge of fat and water T1 times allows efficient fat and water signal suppression. The resulting values were 15%-30% higher than in a similar study using 3T (18), but the earlier one used only two-time points for T2. Very few participants (5-6). Better contrast against fat visible for very dense fibroglandular tissue.
T1 and T2 temperature dependence of female human breast adipose tissue at 1.5T: groundwork for monitoring thermal therapies in the breast (29)	Contained 30 voxels (2 × 2 × 10 mm)?, manual placement in adipose regions	Dependence of T1 and T2 on temperature showed little inter-sample variation
Changes of T2 relaxation time from neoadjuvant chemotherapy in breast cancer lesions (30)	Consensus between two radiologists; drawn on T2 images if lesions were visible or on a fusion of T2 and DWI images. Avoiding necrotic and cystic areas. Mean value from three different regions of a lesion.	Significant differences before and after; also significant between responders and non-responders after, but not before
Role of quantitative analysis of T2 relaxation time in differentiating benign from malignant breast lesions (31)	1-9 cm^2, depending on tumor size, based on STIR	Large overlap between T2 range for benign and malignant samples. Significant differences between malignant and benign samples, but not between different types of malignancies; shorter T2 for malignant. Large overlap between groups makes clustering inefficient.
Lung		
T2 mapping of CT remodeling patterns in interstitial lung disease (32)	"as large as possible, not less as 100 mm2, and positioned to avoid major heterogeneities: large blood vessels, main airways, or motion mismatch that could not be corrected", with reference to CT scans	Significant differences in T2 times depending on the amount of fibrous tissue; differentiation between healthy and pathological tissue.Significant differences between normal and pathological tissues; between values of parenchymal features (GGO, RE, HC);also, between left and right lung for GGO and RE
Prostate		
Measurements of T1-relaxation in *ex-vivo* prostate tissue at 132μT (33)	Classified into normal and cancerous based on expert's observation. Mix of qMRI and NMR to obtain better SNR	T1 contrast is increased at very low field (below 1mT), but low SNR is an issue. Large T1 variability between patients, but it should be sufficient to have enough intrapatient contrast to tell cancer apart from normal tissue. Shorter T1 for cancer.
Relationship between T2 relaxation and apparent diffusion coefficient in malignant and non-malignant prostate regions and the effect of peripheral zone fractional volume (34)	Freehand ROI around dominant tumor nodule, with reference to ADC maps	Significant differences in T2 for prostate zones and tumorSignificant differences between tissue areas around tumors, and additionally correlation between T2 values and diffusion coefficients.
Changes in apparent diffusion coefficient and T2 relaxation during radiotherapy for prostate cancer (35)	By a radiologist/based on the decreased intensity of T2 signal, DCE, and consistency with the previous biopsy	In some cases, therapy resulted in significant changes in T2
Rotating frame relaxation imaging of prostate cancer: Repeatability, cancer detection, and Gleason score prediction (36)	"drawn on TRAFF, T1rcw,and T2 images using anatomical T2wi and prostatectomysections as the reference"	T2 can be used as a parameter to discriminate between healthy and malignant prostate tissues.
Skin		
In vivo morphological characterization of skin by MRI micro-imaging methods (37)	"linear regions of interest of 40 × 1 pixels parallel to the surface (6 mm × 50 μm) located in the center of the area imaged to a depth of 2.5mm"	qMRI, especially at high resolution, allows for efficient skin assessment.Mean values for skin layers are provided, but no statistical analysis was performed.
Kidney		
Quantitative versus qualitative methods in evaluation of T2 signal intensity to improve accuracy in diagnosis of pheochromocytoma (38)	Images evaluated by two radiologists; lesions classified as homo-or heterogeneous. ROIs were approaching, but not including borders of lesions.	Lesion intensity was analyzed as relative to CSF and other organs, with no values assumed as arbitrary. An unusual approach among other studies, but allowed to achieve good classification results (adrenal to muscle (81%) and adrenal to liver best) when distinguishing from other adrenal lesions.

(Continued)

TABLE 2 Continued

Publication	ROI selection	Conclusions
		Alternatively, in some cases, lesion discovery failed at all, whereas they were visible in qualitative analysis.
Liver		
Characterization of focal liver lesions using quantitative techniques: comparison of apparent diffusion coefficient values and T2 relaxation times (39)	ROI included the largest possible part of a lesion, avoiding blood vessels, necrosis, artifacts, and partial volume effects. ROC to define ADC and T2 cutoff between benign and malignant lesions.	T2 had much better AUC than ADC for differentiating between benign and malignant lesions.Mean T2 was lower for malignant than benign lesions.AUC for diagnosing malignancy was 0.932 with sensitivity of 99% and specificity of 80.8%. No possibility of differentiating between specific lesion types due to large overlap.
Differentiation of Hepatocellular Carcinoma and Hepatic Metastasis From Cysts and Hemangiomas With Calculated T2 Relaxation Times and the T1/T2 Relaxation Times Ratio (40)	Manual placement of the largest rectangular ROI within a lesion, by a single investigator. If regions of increased T2 (cysts/necrosis) were present, the ROI was placed along the border, without the center.	The T1/T2 ratio allows differentiating between cysts, hemangiomas, and solid lesions. No overlap between ratios for benign and malignant lesions. Almost 100% sensitivity and specificity for classification based on a ratio. Only lesions larger than 1cm in the largest dimension.
Discrimination of benign from malignant hepatic lesions based on their T2-relaxation times calculated from moderately T2-weighted turbo SE sequence (41)	Two measurements for each lesion. Largest possible ROI, excluding cysts and necrotic regions. Also, reference ROIs from normal liver tissue.	Differentiation with the threshold between 67 and 116 ms resulted in a classification of benign and malignant lesions with a sensitivity of 90% and specificity of 94%. Statistically significant differences between malignant and benign lesions. Despite this fact, the authors still recommend using qMRI as support for other techniques, such as DCE-MRI due to T2 range overlap.
Hepatic malignant tumor vs. cavernous hemangioma: differentiation on multiple breath-hold turbo spin echo MRI sequences with different T2-weighting and T2-relaxation time measurements on a single slice multi-echo sequence (42)	Largest lesion of a patient selected for assessment. Section with the largest tumor dimension. An elliptical or circular region with maximal inclusion of tumor but excluding partial volume averaging with the surrounding liver.	Not only T2 values were compared but also changes in signal intensity along the entire timeline, and intensity at certain echo times showed better clustering potential than others. Fat suppression provided better results than when not applied.
Recurrent hepatocellular carcinoma versus radiation-induced hepatic injury: differential diagnosis with MR imaging (43)	ROIs included parenchyma, irradiated areas, and HCCs.	Patterns of hypo- and hyper-intensity are different for studies with and without contrast. To make HCCs stand out in the irradiated area (after radiation therapy), the use of contrast was beneficial (no significant differences in intensity without it).
Differentiation of focal hepatic lesions in MR imaging with the use of combined quantitative and qualitative analysis (44)	For each of two echoes two measurements were taken and averaged. ROI covering as much tumor as possible, limited to solid parts of tissue	Qualitative analysis was used to differentiate between solid tumor types by the agreement of two radiologists.Significant differences between solid tumors and other lesion types – optimal threshold of 116 ms (96% sensitivity, 93% spec.)

discarded due to using nuclear magnetic resonance (NMR) or magnetic resonance spectroscopy (MRS) instead of qMRI.

The review was primarily targeted toward breast and epithelial tissue, so specific searches were done to investigate these matters in detail. It seems possible that specific searches toward other organs and tissues would result in an even more extensive collection of original works of interest. Also, many additional results were retrieved when using the last query ("quantitative magnetic resonance imaging") instead of the abbreviation (qMRI). It could suggest that using the full names of technologies in queries might be a good practice. Unfortunately, in this case, it resulted in 228 positions either not being related to relaxometry at all (interpreting only "magnetic resonance imaging" probably without a match for "quantitative") or involving the use of contrast agents (about 42% of 228). The rest of the publications did not meet the other requirements, leaving only two papers seemingly relevant: one about breast cancer-related lymphedema (BCRL) and the other one (discarded) based on human cell culture and rat xenografts.

The Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) diagram (45) was included to provide an overview of the search process and is presented in Figure 2. The diagram shows results obtained after using the initial set of queries listed in point 2.4 and results from the query "quantitative magnetic resonance," which was added later. No automation tools were used to perform the identification or screening of retrieved publications.

4 Literature review

4.1 Reviews and systematic reviews so far

Five reviews that meet the search criteria have been identified (46–50). The first one summarizes current techniques of T_1 and T_2 time calculation at a magnetic field of 3T. It also reports T_1 and T_2 values for different tissue types—white/gray matter, CSF, muscle, myocardium, fat, and others. The authors describe multiple aspects to consider when

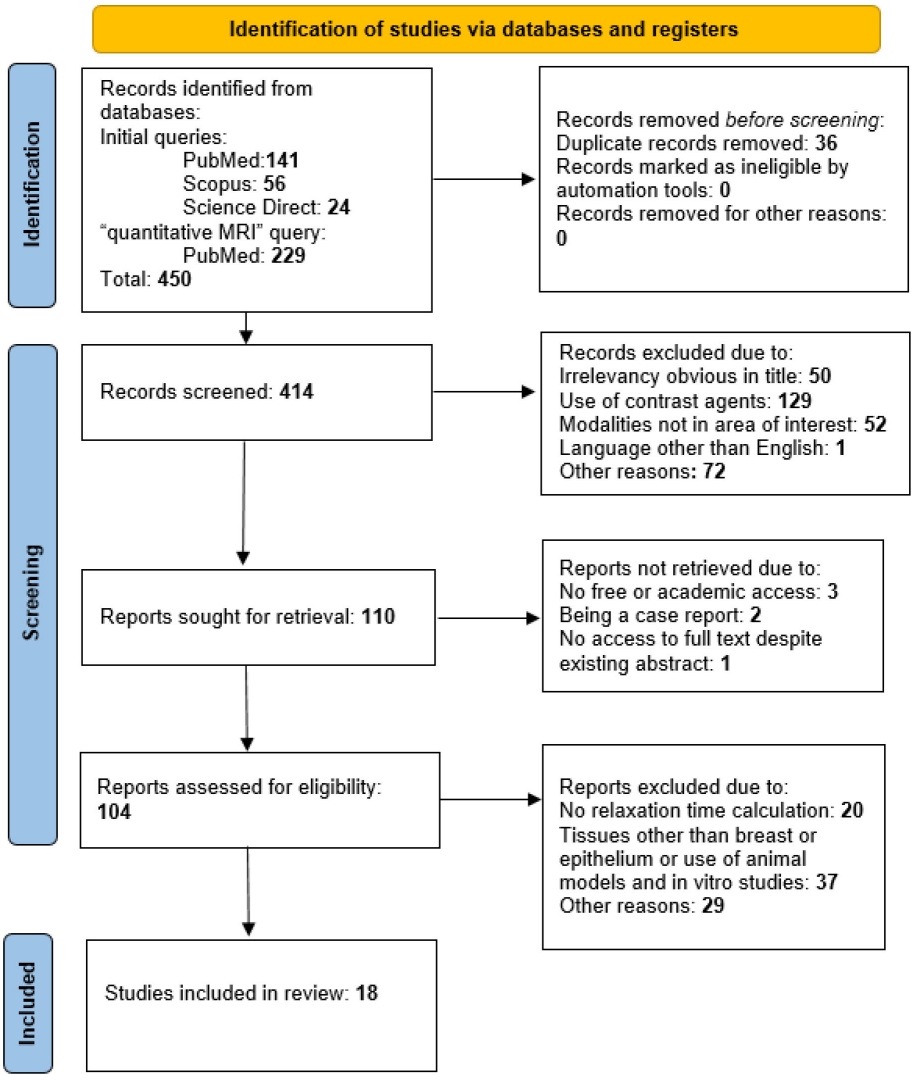


FIGURE 2
The review process is summarized on a PRISMA diagram.

calculating relaxation constants, such as signal noise, partial volume effects, or magnetic field inhomogeneity as sources of large differences between reported T_1 and T_2 values for different tissues.

Another review has been mentioned as a reference in (46), related to methods of calculation of T_1 (51). It is not related to tissue classification and is dated outside the chosen time range (2000–now), but it provides a complex overview of the literature describing approaches up until 1999. It might be of interest to those investigating differences between T_1 times acquired with different techniques.

Wolf et al. (47) presented a review of relaxation time analysis in kidneys, reporting multiple studies and their results regarding both longitudinal and transverse relaxation. Basic qMRI principles and sequences used for renal imaging are mentioned. Multiple factors affect possible results, such as fasting, hydration or oxygenation level, modulating diuresis speed, and fluid retention in necrotic tissue. Such factors are highly specific and do not apply to the imaging of most other organs. In general, the authors conclude the potential usefulness of renal qMRI, although it is strongly susceptible to physiological and pathological alterations, which should always be accounted for.

For cartilage degradation studies, a review has been published (50) that investigates the impact of preexamination activity and exercise on the results of T2 measurement. The review shows that activity might cause water particle movement in cartilage, which affects its relaxatory parameters. Thus, pre-scan procedures should be implemented to reduce patient

movement, as it could be a source of variability in T2 values obtained by different teams.

The last review, published in two parts, provides complex information about current cardiac imaging standards concerning T1 estimation. The authors summarize methodological aspects of myocardial T1 and ECV, such as sequence choice (with MOLLI or SASHA being the most popular ones) or motion correction. It is emphasized in the paper that precise comparisons between studies are possible when the same protocols are being used to obtain results. Due to that fact, T1 analysis has a significant supportive value in cardiac studies, but it cannot be used as a standalone diagnostic parameter.

4.2 Original publications

The number of relevant sources retrieved was not high but sufficient to draw basic conclusions on the use of relaxation times in diagnostics and tissue differentiation. Based on the material, it can be seen that relaxation analysis is practiced more often on some tissues than others. Most of the 59 publications (25%) were dedicated to the nervous system, particularly the brain. qMRI was also frequently used when studying connective tissues, such as cartilage (17%) or muscle tissue, including myocardium (22%). Another 10% of studies treated the liver. Studies dedicated to quantification of other tissues and organs (breasts, prostate, kidneys, and others) were encountered sporadically when using queries 1–3 (see *Search strategy*), as shown in Figure 2. Queries formulated specifically to retrieve breast-related studies allowed us to identify a higher number of valuable articles (Figure 3).

The structure and data provided by authors of publications were strongly diverse, which suggests no "golden rule" for conducting this kind of study. Tables 3, 4 summarize experimental setups from articles chosen for the detailed review.

4.3 Information required in the review

The content required by review rules was generally provided, which means reporting the mean or median T_1/T_2 value with standard deviation. A single exception was a study reporting only relative signal intensity between adrenal lesions and abdominal organs (38), but it provided interesting insights anyway. Most of the studies utilized magnetic fields of strength of 1.5 or 3 T. Other field strengths were rather uncommon—single cases of 2 T (37) and 4.7 T (52). The MRI scanner model was always reported, in most cases together with coil type and software used for later data analysis.

4.4 Additional information

Age—at least mean or median—of participants was always stated, as well as sex (for mixed groups). Only in a single study (31), the authors tried to classify analyzed tissues and reported classification parameters. Otherwise, when investigating differences between tissues [all except (37)], a test of significance was performed, with p-values reported afterward.

The publications chosen for the detailed review are listed in an attachment (*ReviewResults.ods)*.

Due to anatomical and physio-chemical differences between human tissues, all original publications included in the review were grouped by tissue type or organ type. Because of special interest in them, the breast and selected other tissues were summarized separately from other tissues. Because of special conditions present in cell cultures and their lack of resemblance to the real tissue environment, any studies using cell cultures had to be discarded, even when they used breast cancer cells (52).

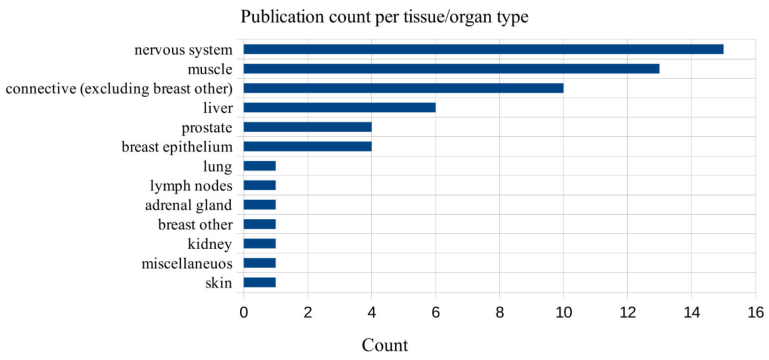


FIGURE 3
The number of studies in a detailed review, grouped by organs and tissues, total = 59.

TABLE 3 Technical aspects of publications chosen for the detailed review.

Publication	Year	Field	Scanner	Sequences			Scanner or data analysis software
				T_1	T_2	Other	
Breast							
Relaxation times of breast tissue at 1.5T and 3T measured using IDEAL (27)	2006	1.53	Echospeed whole-body magnet, GE Healthcare, Waukesha, WI, USA)	FSE-IR	Two Hahn echo scans	IDEAL	Matlab (The MathWorks, Natick, MA)
Longitudinal and Multi-Echo Transverse Relaxation Times of Normal Breast Tissue at 3 Tesla (28)	2010	3	Philips Intera 3T (Philips Healthcare, Best, the Netherlands)	Inversion recovery-prepared multi-shot spin-echo EPI	Spin echo		Matlab (The MathWorks, Natick, MA)
T1 and T2 temperature dependence of female human breast adipose tissue at 1.5T: groundwork for monitoring thermal therapies in the breast (29)	2015	1.5	Achieva (Philips Healthcare, Best, the Netherlands)	Two-dimensional inversion recovery TSE scans with SPIR-WS	Multi-spin echo	T_{2TSE}	
Changes of T2 relaxation time from neoadjuvant chemotherapy in breast cancer lesions (30)	2016	1.5	GE Signa, Milwaukee, WI, USA	FSE, STIR	FSE	T1W, T2W; 3D after contrast	SPSS software, ver. 11.5 (SPSS Inc., Chicago, Il, USA), Functool T2 mapping software
Role of quantitative analysis of T2 relaxation time in differentiating benign from malignant breast lesions (31)	2018	1.5	SIGNA™ Infinity; GE Medical Systems	Axial FSE T1-weighted	Sagittal fat-suppressed T2-WI	Axial DWI, axial short-time inversion recovery (STIR)	Functool (Advantage Workstation 4.3 (GE Medical Systems)); SPSS Inc., Chicago, IL, USA
Lung							
T2 mapping of CT remodeling patterns in interstitial lung disease (32)	2015	1.5	Magnetom Aera, Siemens Medical Systems		Multi-echo single-shot turbo spin echo sequence	CT	MRmap for IDL8.3 Software,R 2.15.1 (R Foundation for Statistical Computing, Vienna)
Prostate							
Measurements of T1-relaxation in *ex-vivo* prostate tissue at 132μT (33)	2012	132μT	Techmag Orion system	SQUID			No data/exponential fit for T1
Relationship between T2 relaxation and apparent diffusion coefficient in malignant and non-malignant prostate regions and the effect of peripheral zone fractional volume (34)	2013	3	Achieva (Philips Medical Systems, Best, Netherlands)		FSE	T1W, T2W	SPSS® v. 19 for Windows (IBM Corporation, Portsmouth, UK)
Changes in apparent diffusion coefficient and T2 relaxation during radiotherapy for prostate cancer (35)	2013	1.5	Signa, General Electric Medical Systems, Waukesha, WI		Magnetization-prepared spiral imaging pulse sequence	2-weighted fast-spin-echo (FSE), DWI	MIPAV (National Institutes of Health, Bethesda, MD), Origin software (OriginLab, Northampton, MA)
Rotating frame relaxation imaging of prostate cancer: Repeatability, cancer detection, and Gleason score prediction (36)	2016	3	Ingenuity PET/MR, Philips, Cleveland, OH		GraSE	T2W, T1rho	GraphPad Prism, version 5.00 (GraphPad Software, San Diego, CA), MATLAB (MathWorks Inc., Natick, MA)
Skin							

(Continued)

TABLE 3 Continued

Publication	Year	Field	Scanner	Sequences			Scanner or data analysis software
				T_1	T_2	Other	
In vivo morphological characterization of skin by MRI micro-imaging methods (37)	2004	2	Magnex Scientific Ltd., Oxford, England	GE		MT, T1W, T2W	Customized imaging console (SMIS Ltd.)
Kidney							
Quantitative versus qualitative methods in evaluation of T2 signal intensity to improve accuracy in diagnosis of pheochromocytoma (38)	2015	1.5	Signa, GE Healthcare		Breath-hold SSFSE or respiratory-triggered frequency-selective fat-suppressed fast recovery FSE	RARE T2W, chemical shift imaging	SPSS (SPSS Inc., Chicago, IL, USA), Matlab (version 2014b, MathWorks).
Liver							
Characterization of focal liver lesions using quantitative techniques: comparison of apparent diffusion coefficient values and T2 relaxation times (39)	2012	1.5	Magnetom Avanto, Siemens Medical Solutions, Erlangen, Germany)		Breath-hold dual echo T2W TSE (with AND without contrast)	DW-SS-EPI, 3D DCE-MRI	ADC: Leonardo Workstation, Siemens Medical Solutions, Erlangen, Germany), Statistica 10.0, Microsoft Excel
Differentiation of Hepatocellular Carcinoma and Hepatic Metastasis From Cysts and Hemangiomas With Calculated T2 Relaxation Times and the T1/T2 Relaxation Times Ratio (40)	2006	1.5	Philips Intera, Philips Medical Systems of North America, Andover, MA, USA)	Mixed-TSE	Mixed-TSE		DICOM processing on PC using MathCAD 2001i (MathSoft, Cambridge, MA, USA), SAS 8.02 (SAS Institute, Cary, NC, USA)
Discrimination of benign from malignant hepatic lesions based on their T2-relaxation times calculated from moderately T2-weighted turbo SE sequence (41)	2002	1.5	Gyroscan ACS NT, Philips, Eindhoven, The Netherlands		Double echo TSE		
Hepatic malignant tumor vs. cavernous haemangioma: differentiation on multiple breath-hold turbo spin echo MRI sequences with different T2-weighting and T2-relaxation time measurements on a single slice multi-echo sequence (42)	2002	1.5	Gyroscan ACS NT, Philips Medical Systems, Best, The Netherlands		8-echo TSE	With or without fat suppression	
Recurrent hepatocellular carcinoma versus radiation-induced hepatic injury: differential diagnosis with MR imaging (43)	2001	1.5	Signa Advantage (GE Medical Systems, Milwaukee, WI, USA)	Spin-echo	Spin-echo	DCE-MRI, T1W, T2W	T2 fitting with scanner built-in software
Differentiation of focal hepatic lesions in MR imaging with the use of combined quantitative and qualitative analysis (44)	2007	1.5	Gyroscan ACS NT, Philips	T1 FSMPGRE	TSE, dual-echo TSE, T2 STIR	Dynamic FSMPGRE	Microsoft Excel, Statistica 6.0 (StatSoft Poland)

4.5 Methodology

Multiple factors can affect the results of relaxation time analysis. Some of them result from equipment choice and experimental conditions, such as strength of magnetic field used or type of coil. There are also other nuances present, important from a diagnostic point of view.

Two scenarios were the most common among analyzed studies: differentiating between tissue types (normal/pathological or organ parts, e.g., white and gray matter) or comparing results before and after therapy. To do so, different approaches were applied, some of them strongly susceptible to human error. As seen in Table 2, often actions were taken to eliminate risk of a radiologist mistakenly selecting an incorrect region for analysis. In most cases, another type of MRI image or modality (CT) was used to identify the region of interest, which then underwent analysis and could potentially be refined based on the results.

Only sporadically, authors mentioned the stage of the menstrual cycle in the case of breast studies, which might affect T_1 or T_2 values obtained due to periodic variations in tissue structure in this area. Such processes might affect not only normal experiments but also those using contrasting agents (53) (Table 5).

10.3389/fonc.2022.1010643

TABLE 4 Biomedical aspects of publications chosen for the detailed review.

Publication	Tissue/ organ	Subject	Environment	Reported T2 values [ms]	Reported T1 values [ms]	Participants: number/age [years]
Breast						
(27)	Breast adipose and fibroglandular tissue	Normal tissue characterization	*In vivo*	Fat 1.5T IDEAL	296.01 ± 12.94	5/36.4 ± 12.6
				Non-IDEAL 53.33	372.04 ± 8.6	
				3T IDEAL	366.78 ± 7.75	
				Non-IDEAL 52.96	449.27 ± 26.09	
				Glandular 1.5T IDEAL	1,266.18 ± 81.8	
				Non-IDEAL 57.51	1,135.98 ± 151.37	
				3T IDEAL	1,444.8 ± 92.7	
				Non-IDEAL 54.36	1,324.42 ± 167.63	
(28)	Adipose and fibroglandular tissue	Normal tissue characterization	*In vivo*	154 ± 9 adipose,71 ± 6 fibroglandular	423 ± 12 adipose,1,680 ± 180 fibroglandular	6/34 ± 6, 38-49
(29)	Connective (adipose) tissue in breast	Dependence of adipose tissue T2 on temperature	*ex vivo*	55 to 100 for temp. 25°C–65°C	200–550 ms for temp. 25-65	21-56, mean 30
(30)	Breast	Lesion response to neoadjuvant therapy measured with T2 constant	*In vivo*	81.34 ± 13.68 ms pretreatment, 64.50 ± 8.71 ms posttreatment		34-70, mean 55.2
(31)	Breast	benign and malignant changes in breasts, mostly invasive ductal carcinomas	*In vivo*	Benign:fibrodenomas: 92.53 ± 22.76papillomas:84.36 ± 14.69fibrocystic adenosis:103.56 ± 4.17Malignant:inv. ductal carc.: 80.64 ± 10.16inv. Lobular carc.: 76.87 ± 14.01ductal carc. *in situ*: 82.29 ± 12.51		67/mean 50.7 ± 17.3,(26-74)
Lung						
(32)	Epithelium—lung	Differentiation and characterization of lung tissues in pneumonia	*In vivo*	Median 41, 38-49		Six men and six women, with a mean age of 62 (47-81)
Prostate						
(33)	Prostate	Contrast between normal and cancerous prostate tissue	*Ex vivo*		41–86, according to authors shorter for cancer, but randomly checked samples often show opposite results	No data, 35 *ex vivo* specimens
(34)	Prostate	Malignant vs. non-malignant prostate regions	*In vivo*	PZ: 149 ± 49 ms, TZ: 125 ± 26 ms, tumor: 97 ± 23 ms PZ = peripheral zone, TZ = transition zone		48–83 years (mean age 67 ± 8 years)
(35)	Prostate	Prostate cancer response to radiotherapy—before and after	*In vivo*	Multiple values for different features reported; approx. 70–114, generally around 80–90		Median age 66; min: 54; max: 77
(36)	Prostate	Prostate cancer detection and differentiation	*In vivo*	79 ± 9 (cancer), 124 ± 38 (peripheral zone), 87 ± 7 (central gland)		63 ± 6 (42-73)
Skin						

(Continued)

TABLE 4 Continued

Publication	Tissue/ organ	Subject	Environment	Reported T2 values [ms]	Reported T1 values [ms]	Participants: number/age [years]
(37)	Epithelium—skin	Identification and characterization of skin layers	*In vivo*/ phantoms		Stratum corneum: = 135 ± 10, epidermis = 347 ± 27, papillary dermis = 356 ± 22, reticular dermis = 290 ± 10 and other	Seven normal subjects, five women, two men with mean SD of 32 ± 6 years
Kidney						
(38)	Adrenal gland	Differentiation between pheochromocytoma and other malignant and benign changes in an adrenal gland	*In vivo*	Only stated as relative to cerebrospinal fluid intensity (CSF) and abdominal organs		74, 39 women, 35 men
Liver						
(44)	Liver	Classification of benign and malignant liver lesions, including carcinomas	*In vivo*	Benign: hemangioma: 124.3Cyst: 1007FNH: 62.8Abscess: 406.8Cystadenoma: 459Malignant: metastasis: 65.3 **HCC (carcinoma): 59.1**Hemangioendothelioma: 64.9Cholangiocarcinoma: 55.7Cystadenocarcinoma: 117.5		73 (34 men, 39 women)/54.2 (18-84)
(40)	Liver	Differentiation between different types of lesions with T2 and T1/T2 ratio	*In vivo*, phantoms	Cysts: 371 ± 118Hemangiomas: 204 ± 70 **HCCs/metastases: 83 ± 17**	Cysts: 691 ± 215Hemangiomas: 653 ± 152 **HCCs/metastases: 609 ± 133**	36 (22 men, 14 women)/60 (30-86)
(41)	Liver	Differentiation of liver lesions with T2	*In vivo*	Liver: 54 ± 8FNH: 66 ± 7Malignant lesions: 85 ± 17Hemangiomas: 155 ± 35Cysts: 583 ± 369		52 (28M, 24F) with 114 lesions
(42)	Liver	Differentiation of malignant and benign liver lesions	*In vivo*	Malignant: 40–102, mean 73Hemangiomas: 68–233, mean 165		34 (26M, 8F)/55 (38-70)
(43)	Liver	Differentiation between HCCs and radiation effects	*In vivo*	Liver parenchyma: 42 ± 1.60Irradiated area: 56.4 ± 3.19 **Hepatocellular carcinoma: 58.7 ± 3.19**		X/65
(44)	Liver	Differentiation of lesions with combined quantitative and qualitative analysis	*In vivo*	Liver: 53 (41-74)Solid tumors: 84.1 (54-148)Other lesions: 250.5 (82-1241) **HCC: 75.3 (56-91)**		292 lesions in 168 patients (80M, 88F)/53 (17-83)

TABLE 5 Relaxation times reported for epithelial tissue.

	Relaxation time [ms]							
	Breast						Lung	
	Identification (depending on the type of pathological changes)		Invasive ductal and lobular carcinomas				Normal	Pathological (pneumonia)
			NAC responders		NAC non-responders			
	Benign	Malignant	Pathological before treatment	Pathological after treatment	Pathological before treatment	Pathological after treatment		
T2	Approx. 84-103	Approx. 77-82	80.93 ± 14.4	63.18 ± 8.37	84.57 ± 6.06	74.62 ± 2.32	41 (28-49) (median)	66.5/74.3/79.5 (median)

TABLE 6 T2 values reported by prostate studies.

		Relaxation time [ms]			
		Study A—1.5 T, exterior torso coil		Study B—3 T, endorectal coil	Study C—3T, cardiac coil
		Before therapy	Week 8	Identification of regions	Identification of regions
T2	Prostate	86 ± 10	78 ± 4	–	–
	Central gland	78 ± 8	76 ± 5	–	87 ± 7
	Peripheral zone	114 ± 27	89 ± 13	149 ± 49 (82-290)	124 ± 38
	Tumor	82 ± 15	75 ± 9	97 ± 23 (62-177)	79 ± 9
	Transition zone	–	–	125 ± 26 (84-186)	–

4.5.1 Tissues other than epithelium

A basic review (consisting of abstract, methodology, and conclusion scan) of relevant literature showed a variety of applications of relaxation time analysis. The most commonly studied subjects were the nervous system, especially the brain, and connective tissues such as cartilage (Table 6).

Techniques used for quantitative imaging differed strongly depending on the body area. Some studies implemented relatively simple protocols, using general-purpose whole-body coils (54), whereas others used complex and tailored approaches to cardiac (55–57) or vessel imaging (58). Moderate magnetic fields dominate the literature, such as 1.5 or 3 T, but attempts were made to visualize human tissue at ultra-low or ultra-high fields, e.g., imaging of the hippocampus at 7 T (59) or of the meniscus at 9.4 T (60). Similarly, in the case of gray matter (61) or prostate study mentioned further (33), this time using very low fields, field manipulation is used to achieve better contrast between features of interest, which might be more visible in specific conditions.

A multiplicity of qMRI applications exists, ranging from the analysis of wear and tear of cartilage with age (60) to assessment of pathological changes responsible for conditions such as Alzheimer's or Parkinson's diseases (62). Relaxation time analysis is often applied to visualize abnormalities present in the brain, and instead of being used as a direct measure of tissue state (healthy/abnormal), imaging is used to calculate organ part volumes, which in neurology or cardiology are considered important indicators of health. Sometimes different sequences showed different efficiencies depending on the organ part imaged, as in the case of one brain study (63).

Not only T1 or T2 analysis was often performed, but correlations were searched for between them and other parameters, such as ADC (64); T2*, T1rho, T1, and T2 obtained after contrast medium application (non-native relaxation times), diffusion tensor or magnetization transfer imaging (65); or optical coherence tomography results (66). Such measurements complemented the diagnosis made by histological or macroscopic sample assessment.

4.5.2 Epithelium, especially in the breast

Out of 59 accepted original publications, 18 were related to epithelial tissue, including two describing the breast. Apart from those, one publication described breast adipose tissue properties. Because of only these few positions, a decision was made to include other tissues in the comparison, such as the prostate and lung. Although lung parenchyma does not seem to be very similar to breast epithelial regions, there might exist scientific background suggesting genetic likeness between them (67–69).

Studies of epithelial lung and breast tissue did not seem comparable. The same magnetic fields were used, but experimental setups and results differed significantly. As seen in Table 7, T_2 ranges of epithelium do not overlap for breast and lung. On the other hand, means of T_2 registered for breast epithelium show significant differences (according to corresponding publications), but actual results still overlap, even when only standard deviations are accounted for. That makes any clustering attempts difficult.

TABLE 7 Review inclusion and exclusion criteria for original studies.

Inclusion criteria	qMRI
	Human tissues *in vivo*/*ex vivo*
	T_1 and/or T_2 time calculation
	Full text available
	Original publications since the year 2000
	Final publication stage
	Languages: English
Exclusion criteria	Animal tissues or phantoms, cell cultures
	Use of contrasting agents
	No free or academic access
	MRS, NMR, CT, etc.
	Only parameters other than T_1/T_2 (T_2*, T_1rho, etc.)
	Only T_1/T_2-weighted images
	Case reports, case series

The structural and pathogenic likeness is better established between breast and prostate cancer, which makes them often studied together (70–73). Similar treatment approaches may be used for both (74).

All studies regarding prostate reported the significance of differences between T2 values of at least some analyzed areas, as shown in Table 4. There was no straight correlation between magnetic field strength and measured relaxation times, but their values suggest more similarity to breast tissue than the lung epithelium showed.

There was no consensus on experimental design in this group of studies, but instead, different types of coils were used depending on external circumstances or other choices.

One novel (for breast) publication was found (75), regarding the calculation of T1/T2 ratios in breast tissues. Unfortunately, it seems that a contrasting agent was used during the scan. It suggests that T1/T2 ratios might change proportionally and be significantly correlated with pathological breast cancer stage. The majority of studies, reports, and analyses were based only on one of two constants, which is shown in Table 7. A similar approach, based on T1/T2 ratios, was earlier used in 2006 regarding liver lesions (40), and this one was included in the detailed review as the authors did not use contrasting agents.

The study regarding breast adipose tissue was included in this part because, while it might not be an area susceptible to typical breast carcinomas, it might have the potential to affect relaxation measurements due to its abundant presence. Additionally, there was a chance that a significant difference in relaxation times between epithelium and fat would occur, which could help during ROI selection and tissue differentiation. A series of T1 and T2 values was reported for temperatures ranging from 25°C to 65°C with approximate T2 values for body temperature (37°C) between 65 and 70 ms. The T1 range was between 270 and 320 ms for 37°C. Unfortunately, there were no T1 values reported for breast to compare them with.

In addition to prostate and lung epithelium, other organs were included in this part if their corresponding carcinomas were studied, as they originate from epithelial tissue. Experiments regarding healthy kidney or liver tissue were considered not related to the main topic.

Publication by Cieszanowski et al. (41) provides an exhaustive insight into techniques used for quantitative imaging of liver lesions.

One study (42) presented a slightly different approach than others because not only T2 values were compared but also changes in signal intensity along the entire timeline. Intensity at certain echo times showed better clustering potential than others, as well as relative intensity change compared with the initial tissue state.

5 Conclusions

Reporting only one of the constants might be caused by limited resources or assessment of other parameters than the most known ones. There might be some beliefs present, suggesting better efficiency of one or another parameter when differentiating tissues, as considering the number of available publications, it cannot be at this point verified whether any of them is better than the other. In such a situation, it could be profitable to search for correlations between different tissue behaviors instead of sticking to one. It is also possible that many teams tried to analyze both T1 and T2 data, but negative results could have lower chances of getting published than positive findings, as they do not seem as important or revolutionary. Disappointing or inconclusive results were rarely reported (5 out of 59), suggesting areas in which relaxation analysis is, at least currently, not efficient (58, 76–78). Otherwise, multiple reports of high clustering efficiency have been reported.

In cases when the menstrual cycle was not considered or otherwise noted, it could be a source of high intrapatient variation in breast cancer and healthy tissue studies.

Only two studies reported the use of T1/T2 ratios, one examining the liver (without contrasting agents) and the other breast (with contrast), but presented classification results suggest that such an approach might be worth investigating. It is not common nowadays, especially since most researchers perform only one kind of analysis; it is T1 or T2 only, sometimes supported by other sequences or modalities (ADC, CT, T1W, T2W). Classification sensitivity and specificity were much higher for the T1/T2 ratio (close to 100%) than for any of these parameters used separately (~80%).

According to an analyzed study of liver lesions (39), relaxation time analysis has high potential in differentiating between benign and malignant lesions, for example telling apart carcinomas from cysts and hemangiomas. For a more detailed analysis, especially of solid tumors, additional means are necessary, such as dynamic studies with contrasting agents or confirmation by biopsy. qMRI could be thus used just as one of the steps on the way to presurgical diagnosis and treatment choice.

While there was only a single study that mentioned how a patient's activity before MR scan affects obtained T2 values (50), it should be noted that preexamination procedures might affect studies of different organs depending on mechanical load, oxygen exchange, or blood flow. Additionally, the treatment of *ex vivo* samples could also leave its mark on the samples studied. Such samples also tend to be studied at lower temperatures than body temperature, so correction for that factor could be necessary when attempting comparisons with *in vivo* studies.

Based on analyzed studies, the choice of protocols, technologies, and tissue parameters to assess should be made based on previous publications, which might suggest which approaches have proved successful and what did not work. There are no universal rules for qMRI as too many factors need to be considered when designing an experiment, and thus authors need to try and make informed choices regarding qMRI of any tissue or organ. Some "good practices" are recognized in certain applications, especially for brain and cardiac imaging, so they should be researched beforehand.

Data availability statement

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

Author contributions

Conceptualization, MaM; DA; JS, DB-A; MiM; methodology, MaM; DA; JS, D-BA; MiM; validation, MaM; DA; JS, D-BA; MiM; formal analysis, MaM; DA; JS, D-BA; MiM; resources, MaM; DA; JS, D-BA; MiM; writing-original draft preparation, MaM; DA; JS, D-BA; MiM; writing—review and editing, MaM; DA; JS, D-BA; MiM; supervision, MaM; DA; JS, D-BA; MiM. All authors contributed to the article and approved the submitted version.

Funding

The study was co-financed by the European Union from the European Regional Development Fund under RPO (Regional Operational Programme) for Podkarpackie Region for 2014-2020 "Competitive and Innovative Economy" RPPK.01.02.00-18-0012/18-00 "R+D works on developing expert system supporting data analysis obtained from breast cancer tissue using MRI".

Conflict of interest

Authors MaM, JS and MiM were employed by SoftSystem Sp. z o.o.

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

References

1. Sardanelli F, Boetes C, Borisch B, Decker T, Federico M, Gilbert FJ, et al. Magnetic resonance imaging of the breast: Recommendations from the EUSOMA working group. *Eur J Cancer* (2010) 46(8):1296–316. doi: 10.1016/j.ejca.2010.02.015

2. The American Society of Breast Surgeons. *Official statements* (2017). Available at: https://www.breastsurgeons.org/docs/statements/Consensus-Guideline-on-Diagnostic-and-Screening-Magnetic-Resonance-Imaging-of-the-Breast.pdf (Accessed 20 March 2020).

3. European Commission. *European Guidelines on breast cancer screening and diagnosis* . Available at: https://healthcare-quality.jrc.ec.europa.eu/ecibc/european-breast-cancer-guidelines (Accessed 19 March 2020).

4. European Commission. *Tailored screening in women with high breast density: Magnetic resonance imaging* . Available at: https://healthcare-quality.jrc.ec.europa.eu/european-breast-cancer-guidelines/dense-breast/MRI (Accessed 19 March 2020).

5. Mann RM, Athanasiou A, Baltzer PA, Camps-Herrero J, Clauser P, Fallenberg EM, et al. Breast cancer screening in women with extremely dense breasts recommendations of the European society of breast imaging (EUSOBI). *Eur Radiol* (2022) 32(6):4036–45. doi: 10.1007/s00330-022-08617-6

6. Partridge SC, Nissan N, Rahbar H, Kitsch AE, Sigmund EE. Diffusion-weighted breast MRI: Clinical applications and emerging techniques. *J magnetic resonance imaging: JMRI* (2017) 45(2):337–55. doi: 10.1002/jmri.25479

7. Chen X, Li WL, Zhang YL, Wu Q, Guo YM, Bai ZL. Meta-analysis of quantitative diffusion-weighted MR imaging in the differential diagnosis of breast lesions. *BMC Cancer* (2010) 10:693. doi: 10.1186/1471-2407-10-693

8. Zhang L, Tang M, Min Z, Lu J, Lei X, Zhang X. Accuracy of combined dynamic contrastenhanced magnetic resonance imaging and diffusion-weighted imaging for breast cancer detection. *Acta Radiol* (2015) 57(6):651–60.doi: 10.1177/0284185115597265

9. Choi SY, Chang Y-W, Park HJ, Kim HJ, Hong SS, Seo DY. Correlation of the apparent diffusion coefficiency values on diffusion-weighted imaging with prognostic factors for breast cancer. *Br J Radiol* (2012) 45(2):337–355. doi: 10.1259/bjr/79381464

10. Constantini M, Belli P, Rinaldi P, Bufi E, Giardina G, Franceschini G, et al. Diffusion-weighted imaging in breast cancer: relationship between apparent diffusion coefficient and tumour aggressiveness. *Clin Radiol* (2010) 1005-1012:1005–12.doi: 10.1016/j.crad.2010.07.008

11. Rahbar H, Partridge SC, Eby PR, Demartini WB, Gutierrez RL, Peacock S, et al. Characterization of ductal carcinoma *in situ* on diffusion weighted breast MRI. *Eur Radiol* (2011) 21(9):2011–9.doi: 10.1007/s00330-011-2140-4

12. Galbán CJ, Ma B, Malyarenko D, Pickles MD, Heist K, Henry NL, et al. Multi-site clinical evaluation of DW-MRI as a treatment response metric for breast cancer patients undergoing neoadjuvant chemotherapy. *PLoS One* (2015) 10(3):e0122151. doi: 10.1371/journal.pone.0122151

13. Yabuuchi H, Matsuo Y, Sunami S, Kamitani T, Kawanami S, Setoguchi T, et al. Detection of non-palpable breast cancer in asymptomatic women by using unenhanced diffusion-weighted and T2-weighted MR imaging: comparison with mammography and dynamic contrast-enhanced MR imaging. *Eur Radiol* (2011) 21(9):11–7.doi: 10.1007/s00330-010-1890-8

14. Amornsiripanitch N, Bickelhaupt S, Shin HJ, Dang M, Rahbar H, Partridge SC. Diffusion-weighted MRI for unenhanced breast cancer screening. *Radiology* (2019) 293(3):504–20. doi: 10.1148/radiol.2019182789.

15. Partridge SC, Demartini WB, Kurland BF, Eby PR, White SW, Lehman CD. Differential diagnosis of mammographically and clinically occult breast lesions on diffusion-weighted MRI. *J magnetic resonance imaging: JRMI* (2010) 31(3):562–70.doi: 10.1002/jmri.22078

16. Pinker K, Moy L, Sutton EJ, Mann RM, Weber M, Thakur SB, et al. Diffusion-weighted imaging with apparent diffusion coefficient mapping for breast cancer detection as a stand-alone parameter: Comparison with dynamic contrast-enhanced and multiparametric magnetic resonance imaging. *Invest Radiol* (2018) 53(10):587–95.

17. Kazama T, Kuroki Y, Kikuchi M, Sato Y, Nagashima T, Miyazawa Y, et al. Diffusion-weighted MRI as an adjunct to mammography in women under 50 years of age: an initial study. *J magnetic resonance imaging: JMRI* (2012) 36(1):139–44.

18. Pinker K, Helbich TH, Morris EA. The potential of multiparametric MRI of the breast. *Br J Radiol* (2017) 90(1069):20160715.

19. Daar D, Bernardo M, Choyke PL, McKinney Y, Turkbey B. Prostate multiparameter MR imaging. *Radiol Technol* (2011) 1:563–8.

20. Ferda J, Ferdová E, Hes O, Mraček J, Kreuzberg B, Baxa J. PET/MRI: Multiparametric imaging of brain tumors. *Eur J Radiol* (2017) 94:A14–A25.

21. Abramson RG, Arlinghaus L, Dula A, Quarles CC, Stokes A, Weis J, et al. MRI Biomarkers in oncology clinical trials. *Magnetic resonance Imaging Clinics North America* (2016) 78–88.

22. Cumpston M, Li T, Page MJ, Chandler J, Welch VA, Higgins JP, et al. Updated guidance for trusted systematic reviews: a new edition of the cochrane handbook for systematic reviews of interventions. *Cochrane Database Systematic Rev* (2019) 10:ED000142.

23. BioMed Central. *Systematic reviews*. Available at: https://systematicreviewsjournal.biomedcentral.com/submission-guidelines/preparing-your-manuscript/methodology (Accessed 09 March 2020).

24. National Center for Biotechnology Information. *PubMed* . Available at: https://pubmed.ncbi.nlm.nih.gov/about/ (Accessed 05 March 2020).

25. Elsevier. *Scopus* . Available at: https://www.scopus.com/home.uri (Accessed 06 March 2020).

26. Elsevier. *Science direct* . Available at: https://www.sciencedirect.com/ (Accessed 06 March 2020).

27. Rakow-Penner R, Daniel B, Yu H, Sawyer-Glover A, Glover GH. Relaxation times of breast tissue at 1.5T and 3T measured using IDEAL. *J Magnetic Resonance Imaging* (2005) 1:87–91.

28. Edden RA, Smith SA, Barker PB. Longitudinal and multi-echo transverse relaxation times of normal breast tissue at 3 Tesla. *J magnetic resonance imaging: JMRI* (2010) 32(4):982–7.

29. Baron P, Deckers R, Knuttel FM, Bartels LW. T1 and T2 temperature dependence of female human breast adipose tissue at 1.5 T: groundwork for monitoring thermal therapies in the breast. *NMR biomed* (2015) 44(4):1463–70.

30. Liu L, Yin B, Geng DY, Lu YP, Peng WJ. Changes of T2 relaxation time from neoadjuvant chemotherapy in breast cancer lesions. *Iranian J Radiol Q J published by Iranian Radiol Soc* (2016) 13(3):e24014.

31. Liu L, Yin B, Shek K, Geng D, Lu Y, Wen J, et al. Role of quantitative analysis of T2 relaxation time in differentiating benign from malignant breast lesions. *J Int Med Res* (2018) 46(5):1928–35.

32. Buzan MT, Eichinger M, Kreuter M, Kauczor H-U, Herth FJ, Warth A, et al. T2 mapping of CT remodelling patterns in interstitial lung disease. *Eur Radiol* (2015) 25(11):3167–74.

33. Busch S, Hatridge M, Mößle M, Myers W, Wong T, Mück M, et al. Measurements of T 1-relaxation in *ex vivo* prostate tissue at 132 μt. *Magnetic resonance Med* (2012) 67(4):1138–45.

34. Simpkin CJ, Morgan VA, GIles SL, Riches SF, Parker C, deSouza NM. Relationship between T2 relaxation and apparent diffusion coefficient in malignant and non-malignant prostate regions and the effect of peripheral zone fractional volume. *Br J Radiol* (2013) 86(1024):20120469.

35. Foltz WD, Wu A, Chung P, Catton C, Bayley A, Milosevic M, et al. Changes in apparent diffusion coefficient and T2 relaxation during radiotherapy for prostate cancer. *J magnetic resonance imaging: JMRI* (2012) 37(4):909–16.

36. Jambor I, Pesola M, Taimen P, Merisaari H, Boström PJ, Minn H, et al. Rotating frame relaxation imaging of prostate cancer: Repeatability, cancer detection, and Gleason score prediction. *Magnetic resonance Med* (2016) 75 (1):337–44. doi: 10.1002/mrm.25647

37. Mirrashed F, Sharp JC. *In vivo* morphological characterisation of skin by MRI micro-imaging methods. *Skin Res Technol* (2004) 1:149–60. doi: 10.1111/j.1600-0846.2004.00071.x

38. Borhani AA, Hosseinzadeh K. Quantitative versus qualitative methods in evaluation of T2 signal intensity to improve accuracy in diagnosis of pheochromocytoma. *AJR Am J roentgenol* (2015) 205(1):302–10. doi: 10.2214/AJR.14.13273

39. Cieszanowski A, Anysz-Grodzicka A, Szeszkowski W, Kaczynski B, Maj E, Gornicka B, et al. Characterization of focal liver lesions using quantitative techniques: comparison of apparent diffusion coefficient values and T2 relaxation times. *Eur Radiol* (2012) 22(11):2514–24. doi: 10.1007/s00330-012-2519-x

40. Farraher SW, Jara H, Chang KJ, Ozonoff A, Soto JA. Differentiation of hepatocellular carcinoma and hepatic metastasis from cysts and hemangiomas with calculated T2 relaxation times and the T1/T2 relaxation times ratio. *J magnetic resonance imaging: JMRI* (2006) 24 (6):1333–41. doi: 10.1002/jmri.20758

41. Cieszanowski A, Szeszkowski W, Gołębiowski M, Bielecki DK, Grodzicki M, Pruszynski B. Discrimination of benign from malignant hepatic lesions based on their T2-relaxation times calculated from moderately T2-weighted turbo SE sequence. *Eur Radiol* (2002) 12(9):2273–9.

42. Chan YL, Lee SF, Yu SC, Lai P, Ching AA. Hepatic malignant tumour versus cavernous haemangioma: Differentiation on multiple breath-hold turbo spin-echo MRI sequences with different T2-weighting and T2-relaxation time measurements on a single slice multi-echo sequence. *Clin Radiol* (2002) 57(4):250–7.

43. Onaya H, Itai Y, Ahmadi T, Yoshioka H, Okumura T, Akine Y, et al. Recurrent hepatocellular carcinoma versus radiation-induced hepatic injury: differential diagnosis with MR imaging. *Magnetic resonance Imaging* (2001) 19(1):41–6.

44. Cieszanowski A, Szeszkowski W, Gołębiowski M, Grodzicki M, Maj E, Palczewski P, et al. Differentiation of focal hepatic lesions in MR imaging with the use of combined quantitative and qualitative analysis. *Polish J Radiol* (2007) 72(1):26–35.

45. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffman TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Br Med J* (2021) 372:n71.

46. Bojorquez JZ, Bricq S, Acquitter C, Brunotte F., Walker PM, Lalande A. What are normal relaxation times of tissues at 3 T? *Magnetic resonance Imaging* (2017) 35:69–80.

47. Wolf M, de Boer A, Sharma K, Boor P, Leiner T, Sunder-Plassmann G, et al. Magnetic resonance imaging T1- and T2-mapping to assess renal structure and function: a systematic review and statement paper. *Nephrol dialysis Transplant* (2018) 33(suppl_2):ii41–ii50.

48. Reiter G, Reiter C, Kräuter C, Fuchsjäger M, Reiter U. Cardiac magnetic resonance T1 mapping. part 1: Aspects of acquisition and evaluation. *Eur J Radiol* (2018) 33(5):223–34.

49. Reiter U, Reiter C, Kräuter C, Fuchsjäger M, Reiter G. Cardiac magnetic resonance T1 mapping. part 2: Diagnostic potential and applications. *Eur J Radiol* (2018) 109:235–47.

50. Mamisch TC, Trattnig S, Quirbach S, Marlovits S, White LM, Welsch GH. Quantitative T2 mapping of knee cartilage: differentiation of healthy control cartilage and cartilage repair tissue in the knee with unloading–initial results. *Radiology* (2010) 254(3):818–26.

51. Kingsley PB. Methods of measuring spin-lattice (T1) relaxation times: An annotated bibliography. *Concepts Magnetic Resonance* (1999) 39:243–76.

52. Feng H, Xia X, Li C, Song Y, Qin C, Zhang Y, et al. TYR as a multifunctional reporter gene regulated by the tet-on system for multimodality imaging: An in vitro study. *Sci Rep* (2015) 5:15502.

53. Ellis RL. Optimal timing of breast MRI examinations for premenopausal women who do not have a normal menstrual cycle. *Am J Roentgenol* (2009) 193 (6):1738–40.

54. Schwendener N, Jackowski C, Schuster F, Persson A, Warntjes MJ, Zech W-D. Temperature-corrected post-mortem 1.5 T MRI quantification of non-pathologic upper abdominal organs. *Int J legal Med* (2017) 131(5):1369–76.

55. Rogers T, Dabir D, Mahmoud I, Voigt T, Schaeffter T, Nagel E, et al. Standardization of T1 measurements with MOLLI in differentiation between health and disease–the ConSept study. *J Cardiovasc magnetic resonance* (2013) 15(1):78.

56. Mordi I, Carrick D, Bezerra H, Tzemos N. T1 and T2 mapping for early diagnosis of dilated non-ischaemic cardiomyopathy in middle-aged patients and differentiation from normal physiological adaptation. *Eur Heart J Cardiovasc Imaging* (2016) 17(7):797–803.

57. Granitz M, Motloch LJ, Granitz C, Meissnitzer M, Hitzl W, Hergan K, et al. Comparison of native myocardial T1 and T2 mapping at 1.5T and 3T in healthy volunteers. *Wiener klinische Wochenschrift* (2019) 143–55.

58. Mitsouras D, Owens CD, Conte MS, Ersoy H, Creager MA, Rybicki FJ, et al. *In vivo* differentiation of two vessel wall layers in lower extremity peripheral vein bypass grafts: application of high-resolution inner-volume black blood 3D FSE. *Magnetic resonance Med* (2009) 131(5):607–15. doi: 10.1002/mrm.22037

59. Gillmann C, Coras R, Rössler K., Doerfler A, Uder M, Blümcke I, et al. Ultra-high field MRI of human hippocampi: Morphological and multiparametric differentiation of hippocampal sclerosis subtypes. *PloS One* (2018) 13(4):e0196008. doi: 10.1371/journal.pone.0196008

60. Olsson E, Folkesson E, Peterson P, Önnerfjord P, Tjörnstrand J, Hughes HV, et al. Ultra-high field magnetic resonance imaging parameter mapping in the posterior horn of *ex vivo* human menisci. *Osteoarthritis cartilage* (2019) 27(3):476–83. doi: 10.1016/j.joca.2018.12.003

61. Fan G, Wu Z, Pan and Q. Guo S. Quantitative study of MR T1 and T2 relaxation times and 1HMRS in gray matter of normal adult brain. *Chin Med J* (2003) 128:400–4.

62. Fermin-Delgado R, Roa-Sanchez P, Speckter H, Perez-Then E, Rivera-Meija D, Foerster B, et al. Involvement of globus pallidus and midbrain nuclei in pantothenate kinase-associated neurodegeneration: measurement of T2 and T2* time. *Clin neuroradiol* (2013) 3:11–5. doi: 10.1007/s00062-011-0127-9

63. Stehling C, Niederstadt T, Krämer S, Kugel H, Schwindt W, Heindel W, et al. Comparison of a T1-weighted inversion-recovery-, gradient-echo- and spin-echo sequence for imaging of the brain at 3.0 Tesla TT. *RöFo Fortschr auf dem Gebiete der Röntgenstrahlen und der Nuklearmedizin* (2005) 177(4):536–42. doi: 10.1055/s-2005-857902

64. Biffar A, Baur-Melnyk A, Schmidt GP, Reiser MF, Dietrich O. Multiparameter MRI assessment of normal-appearing and diseased vertebral bone marrow. *Eur Radiol* (2010) 20(11):2679–89. doi: 10.1007/s00330-010-1833-4

65. Piper RJ, Mikhael S, Wardlaw JM, Laidlaw DH, Whittle IR, Bastin ME. Imaging signatures of meningioma and low-grade glioma: a diffusion tensor, magnetization transfer and quantitative longitudinal relaxation time MRI study. *Magnetic resonance Imaging* (2016) 45(4):596–602. doi: 10.1016/j.mri.2015.12.006

66. Nebelung S, Brill N, Tingart M, Pufe T, Kuhl C, Jahr H, et al. Quantitative OCT and MRI biomarkers for the differentiation of cartilage degeneration. *Skeletal Radiol* (2016) 45(5):505–16. doi: 10.1007/s00256-016-2334-6

67. Toyooka KO, Toyooka S, Virmani AK, Sathyanarayana UG, Euhus DM, Gilcrease M, et al. Loss of expression and aberrant methylation of the CDH13 (H-cadherin) gene in breast and lung carcinomas. *Cancer Res* (2001) 61 (11):4556–60.

68. Starrett GJ, Luengas EM, McCann JL, Ebrahimi D, Temiz NA, Love RP, et al. The DNA cytosine deaminase APOBEC3H haplotype i likely contributes to breast and lung cancer mutagenesis. *Nat Commun* (2016) 7:129183(3):FSO197. doi: 10.1038/ncomms12918

69. Miao Z, Luker KE, Summers BC, Beranovich R, Bhojani MS, Rehemtulla A, et al. CXCR7 (RDC1) promotes breast and lung tumor growth *in vivo* and is expressed on tumor-associated vasculature. *Proc Natl Acad Sci United States America* (2007) 1:15735–40. doi: 10.1073/pnas.0610444104

70. Glinsky VV, Glinsky GV, Rittenhouse-Olson K, Huflejt ME, Glinskii OV, Deutscher SL, et al. The role of thomsen-friedenreich antigen in adhesion of human breast and prostate cancer cells to the endothelium. *Cancer Res* (2001) 61 (12):4851–7.

71. López-Otín C, Diamandis EP. Breast and prostate cancer: An analysis of common epidemiological, genetic, and biochemical features. *Endocrine Rev* (1998) 19(4):365–96.

72. Spratt DE, Jagsi R. Breast and prostate cancer: Lessons to be shared. *Int J Radiat oncol biol Phys* (2017) 98(2):263–8. doi: 10.1016/j.ijrobp.2017.02.015

73. Zhen JT, Syed J, Nguyen KA, Leapman MS, Agarwal N, Birerley K., et al. Genetic testing for hereditary prostate cancer: Current status and limitations. *Cancer* (2018) 124(15):3105–17.

74. Ray KJ, Sibson NR, Kiltie AE. Treatment of breast and prostate cancer by hypofractionated radiotherapy: Potential risks and benefits. *Clin Oncol J R Coll Radiol* (2015) 27(7):420–6.

75. Malikova MA, Tkacz JN, Slanetz PJ, Guo C-Y, Aakil A, Jara H. Evaluation of T1/T2 ratios in a pilot study as a potential biomarker of biopsy: Proven benign and malignant breast lesions in correlation with histopathological disease stage. *Future Sci OA* (2017) 3(3):FSO197.

76. Huber AT, Bravetti M, Lamy J, Bacoyannis T, Roux C, de Cesare A, et al. Non-invasive differentiation of idiopathic inflammatory myopathy with cardiac involvement from acute viral myocarditis using cardiovascular magnetic resonance imaging T1 and T2 mapping. *J Cardiovasc magnetic resonance* (2018) 20(1):11.

77. Grünewald RA, Farrow T, Vaughan P, Rittey CD, Mundy J. A magnetic resonance study of complicated early childhood convulsion. *J neurol neurosurg Psychiatry* (2001) 71(5):638–42.

78. Farrow TF, Dickson J, Grünewald RA. A six-year follow-up MRI study of complicated early childhood convulsion. *Pediatr Neurol* (2006) 27(7):257–60. doi: 10.1016/j.pediatrneurol.2006.04.002

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 09 January 2023
DOI 10.3389/fonc.2022.1087792

OPEN ACCESS

EDITED BY
Bahram Mohajer,
Johns Hopkins Medicine, United States

REVIEWED BY
Vetri Sudar Jayaprakasam,
Memorial Sloan Kettering Cancer Center, United States
Fang Xie,
Fudan University, China

*CORRESPONDENCE
Xuejuan Wang
xuejuan_wang@hotmail.com
Zhi Yang
pekyz@163.com
Aiwen Wu
wuaw@sina.com

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

RECEIVED 02 November 2022
ACCEPTED 19 December 2022
PUBLISHED 09 January 2023

CITATION
Lin X, Li Y, Wang S, Zhang Y, Chen X, Wei M, Zhu H, Wu A, Yang Z and Wang X (2023) Diagnostic value of [^{68}Ga]Ga-FAPI-04 in patients with colorectal cancer in comparison with [^{18}F]F-FDG PET/CT.
Front. Oncol. 12:1087792.
doi: 10.3389/fonc.2022.1087792

Diagnostic value of [^{68}Ga]Ga-FAPI-04 in patients with colorectal cancer in comparison with [^{18}F]F-FDG PET/CT

Xinfeng Lin[1], Yingjie Li[2], Shuailiang Wang[1], Yan Zhang[1], Xuetao Chen[1], Maomao Wei[1], Hua Zhu[1], Aiwen Wu[2*], Zhi Yang[1*] and Xuejuan Wang[1*]

[1]Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education/Beijing), NMPA Key Laboratory for Research and Evaluation of Radiopharmaceuticals (National Medical Products Administration), Department of Nuclear Medicine, Peking University Cancer Hospital and Institute, Beijing, China, [2]Department of Gastrointestinal Cancer Centre, Unit III, Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education/Beijing), Peking University Cancer Hospital and Institute, Beijing, China

Purpose: This study aimed to compare the diagnostic performance of [^{68}Ga]Ga-FAPI-04 PET/CT and [^{18}F]F-FDG PET/CT in primary and metastatic colorectal cancer (CRC) lesions.

Methods: This single-center preliminary clinical study (NCT04750772) was conducted at the Peking University Cancer Hospital & Institute and included 61 participants with CRC who underwent sequential evaluation through PET/CT with [^{18}F]F-FDG and [^{68}Ga]Ga-FAPI-04. Their PET/CT images were analysed to quantify the uptake of the two tracers in the form of maximum standardised uptake (SUV_{max}) values and target-to-background ratio (TBR), which were then compared using Wilcoxon's signed-rank test. The final changes in the tumour–node–metastasis (TNM) stage of all participants were recorded.

Results: Of all the participants, 21 were treatment naïve and 40 had been previously treated. In primary CRC lesions, the average TBRs of [^{68}Ga]Ga-FAPI-04 and [^{18}F]F-FDG were 13.3 $\pm$ 8.9 and 8.2 $\pm$ 6.5, respectively. The SUV_{max} of [^{68}Ga]Ga-FAPI-04 in signet-ring/mucinous carcinomas (11.4 $\pm$ 4.9) was higher than that of [^{18}F]F-FDG (7.9 $\pm$ 3.6) (P = 0.03). Both median SUV_{max} in peritoneal metastases and TBR in liver metastases of [^{68}Ga]Ga-FAPI-04 were higher than those of [^{18}F]F-FDG (5.2 vs. 3.8, $P < 0.001$; 3.7 vs. 1.9, $P < 0.001$, respectively). Compared with [^{18}F]F-FDG PET/CT, clinical TNM staging based on [^{68}Ga]Ga-FAPI-04 PET/CT led to upstaging and downstaging in 10 (16.4%) and 5 participants (8.2%), respectively. Therefore, the treatment options were changed in 13 participants (21.3%), including 9 with additional chemo/radiotherapy and/or surgery and others with avoidance or narrowed scope of surgery.

Conclusion: [^{68}Ga]Ga-FAPI-04 showed potential as a novel PET/CT tracer to detect lymph nodes and distant metastases, which improved CRC staging, thus prompting the optimisation or adjustment of treatment decisions.

KEYWORDS

fibroblast-activation protein inhibitor, colorectal cancer, fibroblast-activation protein, inhibitor, positron emission tomography

1 Introduction

Colorectal cancer (CRC) was reported as the fifth most common cause of cancer-related deaths in the United States in 2022 (1). At diagnosis, 22% of patients with CRC have metastases, and 50% develop metastases during their lifetime. The overall 5-year survival of patients with CRC largely depends on the stage at presentation, varying from 80%–90% in the early stages to 13% in the advanced stage (2, 3). Currently, the key challenge is to establish optimal treatment plans according to the patients' disease stage. Optimal imaging for CRC is crucial for accurate initial staging and the selection of primary therapy as well as during follow-up examinations for the accurate and timely detection of local recurrence and/or metastasis.

Non-invasive molecular imaging novel PET tracers is being increasingly used in the field of clinical oncology. Flourine-18 fluorodeoxyglucose ([^{18}F]F-FDG) PET/CT, which uses altered glucose metabolism in cancer cells, is a valuable imaging modality in CRC management (4, 5). Compared with the routinely recommended conventional imaging modalities, [^{18}F]F-FDG PET/CT can reflect cancer cells activity and the whole-body tumour burden. However, [^{18}F]F-FDG PET/CT has several limitations, including low specificity, inability to detect small-volume lesions and lack of isotope uptake in mucinous and signet-ring cell carcinomas (6, 7). Tumour microenvironment imaging beyond the detection of tumour metabolism is a novel approach to elucidate *in vivo* tumour biology, with potential translational implications in clinical oncology. Fibroblast-activation protein (FAP) is a membrane-anchored peptidase that is highly expressed in cancer-associated fibroblasts (CAFs) in >90% of epithelial tumours, including CRC, and contributes to disease progression and worsening prognosis in various cancers (8–11). Several recently developed quinolone-based FAP inhibitors (FAPIs) coupled to chelators, including gallium-68 (^{68}Ga)-labelled FAPI, are advantageous in staging and restaging many cancers, including peritonitis carcinomatosis, compared with [^{18}F]F-FDG PET/CT (9–11). Koerber et al. (12) and Pang et al. (13) reported the avidity of [^{68}Ga]Ga-FAPI in PET/CT imaging of CRC, supporting the potential use of FAP-targeted imaging in advanced CRC. We hypothesised that [^{68}Ga]Ga-FAPI-04 PET/CT could contribute to CRC staging than conventional [^{18}F]F-FDG PET/CT. Thus, we assessed the avidity of [^{68}Ga]Ga-FAPI-04 in patients with CRC to compare the clinical impact of [^{68}Ga]Ga-FAPI-04 PET/CT on tumour–node–metastasis (TNM) staging with that of standard-of-care [^{18}F]F-FDG PET/CT imaging in participants with primary and recurrent/metastatic CRC.

2 Materials and methods

2.1 Participants

This preliminary clinical trial was approved by our Institutional Review Board (no. 2019KT95) and registered on ClinicalTrials.gov (NCT04750772). Written informed consent was obtained from all participants who were consecutively recruited from the study institution. The inclusion criteria were as follows: age >18 years, histologically confirmed CRC referral to the Nuclear Medicine Department for both [^{18}F]F-FDG and [^{68}Ga]Ga-FAPI-04 PET/CT scans for staging or restaging to aid optimal clinical decision making and provision of written informed consent. Pregnant or lactating women and those with severe liver or kidney dysfunction were excluded. The final cohort comprised 61 participants with CRC. The diagnosis was confirmed through surgery in 25 participants and through endoscopic biopsy in others. All suspicious metastatic lesions were confirmed by histopathology or follow-up for 3–6 months. Histopathology was served as a gold standard reference for the confirmation of the imaging findings by the means of rebiopsy or surgery. If pathological diagnosis was not applicable, conventional imaging (such as CT and MRI, etc.) follow-up for anatomical abnormalities of lesions was performed. Lesions were diagnosed of malignant based on any of the following follow-up criteria: 1) typical malignant features demonstrated by other imaging, especially the contrast-enhanced CT/MRI referred to as the first-line imaging. 2) posttreatment shrinkage or expansion of a suspicious lesion on follow-up imaging indicating improvement or progression, periodic plain CT/MRI scan used as the second-line imaging. 3) Density changes of lesions, such as cortical breakthrough for bone metastases. The finial observation of significant malignant features of suspicious lesions was the follow-

up imaging end-point. The study flowchart of participant enrolment is presented in Supplementary Figure 1.

2.2 Synthesis of [^{18}F]F-FDG and [^{68}Ga]Ga-FAPI-04

[^{18}F]F-FDG was manufactured in accordance with the standard method described by our laboratory using the coincidence [^{18}F]F-FDG synthesis module. The FAPI precursor (DOTA-FAPI-04) was purchased from Huayi Technology Co., Ltd. (China), and synthesis and radiolabelling of [^{68}Ga]Ga-FAPI-04 were performed as previously described (14). The radiochemical purity exceeded 95% for both [^{18}F]F-FDG and [^{68}Ga]Ga-FAPI-04. The final products underwent sterility testing before administration *via* intravenous injection.

2.3 PET/CT imaging

All enrolled participants underwent routine [^{18}F]F-FDG PET/CT and subsequent [^{68}Ga]Ga-FAPI-04 PET/CT within 1 week. All participants fasted for at least 6 h before [^{18}F]F-FDG PET/CT, and a blood glucose level of <10 mmol/L was confirmed before tracer injection. Contrastingly, participants on a normal diet were intravenously injected with 1.85–2.96 MBq/kg [^{68}Ga]Ga-FAPI-04 and underwent imaging using a hybrid PET/CT scanner (Biograph mCT Flow 64; Siemens Healthineers USA, Knoxville, TN, USA) after approximately 1 h. The acquisition was commenced in 6–8 bed positions (1 min/bed) covering the area between the top of the skull and upper thigh. Non-contrast-enhanced CT was performed using 100-mA modulation at 120 kV with a 3-mm slice thickness for attenuation correction and anatomical localisation. All data were transferred to the Syngo MultiModality Workplace (version VE40F; Siemens Healthineers) and reconstructed using the ordered subset expectation maximum algorithm to construct display images in the coronal, axial and sagittal planes.

2.4 Safety

Vital parameters, including blood pressure, heart rate, temperature and respiration rate, of all participants were carefully monitored during the examination. Any abnormal symptoms (e.g. allergy) were addressed as soon as possible.

2.5 Image analysis

All the images were reviewed by two groups of physicians with at least 10 years of experience in nuclear medicine and radiology. The physicians in group 1 (X.C. and M.W.) and group 2 (X.L. and X.W.) independently reviewed the [^{18}F]F-FDG and [^{68}Ga]Ga-FAPI-04 PET/CT images, respectively. Reference information from the other group and all other images and clinical data, including CT, MRI, endoscopic and pathological results, were absent. All differing opinions were interpreted and discussed within the groups until a consensus was reached. The inter-reader agreement within the two groups was expressed using the κ value.

Visual assessment was performed, and positive uptake was defined as focal tracer uptake exceeding background uptake. Circular volumes of interest within tumour lesions and healthy tissues were used to quantify radiotracer biodistribution. Tracer uptake was quantified using maximum standardised uptake (SUV_{max}) values, which was measured by drawing regions of interest around the tumours on transaxial slices that were automatically adapted to a three-dimensional volume of interest with the system software at an 80% isocontour. The normal organs were evaluated using a 1–2-cm-diameter circular sphere. Primary lesions, lymph nodes and distant metastases were analysed. The lymph nodes were classified according to their location as cervical–supraclavicular, thoracic, abdominal and pelvic. The target-to-background ratios (TBRs) of the primary lesions; lymph nodes; and liver, lung, bone and peritoneal metastatic tumours were calculated (the normal transverse colon without physiological uptake, blood pool of the aorta, normal liver tissue, normal lung tissue, L5 and normal mesenterium were used as backgrounds, respectively).

We used TNM classification based on the National Comprehensive Cancer Network *(NCCN)* guidelines (15, 16). In all participants, changes in the TNM stage, metastases localisation and previous oncologic or radio-oncologic management history were recorded.

2.6 Statistical analyses

All statistical analyses were performed using SPSS 23.0 (IBM, Armonk, NY, USA). Inter-reader agreement was evaluated using Kappa test. The uptakes of positive lesions in [^{18}F]F-FDG and [^{68}Ga]Ga-FAPI-04 PET/CT were compared using Wilcoxon's signed-rank test. SUV_{max} and TBR were the main parameters for evaluating the two PET/CT scans, and normally distributed and skewed variables were expressed as means (95% confidence intervals) and medians (ranges), respectively. A two-tailed *P* value of <0.05 was considered statistically significant.

3 Results

3.1 Participant characteristics

Table 1 summarises the clinical characteristics of the 61 participants. The median age of the participants was 62 (range, 32–81) years, and 42 (68.9%) participants were men. The most

TABLE 1 Participants' characteristics.

Characteristic	Value
Participants (*n*)	61
Age (years), median (range)	62 (32–81)
Sex (male:female)	42:19
Colon cancer	26 (42.6%)
Rectal cancer	35 (57.4%)
Treatment status	
Treatment-naïve	21
Neoadjuvant treatment	15
Chemotherapy	5
Radiotherapy	3
Chemotherapy + radiotherapy	7
Chemotherapy/radiotherapy after surgery	25
Pathology	
Adenocarcinoma (poorly differentiated)	8
Adenocarcinoma (moderately differentiated)	35
Adenocarcinoma (well-differentiated)	3
Mucinous/signet-ring cell carcinoma	8
Adenocarcinoma with mucinous component	7
Purpose of PET/CT	
Staging	21
Restaging/therapeutic evaluation	40

CT, computed tomography; PET, positron emission tomography.

common histologic grade was moderate differentiation in 35 (57.4%) participants, whereas 8 (13.1%) and 7 (11.5%) participants had mucinous/signet-ring cell carcinoma and adenocarcinoma with a mucinous component, respectively. Overall, 21 (34.4%) participants underwent PET/CT for initial assessment and staging; the remaining 40 (65.6%) underwent PET/CT for restaging or therapeutic evaluation.

3.2 Safety

All participants tolerated [^{68}Ga]Ga-FAPI-04 PET/CT without any complications. No signs of drug-related side effects were reported during the entire observation period of >5 h.

3.3 Distribution

The inter-reader agreement between groups 1 and 2 was nearly perfect, and the κ value was >0.8 (Supplementary Tables 1–4). SUV_{max} was determined for normal tissues/organs and primary tumours after [^{68}Ga]Ga-FAPI-04 and [^{18}F]F-FDG PET/CT, which were sequentially performed for all the participants. [^{68}Ga]Ga-FAPI-04 activity was significantly lower than [^{18}F]F-FDG activity in several normal organs ($P < 0.001$), especially the brain (SUV_{max}, 0.3 ± 0.2 vs. 10.0 ± 2.8, $P < 0.001$) and liver (SUV_{max}, 1.4 ± 0.4 vs. 2.9 ± 0.5, $P < 0.001$), leading to significantly high TBRs of >2 in both organs ($P < 0.001$). Further, [^{68}Ga]Ga-FAPI-04 uptake was higher than [^{18}F]F-FDG uptake in the salivary and thyroid glands and the pancreas ($P < 0.001$ for all). The detailed distribution of [^{18}F]F-FDG and [^{68}Ga]Ga-FAPI-04 uptakes is presented in Figure 1.

3.4 Primary tumours

The histopathological data of the primary tumours were available for all treatment-naïve participants ($n = 21$) and for those who received neoadjuvant treatment ($n = 15$). Only 1 of these 36 participants had two primary lesions. Therefore, 37 primary lesions were measured. The sensitivity was 100% (37/37) for both [^{68}Ga]Ga-FAPI-04 and [^{18}F]F-FDG PET/CT. The average SUV_{max} and median SUV_{max} (range) of all primary lesions were 9.7 ± 5.4 and 9.7 (2.0–25.5), respectively, on [^{68}Ga]Ga-FAPI-04 PET/CT and 11.4 ± 7.6 and 10.3 (2.4–35.1), respectively, on [^{18}F]F-FDG PET/CT ($P = 0.09$). The average TBR of all 37 primary lesions was significantly higher on [^{68}Ga]Ga-FAPI-04 PET/CT than on [^{18}F]F-FDG PET/CT (13.3 ± 8.9 and 8.2 ± 6.5, respectively; $P < 0.001$). The average SUV_{max} values for [^{68}Ga]Ga-FAPI-04 and [^{18}F]F-FDG were 12.3 ± 4.6 and 14.1 ± 7.3, respectively, in the treatment-naïve group ($P =$

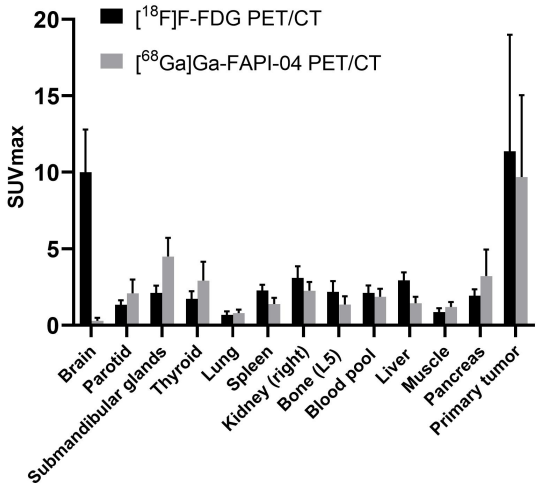


FIGURE 1
PET-based biodistribution analysis of 61 participants evaluated using [^{68}Ga]Ga-FAPI-04 PET/CT and [^{18}F]F-FDG PET/CT 1 h after tracer injection.

0.21) and 5.9 ± 4.1 and 7.4 ± 6.5, respectively, in the neoadjuvant radio-chemotherapy group (P = 0.18). The average TBR of the treatment-naïve lesions was significantly higher on [^{68}Ga]Ga-FAPI-04 PET/CT than on [^{18}F]F-FDG PET/CT (17.6 ± 8.5 vs. 10.5 ± 7.2, P = 0.002), whereas the average TBR was not different between the imaging modalities for the post-treatment lesions (7.0 ± 5.0 vs. 5.0 ± 3.4, P = 0.061).

The analysis of treatment-naïve primary tumours revealed that the avidity of [^{68}Ga]Ga-FAPI-04 (11.4 ± 4.9) was significantly higher than that of [^{18}F]F-FDG (7.9 ± 3.6) in signet-ring/mucinous carcinomas (P = 0.03; Figure 2). Additionally, [^{68}Ga]Ga-FAPI-04 uptake was significantly lower than [^{18}F]F-FDG uptake in poorly differentiated carcinomas (average SUV_{max}, 12.7 ± 3.7 vs. 18.1 ± 4.1; P = 0.04). There was also a significant difference in the SUV_{max} of primary lesions between [^{68}Ga]Ga-FAPI-04 and [^{18}F]F-FDG PET/CT among well-differentiated and moderately differentiated carcinomas (average SUV_{max}, 10.8 ± 3.0 vs. 16.2 ± 8.6; P = 0.025). Although the uptake of [^{68}Ga]Ga-FAPI-04 was higher in poorly differentiated carcinomas (12.7 ± 3.7) than in moderately-well differentiated carcinomas (10.8 ± 3.0), no significant difference was noted between them (P = 0.074). Interestingly, in the neoadjuvant chemotherapy group, there was no significant difference in the SUV_{max} of primary lesions between [^{68}Ga]Ga-FAPI-04 and [^{18}F]F-FDG PET/CT among all carcinoma types (P = 0.182).

3.5 Changes in TNM stage

Compared with staging based on [^{18}F]F-FDG PET/CT, [^{68}Ga]Ga-FAPI-04 PET/CT revealed additional findings in 34 of the 61 participants, which led to changes in staging in 15 participants. Particularly, 6 (28.6%) of the 21 treatment-naïve participants were upstaged and 2 (9.5%) were downstaged (Table 2).

Among all participants who were upstaged based on [^{68}Ga]Ga-FAPI-04 PET/CT findings (10/61, 16.4%), the changes were based on the detection of new or additional distant metastases in one or more organ systems. All additional findings were confirmed through biopsy or surgery (10/34, 29.4%) or through other conventional imaging modalities (24/34, 70.6%). Moreover, in 16 participants, new lymph node metastases were detected but did not lead to changes in the stage (Table 2). The median uptake of [^{68}Ga]Ga-FAPI-04 was higher than that of [^{18}F]F-FDG in both abdominal (6.4 vs. 4.2, P < 0.001) and pelvic lymph nodes (4.6 vs. 2.7, P < 0.001; Table 3, Figure 3). The TBRs of both abdominal (3.5 vs. 2.1, P < 0.001) and pelvic lymph nodes (2.9 vs. 1.2, P < 0.001) were also significantly higher in [^{68}Ga]Ga-FAPI-04 PET/CT images than in [^{18}F]F-FDG PET/CT images. However, the median uptake (3.0 vs. 4.7, P < 0.001) and TBR (1.4 vs. 2.4, P < 0.001) of [^{68}Ga]Ga-FAPI-04 were lower than those of [^{18}F]F-FDG in thoracic lymph nodes (Table 3).

Imaging with [^{68}Ga]Ga-FAPI-04 PET/CT led to upstaging based on the detection of peritoneal and bone metastases in four

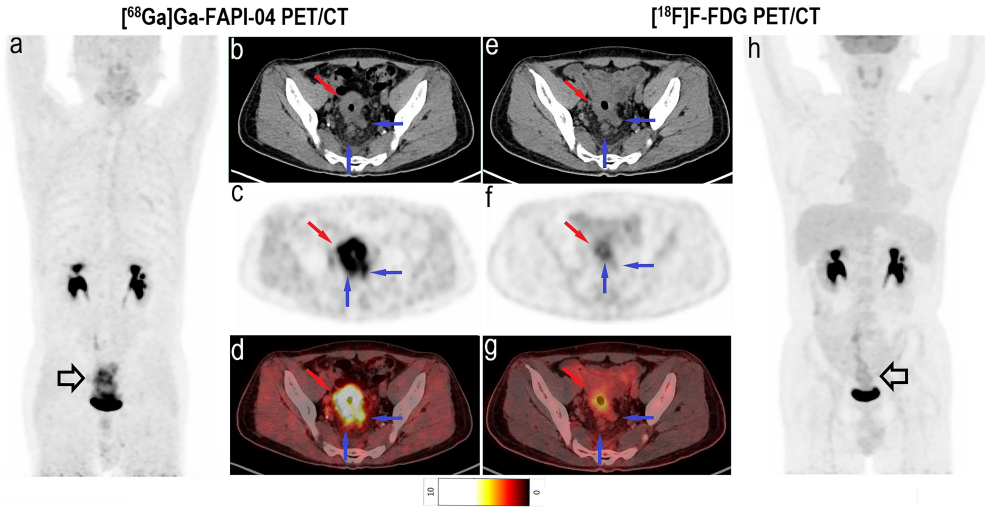


FIGURE 2
Images of a 49-year-old man with rectal mucinous carcinoma for staging. **(A)** Mean intensity projection images of [^{68}Ga]Ga-FAPI-04 PET/CT scan. **(B–D)** Axial CT, PET and fused images of [^{68}Ga]Ga-FAPI-04 PET/CT scan. **(E–G)** Axial CT, PET and fused images of [^{18}F]F-FDG PET/CT scan. **(H)** Mean intensity projection images of [^{18}F]F-FDG PET/CT scan. Compared with [^{18}F]F-FDG PET/CT, [^{68}Ga]Ga-FAPI-04 PET/CT exhibits the primary lesion (hollow black and red arrows) and suspicious lymph nodes (hollow black and blue arrows) more clearly because of the higher tracer uptake. The SUV_{max} of primary lesions was 14.6 for [^{68}Ga]Ga-FAPI-04 and 4.4 for [^{18}F]F-FDG PET/CT. The [^{68}Ga]Ga-FAPI-04 PET/CT images also depict more intense tracer uptake in suspicious lymph nodes than [^{18}F]F-FDG PET/CT images, with SUV_{max} values of 5.6–6.0 and 1.5–2.0, respectively.

10.3389/fonc.2022.1087792

TABLE 2 Comparison of FDG PET-based and FAPI PET-based TNM staging of 34 CRC participants with additional findings.

Number	Treatment status	TNM stage (FDG PET-based)	TNM stage (FAPI PET-based)	Additional finding in FAPI PET	Reference standard	Change in staging
2	Treatment-naïve	T4aN2bM0 (III B)	T4aN2bM1c (IV C)	Peritoneal metastasis	Contrast-enhanced MRI	Upstage
4	Treatment-naïve	T4aN2bM1c (IV C)	T4aN2bM1c (IV C)	2 LNs	Shrinkage after treatment (contrast CT)	None
5	Treatment-naïve	T1/2N0M1a (IV A)	T1/2N0M0 (I)	No rectal metastasis	Endoscopic biopsy	Downstage
6	Treatment-naïve	T4aN2aM1c (IV C)	T4aN2aM1c (IV C)	2 LNs	Shrinkage after treatment (contrast CT)	None
11	Neoadjuvant treatment	yT0N0M1a (IV A)	yT0N0M0	No small intestinal metastasis	Endoscopic biopsy	Downstage
15	Treatment-naïve	T3N1bM1a (IV A)	T3N1bM1a (IV A)	1 LN	Shrinkage after treatment (contrast CT)	None
16	Treatment-naïve	T4bN2bM0 (III C)	T4bN2bM0 (III C)	5 LNs	Shrinkage after treatment (contrast CT)	None
17	Treatment-naïve	T4aN1bM0 (III B)	T4aN1bM0 (III B)	1 LN	Shrinkage after treatment (contrast CT)	None
20	Treatment-naïve	T3N2aM1a (IV A)	T3N2aM1a (IV A)	1 LN	Shrinkage after treatment (contrast CT)	None
21	Treatment-naïve	T4aN1bM0 (III B)	T4aN1bM0 (III B)	1 LN	Surgery	None
22	Treatment-naïve	T3N1bM0 (III B)	T3N1bM1c (IV C)	1 LN + Peritoneal metastasis	Surgery	Upstage
23	Treatment-naïve	T3N1bM0 (III B)	T3N1bM0 (III B)	1 LN	Surgery	None
24	Post-operation	M1	M0	No metastatic recurrence in anastomotic stoma	Endoscopic biopsy	Downstage
26	Treatment-naïve	T3N1bM0 (III B)	T3N1bM0 (III B)	1 LN	Shrinkage after treatment (contrast CT)	None
27	Treatment-naïve	T3N0M0 (II A)	T3N1aM0 (III B)	1 LN	Surgery	Upstage
28	Treatment-naïve	T3N1cM0 (III B)	T3N1cM1a (IV A)	1 LN + Left acetabulum metastasis	Staging surgery and shrinkage after treatment (contrast CT)	Upstage
30	Neoadjuvant treatment	yT3N1aM0 (III B)	yT3N1bM0 (III B)	1 LN	Shrinkage after treatment (contrast CT)	None
33	Treatment-naïve	T4bN2bM0 (III C)	T4bN2bM0 (III C)	2 LNs	Shrinkage after treatment (contrast CT)	None
35	Treatment-naïve	T4aN1aM0 (III B)	T4aN2bM0 (III C)	7 LNs	Shrinkage after treatment (contrast CT)	Upstage
36	Treatment-naïve	T4aN2aM0 (III C)	T4aN2aM1a (IV A)	2 LNs + Liver metastasis	Contrast-enhanced MRI	Upstage
37	Post-operation	M1c	M1c	No 2 right axillary and internal mammary LN metastases	No change in size (follow-up by CT)	None
38	Treatment-naïve	T4aN1aM0 (III B)	T4aN1bM0 (III B)	1 LN	Surgery	None

(Continued)

TABLE 2 Continued

Number	Treatment status	TNM stage (FDG PET-based)	TNM stage (FAPI PET-based)	Additional finding in FAPI PET	Reference standard	Change in staging
40	Post-operation	M1c	M1c	No 1 right hilar LN metastasis	No change in size (follow-up by CT)	None
41	Post-operation	M1b	M0	No 7 mediastinal and hilar LN metastases	No change in size (follow-up by CT)	Downstage
44	Post-operation	M1c	M1c	1 left hilar LN metastasis	Shrinkage after treatment (contrast CT)	None
46	Post-operation	M1c	M1c	No metastatic recurrence in anastomotic stoma	Endoscopic biopsy	None
47	Post-operation	M1b	M1c	Peritoneal metastases	Expansion after drug resistance	Upstage
49	Post-operation	M0	M1a	Left 11th rib metastasis	Changes in bone density (contrast CT)	Upstage
50	Post-operation	M1a	M1c	Peritoneal metastases (liver capsule)	Contrast-enhanced MRI	Upstage
52	Post-operation	M1c	M1c	> 7 peritoneal metastases	Shrinkage after treatment (contrast CT)	None
53	Neoadjuvant treatment	yT3N0M0 (II A)	yT3N1bM1b (IV B)	9 LNs	Shrinkage after treatment (contrast CT)	Upstage
55	Treatment-naïve	T3N0M1a (IV A)	T3N0M0 (II A)	No 1 left hilar LN metastasis	No change in size (follow-up by CT)	Downstage
58	Neoadjuvant treatment	yT1/2N0M1b (IV B)	yT1/2N0M1a (IV A)	No 4 cervical LN metastases	No change in size (follow-up by CT)	None
59	Post-operation	M1c	M1c	No metastatic recurrence in anastomotic stoma	Endoscopic biopsy	None

T1/2, inability to differentiate T1 and T2 stages using PET/CT; LN, lymph node; FDG, fluorodeoxyglucose; FAPI, fibroblast-activation protein inhibitor; SUV_{max}, maximum standardised uptake value; TNM, tumour–node–metastasis.

and two participants, respectively (Table 2). In all 20 participants, the SUV_{max} (5.2 vs. 3.8, $P < 0.001$) and TBR (6.9 vs. 3.8, $P < 0.001$) of [^{68}Ga]Ga-FAPI-04 were higher than those of [^{18}F]F-FDG in peritoneal metastases (Table 3). Compared with the [^{18}F]F-FDG PET/CT images, the peritoneal metastases were clearly visible in the [^{68}Ga]Ga-FAPI-04 PET/CT images (Figures 4, 5). There was no significant difference in SUV_{max} between [^{68}Ga]Ga-FAPI-04 and [^{18}F]F-FDG in positive lung lesions ($P = 0.484$), but the TBR of [^{68}Ga]Ga-FAPI-04 was significantly lower than that of [^{18}F]F-FDG ($P = 0.017$; Table 3).

Although there was no significant difference in the SUV_{max} of [^{68}Ga]Ga-FAPI-04 and [^{18}F]F-FDG in positive liver lesions (3.9 vs. 4.6, $P = 0.951$), the number of positive liver lesions detected using [^{68}Ga]Ga-FAPI-04 PET/CT was higher than that detected by [^{18}F]F-FDG PET/CT because of the lower background SUV_{max} of [^{68}Ga]Ga-FAPI-04. Further, 30 positive liver lesions detected using [^{68}Ga]Ga-FAPI-04 PET/CT were confirmed as metastases through surgery/biopsy and other imaging modalities. Only 16 of these positive liver lesions were detected by [^{18}F]F-FDG PET/CT (53.3%, 16/30). The TBR of [^{68}Ga]Ga-FAPI-04 was significantly higher than that of [^{18}F]F-FDG (3.7 vs. 1.9, $P < 0.001$), and the liver metastases were clearly visible in [^{68}Ga]Ga-FAPI-04 PET/CT images and finally demonstrated using contrast-enhanced MRI (Figure 6). Table 3 presents the detailed comparison of liver metastases detected using [^{68}Ga]Ga-FAPI-04 and [^{18}F]F-FDG PET/CT.

The final staging changes in 15 participants were verified based on the reference standards; thus, 13 participants' treatment options were changed from their pre-examination or originally planned regimens based on the changed stage. Chemo/radiotherapy and/or surgery was added in nine participants (#2, #22, #28, #35, #47, #49, #50, #53 and #55), and surgery was avoided or its scope was narrowed in the other participants (#5, #11, #24 and #41).

4 Discussion

FAP is an excellent target for tumour stroma, and ^{68}Ga-FAPIs, as newer imaging tracers, present a promising alternative to [^{18}F]F-FDG. This preliminary clinical study investigated the avidity of [^{68}Ga]Ga-FAPI-04 in CRC and explored the potential

TABLE 3 Comparison of [^{18}F]F-FDG and [^{68}Ga]Ga-FAPI-04 uptake in colorectal tumour sites.

Tumour sites and parameters	[^{18}F]F-FDG PET/CT	[^{68}Ga]Ga-FAPI-04 PET/CT	*P*-value
Primary tumour			
No. of lesions (participants)	37 (36)	37 (36)	
Mean SUV_{max} (95% CI)	11.4 (9.1, 13.8)	9.7 (8.0, 11.4)	0.09
Mean TBR (95% CI)	8.2 (6.4, 10.6)	13.3 (10.5, 16.2)	<0.001
Involved lymph nodes			
Cervical–supraclavicular			
No. of lesions (participants)	6 (2)	4 (2)	
Median SUV_{max} (range)	3.2 (2.1–4.3)	2 (1.8–2.5)	0.027
Median TBR (range)	1.6 (1.3–2.3)	1.2 (1.1–1.5)	0.027
Thoracic*			
No. of lesions (participants)	26 (13)	18 (13)	
Median SUV_{max} (range)	4.7 (2.4–10.2)	3.0 (1.9–12.9)	<0.001
Median TBR (range)	2.4 (1.0–4.9)	1.4 (1.0–5.4)	<0.001
Abdominal†			
No. of lesions (participants)	26 (9)	38 (9)	
Median SUV_{max} (range)	4.2 (2.1–9.4)	6.4 (2.7–20.5)	<0.001
Median TBR (range)	2.1 (1.2–6.3)	3.5 (1.5–13.7)	<0.001
Pelvic§			
No. of lesions (participants)	37 (15)	72 (15)	
Median SUV_{max} (range)	2.7 (2.0–6.6)	4.6 (2.5–17.6)	<0.001
Median TBR (range)	1.2 (0.7–3.3)	2.9 (1.5–9.8)	<0.001
Involved distant lesions			
Liver			
No. of lesions (participants)	16 (9)	30 (9)	
Median SUV_{max} (range)	4.6 (2.5–9.9)	3.9 (1.7–12.2)	0.951
Median TBR (range)	1.9 (0.6–4.7)	3.7 (1.6–7.6)	<0.001
Lung			
No. of lesions (participants)	24 (9)	23 (9)	
Median SUV_{max} (range)	2.1 (0.9–11.0)	2.1 (1.0–11.2)	0.484
Median TBR (range)	4.1 (1.3–22.0)	3.5 (1.2–22.4)	0.017
Bone			
No. of lesions (participants)	6 (5)	8 (5)	
Median SUV_{max} (range)	5.5 (4.0–8.8)	7.9 (2.8–14.0)	0.036
Median TBR (range)	3.4 (2.9–5.0)	8.7 (3.1–10.9)	0.017
Peritoneum			
No. of lesions (participants)	45 (20)	60 (20)	

(Continued)

TABLE 3 Continued

Tumour sites and parameters	[^{18}F]F-FDG PET/CT	[^{68}Ga]Ga-FAPI-04 PET/CT	*P*-value
Median SUV_{max} (range)	3.8 (1.1–16.4)	5.2 (2.1–12.6)	<0.001
Median TBR (range)	3.8 (1.8–29.0)	6.9 (2.6–31.5)	<0.001
Brain			
No. of lesions (participants)	1 (1)	1 (1)	
Median SUV_{max} (range)	6.9 (6.9)	2.1 (2.1)	NA
Median TBR (range)	0.9 (0.9)	21 (21)	NA

*Lymph nodes in the thoracic regions include mediastinal or/and hilar, axillary and internal mammary lymph nodes.
†Lymph nodes in the abdominal regions include para-aortic, retroperitoneal and celiac lymph nodes.
§Lymph nodes in the pelvic regions include pelvic, iliac and inguinal lymph nodes; ^{18}F, fluorine 18; ^{68}Ga, gallium 68; FDG, fluorodeoxyglucose; FAPI, fibroblast-activation protein inhibitor; SUV_{max}, maximum standardised uptake value; NA, not applicable.

utility of [^{68}Ga]Ga-FAPI-04 PET/CT as the sole imaging modality for assessing primary and recurrent/metastatic CRC. Our analyses indicated that [^{68}Ga]Ga-FAPI-04 PET/CT improved tumour staging in patients with CRC owing to favourable tumour/background activity and low tracer uptake in the gastrointestinal tract. Moreover, signet-ring/mucinous carcinomas accumulated more [^{68}Ga]Ga-FAPI-04 than [^{18}F]F-FDG, whereas FAPI avidity was lower than FDG avidity in poorly differentiated carcinomas.

As key constituents of the tumour stroma, CAFs can support the immunosuppressive microenvironment, tumour cell growth, progression and metastatic potential. Expressed by CAFs, FAP is an attractive diagnostic and therapeutic target (8, 10, 17). [^{68}Ga]Ga-FAPI-04 PET/CT is characterised by high tumour/background activity and is more sensitive than [^{18}F]F-FDG PET/CT for identifying primary gastrointestinal carcinoma lesions (12, 13).

The origin, number and distribution of FAP-expressing CAFs and the number of FAP molecules per cell may differ among tumours. Mona et al. (18) reported a strong correlation between tumour [^{68}Ga]Ga-FAPI-46 uptake intensity and histopathological FAP expression in colon cancer. We expected variations in intra-tumoural tracer distribution in treatment-naïve patients with specific histopathologic types of CRC. In this study, we demonstrated additional FAP expression in signet-ring/mucinous carcinomas, which normally exhibit low [^{18}F]F-FDG uptake (6, 7, 19). The results revealed that [^{68}Ga]Ga-FAPI-04 PET/CT would

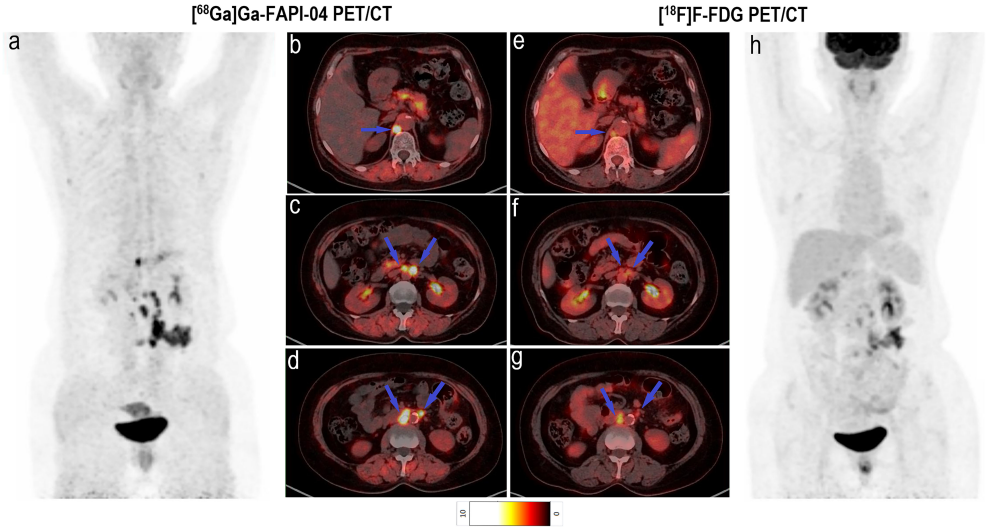


FIGURE 3
Images of a 64-year-old woman with colon cancer for staging. **(A)** Mean intensity projection images of [^{68}Ga]Ga-FAPI-04 PET/CT scans. **(B–D)** Axial fused images of suspicious lymph nodes at different levels of the body in [^{68}Ga]Ga-FAPI-04 PET/CT. **(E–G)** Axial fused images of suspicious lymph nodes at different levels of the body in [^{18}F]F-FDG PET/CT. **(H)** Mean intensity projection images of [^{18}F]F-FDG PET/CT scan. [^{68}Ga]Ga-FAPI-04 uptake (blue arrows in **B–D**) was higher than [^{18}F]F-FDG uptake (blue arrows in **E–G**) in both the diaphragmatic and retroperitoneal metastatic lymph nodes, with SUV_{max} values of 5.9–12.0 and 3.4–4.3, respectively.

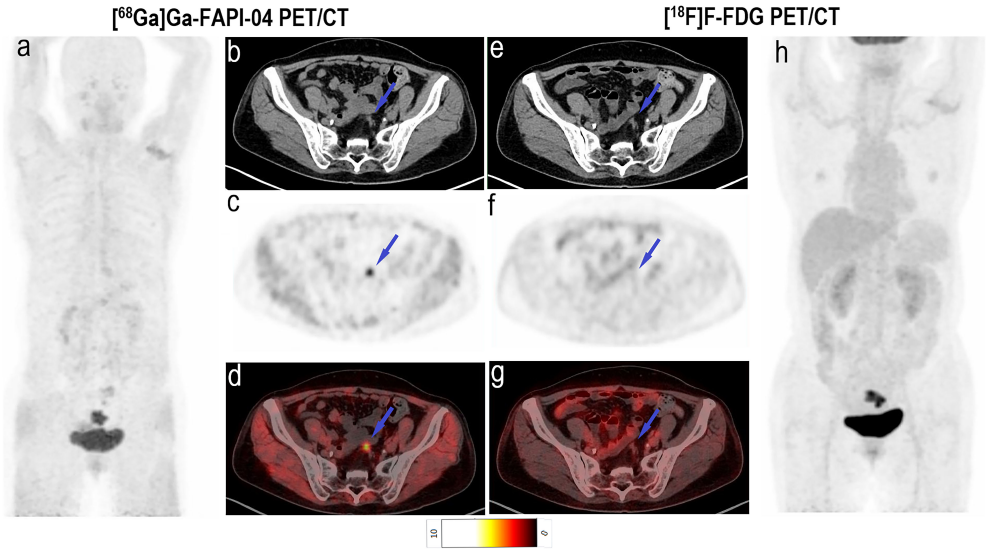


FIGURE 4
Images of a 64-year-old woman with rectal cancer for staging. **(A)** Mean intensity projection images of [^{68}Ga]Ga-FAPI-04 PET/CT scan. **(B–D)** Axial CT, PET and fused images of [^{68}Ga]Ga-FAPI-04 PET/CT scan. **(E–G)** Axial CT, PET and fused images of [^{18}F]F-FDG PET/CT scan. **(H)** Mean intensity projection images of [^{18}F]F-FDG PET/CT scan. Pelvic peritoneal carcinoma is distinctly observed in [^{68}Ga]Ga-FAPI-04 PET/CT images (blue arrows in **b–d**) because of intensive tracer uptake (SUV_{max} 5.5). Conversely, little [^{18}F]F-FDG (SUV_{max} 1.1) uptake results in the small lesion being hardly detectable (blue arrows in **E–G**).

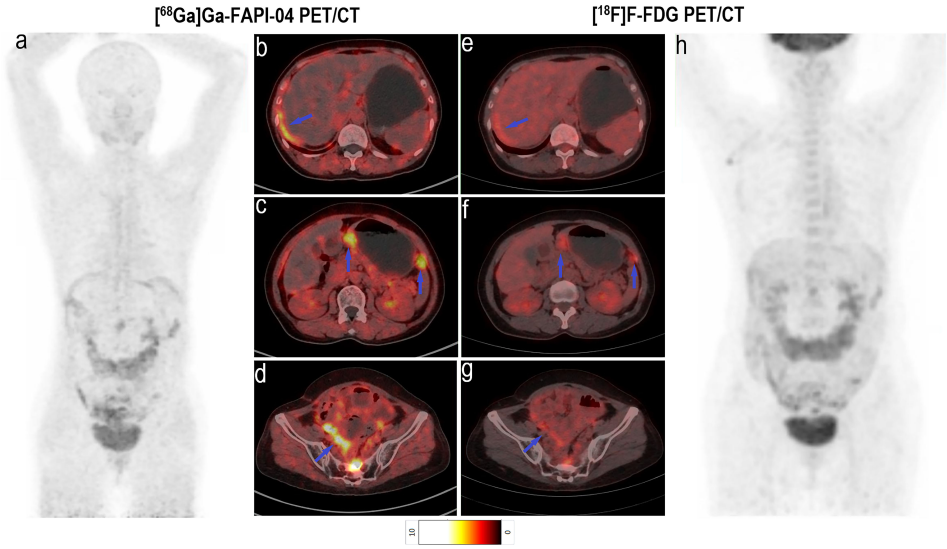


FIGURE 5
Images of a 60-year-old woman with colon cancer and metastatic peritoneal carcinoma for restaging after treatment. **(A)** Mean intensity projection images of [^{68}Ga]Ga-FAPI-04 PET/CT scans. **(B–D)** Axial fused images of metastatic peritoneal carcinomas at different body levels in [^{68}Ga]Ga-FAPI-04 PET/CT. **(E–G)** Axial fused images of metastatic peritoneal carcinomas at different body levels in [^{18}F]F-FDG PET/CT. **(H-h)** Mean intensity projection images of [^{18}F]F-FDG PET/CT scan. [^{68}Ga]Ga-FAPI-04 uptake (blue arrows in **B–D**) was higher than the [^{18}F]F-FDG uptake (blue arrows in **E–G**) in peritoneal carcinomas, and SUV_{max} values for [^{68}Ga]Ga-FAPI-04 and [^{18}F]F-FDG PET/CT were 4.6–8.8 and 3.1–3.5, respectively.

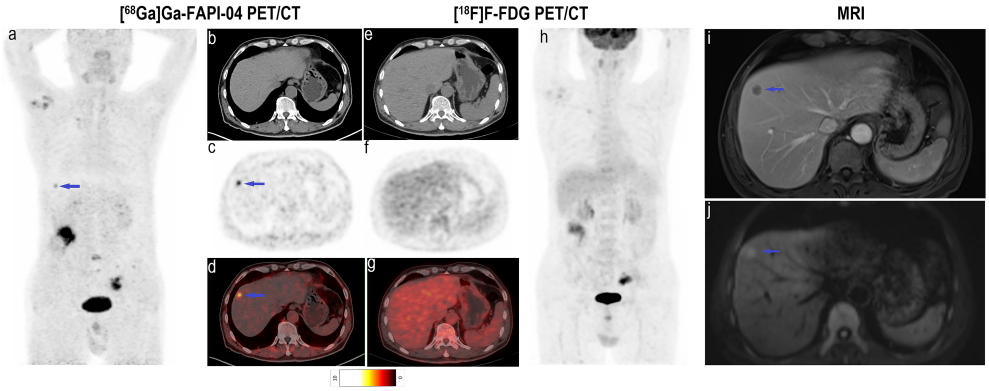


FIGURE 6
Images of a 63-year-old man with colon cancer for staging. **(A)** Mean intensity projection images of [^{68}Ga]Ga-FAPI-04 PET/CT scans. **(B–D)** Axial CT, PET and fused images of [^{68}Ga]Ga-FAPI-04 PET/CT scan. **(E–G)** Axial CT, PET and fused images of [^{18}F]F-FDG PET/CT scan. **(H)** Mean intensity projection images of [^{18}F]F-FDG PET/CT scan. **(I, F)** Images of contrast-enhanced MRI, T2WI (delay phase) and DWI (b = 1000). The suspicious metastatic lesion in the right liver lobe is clearly visible in [^{68}Ga]Ga-FAPI-04 PET/CT images (blue arrows in **A–D**; SUV_{max} 5.0) but absent in [^{18}F]F-FDG PET/CT images. Its presence was confirmed by contrast-enhanced MRI (blue arrows in **I**, **J**).

have a lower false-negative rate than [^{18}F]F-FDG PET/CT in detecting primary and metastatic CRC lesions. Although Solano-Iturri et al. (20) reported that CRC tissues with poor differentiation exhibited a higher percentage of FAP staining than those with moderately-well differentiation, the poorly differentiated carcinomas exhibited moderately higher [^{68}Ga]Ga-FAPI-04 uptake without significant differences between these subtypes in our study. Moreover, we found that poorly differentiated carcinomas exhibited significantly lower [^{68}Ga]Ga-FAPI-04 uptake than [^{18}F]F-FDG uptake, although this subtype showed avidity for both [^{68}Ga]Ga-FAPI-04 and [^{18}F]F-FDG.

The TNM classification provides standard guidelines to classify the extent of cancer metastasis. The degree of tumour progression and invasion at the time of surgical resection as well as patient outcomes are estimated on the basis of this staging system for CRC. This study demonstrated that [^{68}Ga]Ga-FAPI-04 PET/CT could detect both primary tumours and metastases arising from CRC. The sensitivity was 100% for both [^{68}Ga]Ga-FAPI-04 and [^{18}F]F-FDG PET/CT, and no significant differences in SUV_{max} were found between [^{68}Ga]Ga-FAPI-04 and [^{18}F]F-FDG PET/CT images in both treatment-naïve and post-treatment lesions. However, the average TBR of treatment-naïve lesions was significantly higher on [^{68}Ga]Ga-FAPI-04 PET/CT than on [^{18}F]F-FDG PET/CT. This result was consistent with the recently study reported by Halil et al. (21). Thus, [^{68}Ga]Ga-FAPI-04 PET/CT might be more advantageous than [^{18}F]F-FDG PET/CT to improve detecting efficiency in T stage.

The superiority of [^{18}F]F-FDG PET/CT is evident in the detection of lymph node and distant metastases in CRC, and the detection of additional metastases can significantly change treatment plans (22, 23). However, several studies demonstrated that the detection of metastatic regional nodes was low/moderate using [^{18}F]F-FDG PET/CT, illustrating the limitations of this method (24, 25). Several studies suggested that CRC commonly harbours CAF-expressing FAP. Sugai et al. (26) suggested that high FAP expression is correlated with lymph node metastasis in submucosal invasive CRC. Solano-Iturri et al. (20) observed a significant positive correlation between FAP expression in primary CRC tumours and their corresponding local and distant metastases. Thus, this study examined the reliability of [^{68}Ga]Ga-FAPI-04 PET/CT for detecting metastatic CRC lesions. Our results showed that [^{68}Ga]Ga-FAPI-04 PET/CT identified additional findings in 41 metastatic and 15 inflammatory lymph nodes of 24 participants with CRC and improved the N staging in these participants. Additionally, [^{68}Ga]Ga-FAPI-04 uptake was higher than [^{18}F]F-FDG uptake in abdominal and pelvic lymph nodes. However, cervical–supraclavicular and thoracic FDG-avid inflammatory/age-related lymph nodes were FAPI-negative.

According to the M stage, the early detection of isolated metastases in the liver or other sites often improves survival following radical resection (27). Owing to moderate FDG uptake in the liver, [^{18}F]F-FDG PET/CT was not the first choice for identifying liver metastasis. Our data revealed that the hepatic background intensity was significantly lower in [^{68}Ga]Ga-FAPI-04 PET/CT than that in [^{18}F]F-FDG PET/CT, corroborating the findings of previous studies (9, 11, 21). The TBR values of liver metastases were higher in [^{68}Ga]Ga-FAPI-04 than those in [^{18}F]F-FDG PET/CT in this study. Thus, FAPI-imaging might be advantageous for patients with suspected liver metastases, resulting in a potentially high detection rate. In this study, the smallest lesion detected had a diameter of approximately 0.7 cm. However, Halil et al. (21) found that both the SUV_{max} and TBR values of liver metastases were signifcantly higher in [^{18}F]F-FDG than those in [^{68}Ga]Ga-FAPI-04 PET/CT. We believe that this

issue can be clarified with future studies involving larger and homogeneous cohorts. Reportedly, the peritoneum is another common site of CRC metastasis (28), and the degree of peritoneal metastasis determines the choice of treatment (23, 29). The detection rate of peritoneal metastasis using [^{18}F]F-FDG PET/CT is not high, primarily because of intestinal inflammatory uptake, small lesions and other factors, including rare pathological types. [^{68}Ga]Ga-FAPI-04 is an active fibroblast-targeted imaging agent, and the development of peritoneal metastases is primarily because of active fibroblasts (28), which is supported by the significantly higher [^{68}Ga]Ga-FAPI-04 uptake in peritoneal metastatic lesions compared with [^{18}F]F-FDG uptake observed in this study. This result was also consistent with the previous studies (13, 21).

In addition, we found that [^{68}Ga]Ga-FAPI-04 uptake was significantly higher than [^{18}F]F-FDG uptake in the uterus, which may be attributed to the presence of active fibroblasts in the uterus (30); this suggests that the SUV_{max} of lesions (primary or/and metastatic lesions) located in the pelvic area is affected by high [^{68}Ga]Ga-FAPI-04 uptake, a potential limitation of [^{68}Ga]Ga-FAPI-04 PET/CT imaging.

This study has several limitations. First, the cohort size was small, and the number of participants with brain and bone metastases was low. Second, the cohort was heterogeneous and included participants with different treatment statuses, which could have affected the SUV_{max} values of lesions. Third, the period of follow-up was set to 3–6 months based on previous experience. Although most of lesions could be observed obvious changes indicating their benign or malignant features, a few lesions might be taken longer to be followed up. Lastly, we could not pathologically confirm all suspicious lesions without considering ethics; thus, neither accurate sensitivity nor specificity could be established. Future studies with larger and homogeneous cohorts are warranted to provide a more comprehensive analysis of the utility of [^{68}Ga]Ga-FAPI-04 PET/CT in CRC.

5 Conclusion

[^{68}Ga]Ga-FAPI-04 PET/CT has several obvious advantages over [^{18}F]F-FDG PET/CT, including the detection of lymph nodes and distant metastases, thereby improving the staging of patients with CRC. This improved staging is helpful for the timely revision of clinical treatment strategies and improvement of patients' prognoses. Additionally, patients might feel more comfortable during [^{68}Ga]Ga-FAPI-04 PET/CT as it does not require fasting.

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 Institutional Review Board of Beijing Cancer Hospital (no. 2019KT95). The patients/participants provided their written informed consent to participate in this study.

Author contributions

XL, XW, ZY, and AW contributed to the study conception and design. Synthesis of tracer and image acquisition were performed by SW and YZ. XL, YL, XC, MW, and HZ processed and analysed the data. The first draft of the manuscript was written by XL. XW, ZY, and AW reviewed and revised the manuscript. All authors read and approved the final manuscript.

Funding

This work was financially supported by the National Natural Science Foundation of China (no. 82071957 and 82171980), Beijing Hospitals Authority Clinical Medicine Development of special funding support (code: XMLX202120), Beijing Hospitals Authority Dengfeng Project (DFL20191102).

Acknowledgments

The authors gratefully acknowledge all participants.

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.1087792/full#supplementary-material

References

1. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. *CA Cancer J Clin* (2022) 72:7–33. doi: 10.3322/caac.21708

2. Dekker E, Tanis PJ, Vleugels JLA, Kasi PM, Wallace MB. Colorectal cancer. *Lancet* (2019) 394:1467–80. doi: 10.1016/S0140-6736(19)32319-0

3. Leah HB, Deborah S. Diagnosis and treatment of metastatic colorectal cancer: A review. *JAMA* (2021) 325:669–85. doi: 10.1001/jama.2021.0106

4. Kantorova I, Lipska L, Be O, Schneiderova M. Routine ^{18}F-FDG PET preoperative staging of colorectal cancer: Comparison with conventional staging and its impact on treatment decision making. *J Nucl Med* (2003) 44:1784–8.

5. Sobhani I, Itti E, Luciani A, Baumgaertner I, Layese R, André T, et al. Colorectal cancer (CRC) monitoring by 6-monthly ^{18}FDG-PET/CT: an open-label multicentre randomised trial. *Ann Oncol* (2018) 29:931–7. doi: 10.1093/annonc/mdy031

6. Berger KL, Nicholson SA, Dehdashti F, Siegel BA. FDG PET evaluation of mucinous neoplasms: Correlation of FDG uptake with histopathologic features. *Am J Roentgenol* (2000) 174:1005–8. doi: 10.2214/ajr.174.4.1741005

7. Borello A, Russolillo N, Lo Tesoriere R, Langella S, Guerra M, Ferrero A. Diagnostic performance of the FDG-PET/CT in patients with resected mucinous colorectal liver metastases. *Surgeon* (2021) 19:e140–145. doi: 10.1016/j.surge.2020.09.004

8. Lindner T, Loktev A, Altmann A, Giesel F, Kratochwil C, Debus J, et al. Development of quinoline-based theranostic ligands for the targeting of fibroblast activation protein. *J Nucl Med* (2018) 59:1415–22. doi: 10.2967/jnumed.118.210443

9. Chen H, Pang Y, Wu J, Zhao L, Hao B, Wu J, et al. Comparison of [^{68}Ga]Ga-DOTA-FAPI-04 and [^{18}F] FDG PET/CT for the diagnosis of primary and metastatic lesions in patients with various types of cancer. *Eur J Nucl Med Mol Imaging* (2020) 47:1820–32. doi: 10.1007/s00259-020-04769-z

10. Loktev A, Lindner T, Burger EM, Altmann A, Giesel F, Kratochwil C, et al. Development of fibroblast activation protein–targeted radiotracers with improved tumor retention. *J Nucl Med* (2019) 60:1421–9. doi: 10.2967/jnumed.118.224469

11. Chen H, Zhao L, Ruan D, Pang Y, Hao B, Dai Y, et al. Usefulness of [^{68}Ga]Ga-DOTA-FAPI-04 PET/CT in patients presenting with inconclusive [^{18}F]FDG PET/CT findings. *Eur J Nucl Med Mol Imaging* (2021) 48:73–86. doi: 10.1007/s00259-020-04940-6

12. Koerber SA, Staudinger F, Kratochwil C, Adeberg S, Haefner M, Ungerechts G, et al. The role of ^{68}Ga-FAPI PET/CT for patients with malignancies of the lower gastrointestinal tract: First clinical experience. *J Nucl Med* (2020) 61:1331–6. doi: 10.2967/jnumed.119.237016

13. Pang Y, Zhao L, Luo Z, Hao B, Wu H, Lin Q, et al. Comparison of ^{68}Ga-FAPI and ^{18}F-FDG uptake in gastric, duodenal, and colorectal cancers. *Radiology* (2021) 298:393–402. doi: 10.1148/radiol.2020203275

14. Jin X, Wei M, Wang S, Wang G, Lai Y, Shi Y, et al. Detecting fibroblast activation proteins in lymphoma using ^{68}Ga-FAPI PET/CT. *J Nucl Med* (2022) 63:212–7. doi: 10.2967/jnumed.121.262134

15. Benson AB, Venook AP, Al-Hawary MM, Arain MA, Chen YJ, Ciombor KK, et al. NCCN guidelines insights: Rectal cancer, version 6.2020: Featured updates to the NCCN guidelines. *J Natl Comp Cancer Netw* (2020) 18:806–15. doi: 10.6004/jnccn.2020.0032

16. Benson AB, Venook AP, Al-Hawary MM, Arain MA, Chen YJ, Ciombor KK, et al. Colon cancer, version 2.2021, NCCN clinical practice guidelines in oncology. *J Natl Comp Cancer Netw* (2021) 19:329–59. doi: 10.6004/jnccn.2021.0012

17. Kratochwil C, Flechsig P, Lindner T, Abderrahim L, Altmann A, Mier W, et al. ^{68}Ga-FAPI PET/CT: tracer uptake in 28 different kinds of cancer. *J Nucl Med* (2019) 60:801–5. doi: 10.2967/jnumed.119.227967

18. Mona CE, Benz MR, Hikmat F, Grogan TR, Lueckerath K, Razmaria A, et al. Correlation of ^{68}Ga-FAPi-46 PET biodistribution with FAP expression by immunohistochemistry in patients with solid cancers: A prospective translational exploratory study. *J Nucl Med* (2021) 63:1021–26. doi: 10.2967/jnumed.121.262426

19. Barbaro B, Leccisotti L, Vecchio FM, Di Matteo M, Serra T, Salsano M, et al. The potential predictive value of MRI and PET-CT in mucinous and nonmucinous rectal cancer to identify patients at high risk of metastatic disease. *Br J Radiol* (2017) 90:20150836. doi: 10.1259/bjr.20150836

20. Solano-Iturri JD, Beitia M, Errarte P, Calvete-Candenas J, Etxezarraga M, Loizate A, et al. Altered expression of fibroblast activation protein-α (FAP) in colorectal adenoma-carcinoma sequence and in lymph node and liver metastases. *Aging* (2020) 12:10337–58. doi: 10.18632/aging.103261

21. Kömek H, Can C, Kaplan İ, Gündoğan C, Kepenek F, Karaoglan H, et al. Comparison of [^{68}Ga]Ga-DOTA-FAPI-04 PET/CT and [^{18}F]FDG PET/CT in colorectal cancer. *Eur J Nucl Med Mol Imaging* (2022) 49:3898–909. doi: 10.1007/s00259-022-05839-0

22. Yang Z, Liu Z. The efficacy of ^{18}F-FDG PET/CT-based diagnostic model in the diagnosis of colorectal cancer regional lymph node metastasis. *Saudi J Biol Sci* (2020) 27:805–11. doi: 10.1016/j.sjbs.2019.12.017

23. Elekonawo FMK, Starremans B, Laurens ST, Bremers AJA, de Wilt JHW, Heijmen L, et al. Can [^{18}F]F-FDG PET/CT be used to assess the pre-operative extent of peritoneal carcinomatosis in patients with colorectal cancer? *Abdom Radiol* (2020) 45:301–6. doi: 10.1007/s00261-019-02268-w

24. Tsunoda Y, Ito M, Fujii H, Kuwano H, Saito N. Preoperative diagnosis of lymph node metastases of colorectal cancer by FDG-PET/CT. *Jpn J Clin Oncol* (2008) 38:347–53. doi: 10.1093/jjco/hyn032

25. Chen R, Wang Y, Zhou X, Huang G, Liu J. Preoperative PET/CT ^{18}F-FDG standardized uptake by lymph nodes as a significant prognostic factor in patients with colorectal cancer. *Contrast Media Mol Imaging* (2018) 2018:1–7. doi: 10.1155/2018/5802109

26. Sugai T, Yamada N, Osakabe M, Hashimoto M, Uesugi N, Eizuka M, et al. Microenvironmental markers are correlated with lymph node metastasis in invasive submucosal colorectal cancer. *Histopathology* (2021) 79:584–98. doi: 10.1111/his.14388

27. Daza JF, Solis NM, Parpia S, Gallinger S, Moulton CA, Belley-Cote EP, et al. A meta-analysis exploring the role of PET and PET-CT in the management of potentially resectable colorectal cancer liver metastases. *Eur J Surg Oncol* (2019) 45:1341–8. doi: 10.1016/j.ejso.2019.03.025

28. Koumpa FS, Xylas D, Konopka M, Galea D, Veselkov K, Antoniou A, et al. Colorectal peritoneal metastases: A systematic review of current and emerging trends in clinical and translational research. *Gastroenterol Res Pract* (2019) 2019:5180895. doi: 10.1155/2019/5180895

29. Chouillard E, Greco V, Tsiminikakis N. Peritoneal carcinomatosis of colorectal origin: Is it really an end-stage disease? *Tech Coloproctol* (2013) 17:619–21. doi: 10.1007/s10151-013-1063-2

30. Karpathiou G, Chauleur C, Dridi M, Baillard P, Corsini T, Dumollard JM, et al. Histologic findings of uterine niches. *Am J Clin Pathol* (2020) 154:645–55. doi: 10.1093/ajcp/aqaa080

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 13 January 2023
DOI 10.3389/fonc.2022.1073983

OPEN ACCESS

EDITED BY
Jasper Nijkamp,
Aarhus University, Denmark

REVIEWED BY
Wei Wei,
Xi'an Polytechnic University, China
Niu Jinliang,
Second Hospital of Shanxi Medical University, China

*CORRESPONDENCE
Xin Zhao
zhaoxinxct@vip.163.com

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

RECEIVED 19 October 2022
ACCEPTED 28 December 2022
PUBLISHED 13 January 2023

CITATION
Cheng M, Tan S, Ren T, Zhu Z, Wang K, Zhang L, Meng L, Yang X, Pan T, Yang Z and Zhao X (2023) Magnetic resonance imaging radiomics to differentiate ovarian sex cord-stromal tumors and primary epithelial ovarian cancers.
Front. Oncol. 12:1073983.
doi: 10.3389/fonc.2022.1073983

Magnetic resonance imaging radiomics to differentiate ovarian sex cord-stromal tumors and primary epithelial ovarian cancers

Meiying Cheng[1], Shifang Tan[1], Tian Ren[2], Zitao Zhu[3], Kaiyu Wang[4], Lingjie Zhang[1], Lingsong Meng[1], Xuhong Yang[5], Teng Pan[6], Zhexuan Yang[1] and Xin Zhao[1]*

[1]Department of Radiology, Third Affiliated Hospital of Zhengzhou University, Zhengzhou, Henan, China, [2]Department of Information, Third Affiliated Hospital of Zhengzhou University, Zhengzhou, Henan, China, [3]Medical College, Wuhan University, Wuhan, China, [4]Magnetic resonance imaging (MRI) Research, GE Healthcare (China), Beijing, China, [5]Department of Research, Huiying Medical Technology Co., Ltd., Beijing, China, [6]Department of Research, Third Affiliated Hospital of Zhengzhou University, Zhengzhou, Beijing, China

Objective: To evaluate the diagnostic ability of magnetic resonance imaging (MRI) based radiomics and traditional characteristics to differentiate between Ovarian sex cord-stromal tumors (SCSTs) and epithelial ovarian cancers (EOCs).

Methods: We consecutively included a total of 148 patients with 173 tumors (81 SCSTs in 73 patients and 92 EOCs in 75 patients), who were randomly divided into development and testing cohorts at a ratio of 8:2. Radiomics features were extracted from each tumor, 5-fold cross-validation was conducted for the selection of stable features based on development cohort, and we built radiomics model based on these selected features. Univariate and multivariate analyses were used to identify the independent predictors in clinical features and conventional MR parameters for differentiating SCSTs and EOCs. And nomogram was used to visualized the ultimately predictive models. All models were constructed based on the logistic regression (LR) classifier. The performance of each model was evaluated by the receiver operating characteristic (ROC) curve. Calibration and decision curves analysis (DCA) were used to evaluate the performance of models.

Results: The final radiomics model was constructed by nine radiomics features, which exhibited superior predictive ability with AUCs of 0.915 (95%CI: 0.869-0.962) and 0.867 (95%CI: 0.732-1.000) in the development and testing cohorts, respectively. The mixed model which combining the radiomics signatures and traditional parameters achieved the best performance, with AUCs of 0.934 (95%CI: 0.892-0.976) and 0.875 (95%CI: 0.743-1.000) in the development and testing cohorts, respectively.

Conclusion: We believe that the radiomics approach could be a more objective and accurate way to distinguish between SCSTs and EOCs, and the mixed model developed in our study could provide a comprehensive, effective method for clinicians to develop an appropriate management strategy.

KEYWORDS

ovarian sex cord-stromal tumor, epithelial ovarian cancer, magnetic resonance imaging, radiomics, prediction model

Introduction

Ovarian sex cord-stromal tumors (SCSTs) are rare nonepithelial neoplasms, and represent about 7% of all primary ovarian tumors (1, 2). According to the classification of ovarian tumors by World Health Organization (WHO) (2020), SCSTs are divided into the following three clinicopathologic subcategories: pure stromal tumors, pure sex cord tumors, and mixed sex cord-stromal tumors (3). Fibromas, thecomas, and granulosa cell tumors account for the majority of SCSTs.

Morphologically, SCSTs usually present as solid masses (4), resembling malignant tumors. Although some of the SCSTs, such as fibromas, have a few specific characteristics on conventional imaging features, diagnostic dilemmas may often arise if the tumor shows increased cellularity, or due to necrosis, hemorrhage, edema, or cystic degeneration (2, 4). Other types are even more confusing on account of morphological complexity. Moreover, the rarity of SCSTs contributes to a low degree of suspicion, which makes it susceptible to misdiagnosis as the more common epithelial ovarian cancer (EOC). Clinically, SCSTs most commonly present at early stages (I) and are primarily surgically treated with an overall favorable prognosis (2, 5), while epithelial tumors usually present at advanced stages (III or IV) and are treated with chemotherapy and surgical debulking (2). Hence, a more accurate and objective assessment method to identify SCSTs from ovarian cancers is imperative.

Magnetic resonance imaging (MRI) has been widely used to detect and evaluate adnexal lesions, especially for the indeterminate adnexal masses on ultrasonography (6, 7). The high soft-tissue resolution and ability to characterize the composition of different fluid types allow it to characterize the lesion types more accurately (7). However, all the evaluation by MRI requires the subjective interpretation of radiologists, in addition to the complexity and overlap in imaging characteristics of different diseases, it is still a challenge to differentiate the types of tumors with MRI alone (6).

Radiomics is an emerging method for postprocessing any type of medical image and generating new quantification metrics which have proven to provide important insights into tumor biology, shifting radiology from the traditional visual analysis to a more objective and automated analysis (6). Recently, radiomics studies on adnexal tumors have demonstrated some encouraging advances, including clinical outcomes in ovarian cancer (6, 8, 9), tumor category (9, 10), and subtype differentiation (9, 11). However, there are few studies involving SCSTs. In this study, we developed an MR image-based radiomics diagnostic model to prove that SCSTs and EOCs are separable.

Patients and methods

Patient population

This retrospective pilot study was approved by our Institutional Review Board with a waiver of informed consent. From January 2017 to December 2021, we retrospectively retrieved all MRI pelvic examinations referring to ovarian or adnexal lesions from our Institutional Picture Archiving and Communication System (PACS) and obtained 1867 results. Every report was analyzed by the researchers. Then we identified patients who met the following inclusion criteria (1): pathologically confirmed as SCSTs or EOCs; (2) MRI was performed within 1 month before pelvic or laparoscopic surgery at our institution; (3) no chemotherapy, radiotherapy, or previous gynecological operation prior to MRI examination. The exclusion criteria were: (1) poor MR image quality, so that the focus cannot be clearly observed or delineated; (2) multiple nodules that the primary focus cannot be identified due to a mutual fusion and extensive adhesion pattern or any other reason. Finally, we consecutively reviewed a total of 148 patients with 173 tumors (81 SCSTs in 73 patients and 92 EOCs in 75 patients) as the primary cohort. Details of histopathology are presented in Table 1. Then the primary dataset was randomly split into the development and testing cohorts with a fixed ratio of 8:2 in each category, resulting in 137 tumors (64 SCSTs and 73 EOCs) for the development cohort and 36 tumors (17 SCSTs and 19 EOCs) for testing cohort. A flowchart of the patient selection process is shown in Figure 1.

Clinical characteristics such as patient age, menstrual status, endocrine level, cancer antigen 125 (CA125), and risk of ovarian malignancy algorithm (ROMA), were obtained from the hospital information system.

MR image acquisition and interpretation

MR examinations were performed on the 3.0 T system (SIGNA Pioneer, GE Healthcare, and Skyra, Siemens Healthcare) with the phased-array abdominal coil. The conventional MR sequences included T1 weighted imaging (T1WI) in the axial plane, T2 weighted imaging (T2WI) in the axial and sagittal plane, fat-suppressed T2WI (FS-T2WI) in the axial and coronal plane, diffusion-weighted imaging (DWI) with the b value of 1000 s/mm^2 in the axial plane, and multiphase contrast-enhanced fat-suppressed T1WI in the axial, sagittal, and coronal plane. Detailed information about the acquisition parameters is presented in Table S1.

Two radiologists (reader 1 and reader 2, with 7 and 15 years of experience in MRI, respectively) independently recorded the conventional imaging features while blinded to the histological results. The recorded features include (1) Maximum diameter (MD) (measured at the lesion slice with the maximum diameter of the tumor in the three-dimensional measurements); (2) Visibility of hemorrhagic component; (3) Solid and cystic components (predominantly cystic, cystic-solid, and predominantly solid corresponding to less than 1/3, 1/3-2/3, and more than 2/3 solid component, respectively); (4) Signal intensity (SI) of the solid components on T2WI (Hypo-, iso-, or hyperintense was relative to the external myometrium; A few purely cystic lesions were not recorded); (5) Apparent diffusion coefficient (ADC) value (measured manually on the DWI derived ADC maps; regions of interest were placed at target areas of the tumor, and areas such as necrosis, hemorrhage, vascular structures were avoided as much as possible; three measurements were obtained and averaged). Discrepancies were resolved by a consensus, or a third radiologist (reader 3 with more than 20 years of experience in gynecologic imaging) would serve as an arbitrator.

TABLE 1 Histopathological types of the selected samples.

Pathological type	Numbers
SCST	81
Thecoma	4
Fibroma	18
Cellular fibroma	2
Fibrothecoma	28
Granulosa cell tumor	25
Sclerosing stromal tumor	2
Sertoli-Leydig cell tumor	2
EOC	92
High grade serous carcinoma	42
Low grade serous carcinoma	17
Endometrioid carcinoma	12
Clear cell carcinoma	12
Mucinous carcinoma	6
Seromucinous carcinoma	3
Total	173

SCST, sex cord-stromal tumors; EOC, epithelial ovarian cancer

Tumor segmentation

In order to reduce the discrepancies related to various scanning parameters and eliminate the internal dependence of image radiomics features on voxel size, all images were set to a fixed voxel size (1mm×1mm×1mm) with linear interpolation algorithm.

The 3D segmentation was performed on open-source software (ITK-SNAP version 3.8.0, http://www.itksnap.org). The volume of interest (VOI) for each lesion on each slice was manually delineated on FS-T2WI by a reader 1. Then, reader 3 confirmed all the regions. To examine the reproducibility of extracted radiomics features and to obtain more robust radiomics features, the intra-class correlation coefficient (ICC) was used to assess the intra- and inter-observer agreement of VOI delineation. So, four weeks later, 30 patients (15 SCSTs and 15 EOCs) were randomly selected. Reader 1 then re-delineated the VOI for intra-consistency testing, in the meantime, reader 2 outlined the VOI according to the same procedure to test inter-consistency. The feature with an ICC > 0.75 was selected for further analysis (12). ICC can be obtained from the following equation:

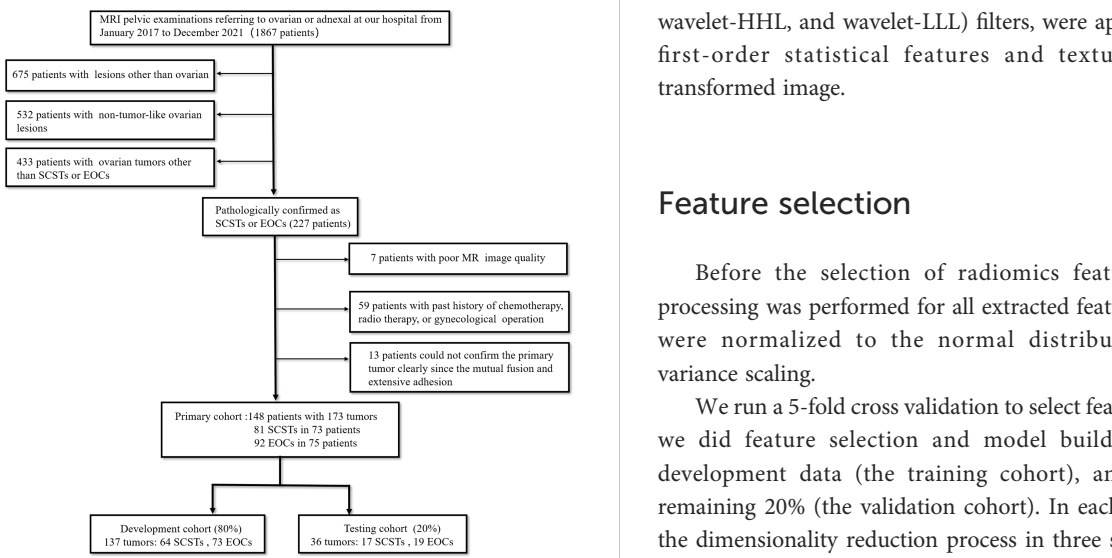


FIGURE 1
Flowchart for the patient selection process in this study.

$$ICC = \frac{(MS_R - MS_E)}{MS_R + (\frac{MS_C - MS_E}{n})} \quad ,$$

where MS_R is mean square for rows, MS_C is the mean square for columns, MS_E is mean square for error and n represents the number of subjects.

Feature extraction

The extraction of radiomics features was conducted in the Radcloud software (Huiying Medical Technology Co., Ltd, Beijing, China). A total of 1,409 features were extracted from the FS-T2WI sequences using the pyradiomics function package (https://pyradiomics.readthedocs.io/). These features could be divided into the following three categories. First, first-order statistical features, such as peak value, mean, and variance, quantitatively describe the voxel intensity distribution of the lesion area in MR images through common basic indicators. Second, two-dimensional morphological features, describe the two-dimensional shape and size of the lesion. Third, texture features, such as Gray Level Co-occurrence Matrix (GLCM), Gray Level Run Length Matrix (GLCM), and Gray Level Size Zone Matrix (GLSZM), quantify the heterogeneity of the lesion texture. In addition, several filters, such as exponential, logarithm, square, square root, and wavelet (including wavelet-LHL, wavelet-LHH, wavelet-HLL, wavelet-LLH, wavelet-HLH, wavelet-HHH, wavelet-HHL, and wavelet-LLL) filters, were applied to calculate the first-order statistical features and texture features of the transformed image.

Feature selection

Before the selection of radiomics features, normalization processing was performed for all extracted features, and the features were normalized to the normal distribution by mean and variance scaling.

We run a 5-fold cross validation to select features, where each fold we did feature selection and model building on 80% of the development data (the training cohort), and evaluated on the remaining 20% (the validation cohort). In each fold, we implement the dimensionality reduction process in three steps. Features with a variance value of >0.8 were first selected. Then SelectKBest was applied to select the features with a *p*-value less than 0.05. Finally, the Lassolars algorithm was used to screen the optimal radiomics

features, among which the top 10 features with the best contribution would be selected. The least absolute shrinkage and selection operator (LASSO) is a regression analysis method that can perform both variable selection and regularization to improve the identification accuracy and interpretability of the model. Regression algorithm least angle regression (LARS) provides variables through the linear combination of high-dimensional data. It is related to positive stepwise regression. Lassolars algorithm used in this paper is a combination of Lars algorithm and lasso model, which can automatically select the optimal parameters λ and have better performance than LASSO algorithm alone (13). Finally, we counted the frequency of each feature selected in the 5 folds, the features that get selected three or more times repeatedly were considered stable features, and different models were built using various combinations of those selected features.

Classifier modeling

Univariate analysis and multivariate logistic regression analysis were successively performed based on development data to screen the clinical and conventional MR features. Based on the selected radiomics features, clinical features, and conventional MR parameters, five prediction models (clinical model, conventional MR model, traditional model, radiomics model and mixed model) were constructed by using the LR classifier. LR classifier includes classification, function establishment, solving optimal model parameters through optimization iteration, and verifying the model performance. In addition, the nomogram was also constructed to visualize the results of the logistic regression. The nomogram develops scoring criteria based on the magnitude of the regression coefficients for all independent variables. It scores each value level of each independent variable, giving an overall score and finally calculating the probability of disease risk for each patient through a conversion function between the score and the outcome probability. The calibration curve was drawn to assess the agreement between the predicted results and actual presence. The detailed process of radiomics analysis is presented in Figure 2.

The performance of the model was estimated by using receiver operating characteristic (ROC) curves and confusion matrix analysis, and the area under the curve (AUC), sensitivity, specificity, and accuracy were calculated. The DeLong test was used to compare the performance of different models. The decision curve analysis (DCA) was used to evaluate the clinical utility of the prediction models by quantifying the net benefits at different threshold probabilities in the dataset.

Statistical analysis

Statistical analysis was performed using SPSS version 25.0 (IBM). Quantitative variables are shown as mean ± standard deviation. Categorical variables were assessed by Chi-square tests or Fisher exact test, and differences in continuous variables were assessed by t-test or Mann-Whitney U test. $P< 0.05$ was considered statistically significant. Model construction was executed using R software 3.5.3 (https://cran.r-project.org/) and Python 2.7 software (https://www.python.org/). The packages of "pyradiomics" (https://pyradiomics.readthedocs.io/), "scikitlearn" (https://scikit-learn.org/), and "matplotlib" (https://matplotlib.org/) were used for feature selection, model building, and plotting in this study.

Results

Patients and tumor characteristics

The comparisons of the clinical data and MR parameters between SCSTs and EOCs groups in the primary and development cohorts are summarized in Table 2. Age, ROMA index, serum CA125, ADC value, MD, SI on T2WI, solid and cystic components showed significant differences between the SCSTs and EOCs groups, while no significant difference was observed in menstrual status, endocrine level, and hemorrhagic component. Among the variables with significant differences, only ROMA index ($P< 0.001$), ADC value ($P = 0.004$), solid and cystic components ($P = 0.043$) remained as independent predictors on the multivariate logistic regression analysis.

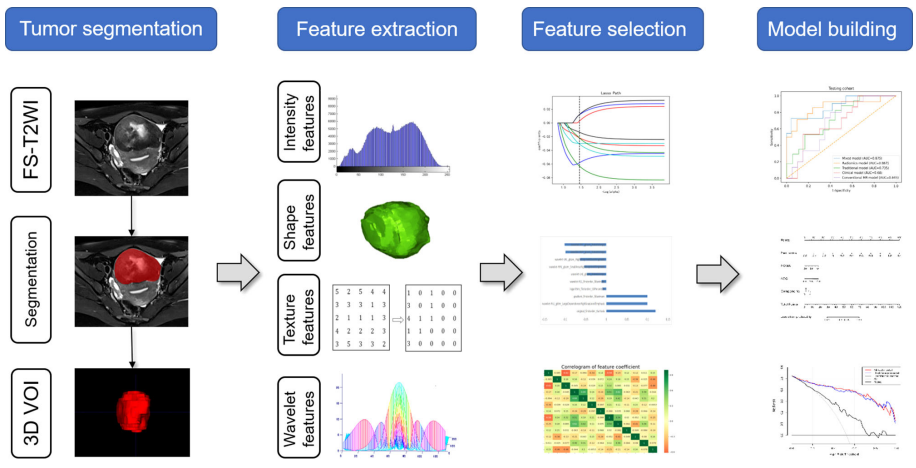


FIGURE 2
Process of radiomics analysis in this study.

TABLE 2 The clinical data and MR parameters of the primary and development cohorts.

Characteristics	Primary cohort			Development cohort		
	EOCs (n = 92)	SCSTs (n = 81)	*P* value	EOCs (n = 73)	SCSTs (n = 64)	*P* value
Age (year) [a]	51.78 ± 10.35	46.33 ± 15.76	0.017*	52.22 ± 10.41	46.58 ± 16.16	0.015*
ROMA (%) [a]	32.07 ± 26.71	16.03 ± 12.20	<.001*	33.65 ± 27.32	16.08 ± 12.91	<.001*
CA125 (μg/L) [b]			<.001*			<.001*
<35	38 (41.3%)	55 (67.9%)		27 (37.0%)	46 (71.9%)	
35-200	26 (28.3%)	23 (28.4%)		23 (31.5%)	15 (23.4%)	
200-500	13 (14.1%)	1 (1.2%)		10 (13.7%)	1 (1.6%)	
>500	15 (16.3%)	2 (2.5%)		13 (17.8%)	2 (3.1%)	
Menstrual status [b]			0.543			0.205
Premenopausal	45 (48.9%)	44 (54.3%)		32 (43.8%)	35 (54.7%)	
Postmenopausal	47 (51.1%)	37 (45.7%)		41 (56.2%)	29 (45.3%)	
Endocrine level [b]			0.788			0.588
Normal	85 (92.4%)	73 (90.1%)		67 (91.8%)	57 (89.1%)	
Abnormal	7 (7.6%)	8 (9.9%)		6 (8.2%)	7 (10.9%)	
ADC ($\times10^{-3}mm^2/s$) [a]	0.98 ± 0.26	1.13 ± 0.34	0.001*	1.00 ± 0.28	1.12 ± 0.34	0.023*
MD (cm) [a]	8.72 ± 4.21	7.03 ± 4.28	0.001*	8.95 ± 4.42	7.31 ± 4.41	0.032*
SI on T2WI [b]			0.002*			0.003*
Hypo-intensity	5 (5.4%)	19 (23.5%)		2 (2.7%)	14 (21.9%)	
Iso-intensity	40 (43.5%)	20 (24.7%)		34 (46.6%)	17 (26.6%)	
Hyperintensity	30 (32.6%)	23 (28.4%)		23 (31.5%)	17 (26.6%)	
Mixed	11 (12%)	15 (18.5%)		9 (12.3%)	13 (20.3%)	
Solid and cystic components [b]			0.004*			0.002*
Predominantly cystic	30 (32.6%)	12 (14.8%)		26 (35.6%)	10 (15.6%)	
Cystic-solid	21 (22.8%)	13 (16.1%)		17 (23.3%)	9 (14.1%)	
Predominantly solid	41 (44.6%)	56 (69.1%)		30 (41.4%)	45 (70.3%)	
Hemorrhage [b]			0.188			0.154
Present	16 (17.4%)	8 (9.9%)		13 (17.8%)	6 (9.4%)	
Absent	76 (82.6%)	73 (90.1%)		60 (82.2%)	58 (90.6%)	

EOC, epithelial ovarian cancer; SCST, sex cord-stromal tumors; ROMA, Risk of Ovarian Malignancy Algorithm; CA125, Cancer antigen 125; ADC, apparent diffusion coefficient; MD, maximum diameter; SI, signal intensity; T2WI, T2 weighted imaging; [a] Data are the mean ± standard deviation, P values calculated by sample t test. [b] Date are the case (%), P values calculated by Chi-square tests or Fisher exact test. * P values< 0.05 were considered statistically significant.

Feature extraction and selection

Of all the extracted radiomics features, the median ICC was 0.90, and 1277 of 1409 features (91%) were robust and were selected for subsequent analysis, with ICC > 0.75. Lassolars algorithms on feature selection for each fold are shown in Figure S1. The selected radiomic features and the corresponding coefficients in each fold are shown in Figure 3 and Table S2. We did model building in every fold, the AUCs for the 5-fold cross validation are reported in Table S3 and Figure S2, and the mean AUCs were 0.883 ± 0.018 in the training cohort and 0.849 ± 0.021 in the validation cohort. Finally, there were three features got selected in 3 of the 5 folds, three features got selected in 4 of the 5 folds, and three features got selected in every fold. The frequencies of the radiomics features are summarized in Table 3.

Construction and performance of the prediction models

Based on the selected clinical variable ROMA index, a clinical model was established. Based on parameter ADC, solid and cystic components, a conventional MR model was established. Then, based on the combination of the above clinical factors and conventional MR parameters, a traditional model was established. The AUCs of the

10.3389/fonc.2022.1073983

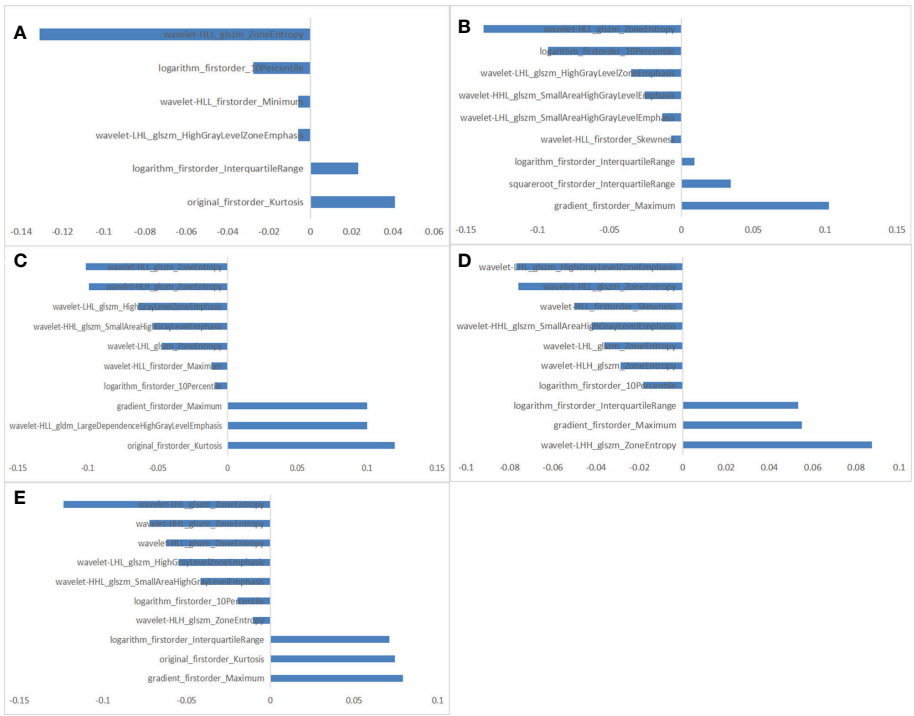


FIGURE 3
The selected radiomic features in each fold. **(A)** 6 radiomic features selected in fold 1. **(B)** 9 radiomic features selected in fold 2. **(C)** 10 radiomic features selected in fold 3. **(D)** 10 radiomic features selected in fold 4. **(E)** 10 radiomic features selected in fold 5.

clinical model, conventional MR model, and traditional model were 0.729 (95%CI: 0.645-0.812), 0.737 (95%CI: 0.654-0.819), 0.755 (95% CI: 0.677-0.834) in the development cohort, respectively, and 0.680 (95%CI: 0.498-0.862), 0.693 (95%CI: 0.515-0.872), 0.735 (95%CI: 0.569-0.902) in the testing cohort, respectively (Figure 4 and Table 4).

Three different radiomics models (Models 1-3) were developed using the following combinations of the radiomics features: 3 features (got selected in every fold), 6 features (got selected 4 or more times repeatedly) and 9 features (got selected 3 or more times repeatedly). Among them, Model 3 showed the best performance, which was determined as the final radiomics model, with the AUCs of 0.915 (95% CI: 0.869-0.962) and 0.867 (95% CI: 0.732-1.000) in the development and testing cohorts, respectively (Figure 4 and Table 4). However, no significant differences were observed between Model 3 and the other two models. The AUCs of the three models and the comparisons in terms of diagnostic performance among them were shown in Table 5. Rad-score based on 9 features were weighted by their respective coefficients, the calculation formula for the Rad-score is provided in the Supplementary Materials page 7.

Finally, we established a mixed model based on the Rad-score, clinical characteristics (ROMA), and conventional MR parameters (ADC, solid and cystic components). The AUCs of the mixed model were 0.934 (95% CI: 0.892-0.976) and 0.875 (95% CI: 0.743-1.000) in the development and testing cohorts, respectively (Figure 4 and Table 4).

TABLE 3 Details of the selected features in the developed radiomics models.

Number	Radiomics feature name	Frequency
1	wavelet-HLL-GLSZM-ZoneEntropy	5
2	wavelet-LHL-GLSZM-HighGrayLevelZoneEmphasis	5
3	logarithm-firstorder-10Percentile	5
4	logarithm-firstorder-InterquartileRange	4
5	gradient-firstorder-Maximum	4
6	wavelet-HL-GLSZM-SmallAreaHighGrayLevelEmphasis	4
7	original-firstorder-Kurtosis	3
8	wavelet-HLH- GLSZM -ZoneEntropy	3
9	wavelet-LHL- GLSZM -ZoneEntropy	3
GLSZM, Gray-Level Size Zone Matrix		

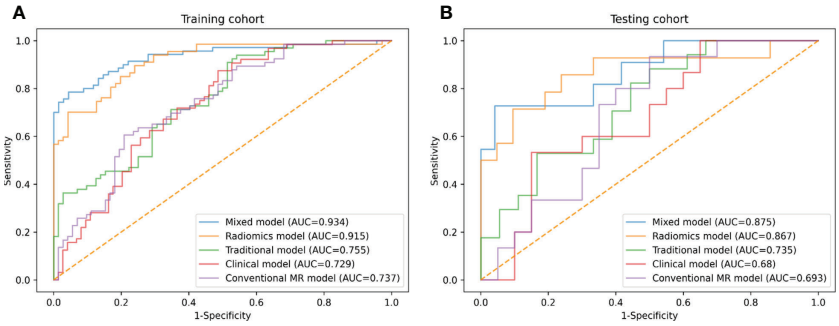


FIGURE 4
ROC curves of conventional MR model, clinical model, traditional model, radiomics model and mixed model within the development **(A)** and testing **(B)** cohorts.

TABLE 4 Predictive performances of the final models in the development and testing cohorts.

Model	Development cohort					Testing cohort				
	AUC (95% CI)	SEN	SPE	ACC	*P* value	AUC (95% CI)	SEN	SPE	ACC	*P* value
Conventional MR model	0.737 (0.654-0.819)	0.758	0.528	0.638	<0.001	0.693 (0.515-0.872)	0.600	0.650	0.629	0.039
Clinical model	0.729 (0.645-0.812)	0.812	0.541	0.667	<0.001	0.680 (0.498-0.862)	0.600	0.650	0.629	0.048
Traditional model	0.755 (0.677-0.834)	0.727	0.542	0.630	<0.001	0.735 (0.569-0.902)	0.706	0.611	0.657	0.017
Radiomics model	0.915 0.869-0.962)	0.701	0.873	0.790	<0.001	0.867 (0.732-1.000)	0.714	0.857	0.800	<0.001
Mixed model	0.934 (0.892-0.976)	0.814	0.868	0.841	<0.001	0.875 (0.743-1.000)	0.727	0.750	0.743	<0.001

AUC, area under curve; SEN, sensitivity; SPE, specificity; ACC, accuracy; p values on the AUCs are difference from AUC=0.5

TABLE 5 Predictive performances of the three developed radiomics models.

Model	Development cohort					Testing cohort				
	AUC (95% CI)	SEN	SPE	ACC	*P* value	AUC (95% CI)	SEN	SPE	ACC	*P* value
Model 1	0.837 (0.765-0.909)	0.672	0.838	0.761	0.073	0.817 (0.662-0.971)	0.800	0.55	0.657	0.631
Model 2	0.873 (0.814-0.932)	0.683	0.853	0.775	0.269	0.846 (0.707-0.985)	0.769	0.818	0.800	0.832
Model 3	0.915 (0.869-0.962)	0.701	0.873	0.790	1.000	0.867 (0.732-1.000)	0.714	0.857	0.800	1.000

AUC, area under curve; SEN, sensitivity; SPE, specificity; ACC, accuracy; P values calculated by Delong test, compared with Model 3.

The difference between the traditional model and radiomics model was statistically significant in the testing cohort (traditional model AUC *vs.* radiomics model AUC: 0.735 *vs.* 0.867, Delong test $P< 0.001$); however, there was no evidence of a difference in the testing set between the radiomics model compared with the mixed model (radiomics model AUC *vs.* mixed model AUC: 0.867 *vs.* 0.875, Delong test $P = 0.561$).

A radiomics nomogram was constructed by using the selected variables from multivariate logistic regression and Rad-score to provide a visualized outcome measure (Figure 5A). The total score for this nomogram was calculated using the formula: Nomo-score = -1.5402 + 11.4118 × Rad-score -0.0335 × ROMA + 1.3928 × ADC + 0.3532 × Components. The calibration curves demonstrated good diagnostic consistency between the predictions of the radiomics nomogram and the actual observations of the samples (Figure 5B).

DCA revealed that the radiomics model and the mixed model provided a better net benefit than the traditional model across the majority of the range of reasonable threshold probabilities (Figure 6).

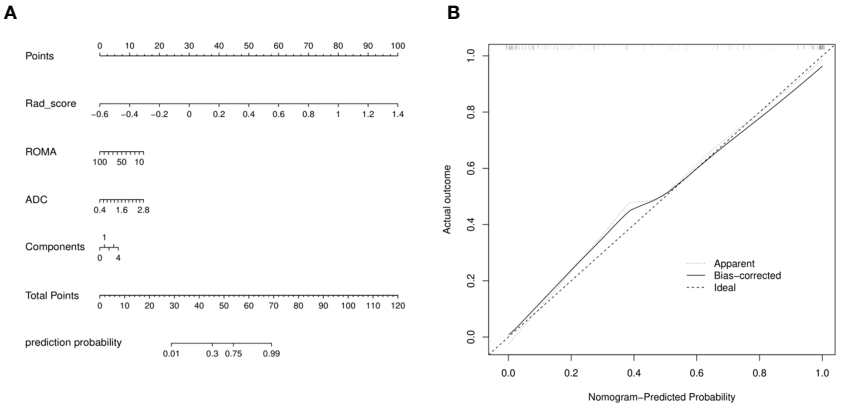


FIGURE 5
(A) Radiomics nomogram with Rad-score, two conventional MR findings and a clinical factor, including components, ADC, and ROMA. **(B)** Calibration curves of radiomics nomogram. The diagonal line represented the perfect prediction of the radiomics nomogram. The black solid line represented the calibration curve of nomogram in the testing cohorts. The calibration curves were close to the diagonal line, which indicated good prediction performance of the nomogram.

Discussion

SCSTs are composed of a heterogeneous group with complicated image manifestation. Although a few cases are mainly cystic, most SCSTs are predominantly solid masses (2, 4, 5). In MRI, the presence of solid tissue in an adnexal lesion is the primary cause of increased risk stratification (7, 14), so the particular lesion that requires to be differentiated is EOC. With the optimization of various diagnostic schemes, multiple methods have been gradually combined to evaluate adnexal tumors (15). In the current study, we assessed the ability of the clinical model, conventional MR model, traditional model, and the radiomics model to distinguish SCSTs and EOCs. As shown in the result, radiomics model yielded prominent predictive performance, significantly higher than that of the traditional model. The mixed model stood out among all the models.

Generally, SCSTs are considered to be clinically different from epithelial tumors to some extent, they occur across a wide age range and some patients may have clinical signs of hormone production (5, 16). In our study, only 8 in 73 patients with SCSTs showed elevated hormone levels, mainly presenting as a slight increase in prolactin. No significant difference was observed in menstrual status or endocrine level between the two groups. In previous studies (17, 18), serum CA125 and ROMA index have been proven to be effective diagnostic markers for EOCs, which could be used to differentiate between benign and malignant lesions. The present results were consistent with the previous studies, revealing that serum CA125 and ROMA index in the SCSTs group were significantly lower than those in EOCs, owing to the blunt nature of SCSTs. And in the multivariate logistic regression analysis, ROMA index was an independent predictor. However, the diagnostic efficacy of the clinical model was not satisfactory (AUCs = 0.729 and 0.680 in the development and testing cohorts, respectively), suggesting that clinical characteristics can only provide limited information for the differential diagnosis.

Conventional MR images could provide abundant information regarding the pathological features of tumors. Yin et al. (19) reported their results of 36 thecomas/fibrothecomas and 40 malignant pelvic

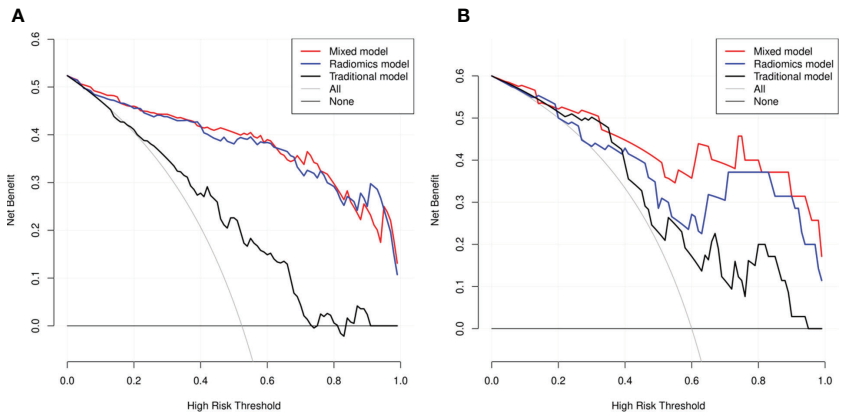


FIGURE 6
Decision curve analysis (DCA) for the three models within the development **(A)** and testing **(B)** cohorts. The net benefit versus the threshold probability was plotted. The x-axis represented the threshold probability, while the y-axis represented the net benefits. The sensitivity and specificity of the model were calculated at each threshold to determine the net benefit. The DCAs showed that the net benefits of the mixed model (black line) and the radiomics model (red line) were superior to the benefits of the traditional model (blue line) with the threshold probability range from 0 to 1.

solid tumors *via* conventional MRI and DWI examinations. They found that signal intensity on T2WI, capsule, and the lowest ADC value were important indicators in discriminating thecomas/fibrothecomas from malignant pelvic solid tumors. Their study lacked other types of sex cord tumors, and they didn't build predictive models. In our study, ADC value, solid and cystic components were independent predictors for differentiating EOCs and SCSTs. The ADC value of SCSTs was relatively higher than that of EOCs, and SCSTs were more likely to be dominated by solid masses. Conventional MR model showed that the discrimination ability of MR parameters was also limited. Moreover, after negotiation, most conventional MR parameters required subjective interpretation by the radiologist, resulting in a poor inter-observer agreement in the assessment of these features.

In this study, we used different combinations of the selected radiomics features to establish three radiomics models, all of which achieved good prediction efficiency. When combining the most efficient group of radiomics features and the traditional parameters, we found that the diagnostic performance of the final radiomics model was comparable to that of the mixed model. To date, in a few studies, radiomics have been applied to patients with adnexal tumors, such as evaluating the ability of texture features to characterize the histopathological classification of ovarian cancer (9–11) and to predict prognosis (8, 9, 20). However, SCSTs have been rarely discussed. To our knowledge, this study was the first one to establish an MR-based radiomics model focusing on the differentiation of SCSTs from others. The results are encouraging, and show great potential to improve the prediction accuracy for ovarian tumors and derive predictive imaging biomarkers. Misdiagnosis could thus be further avoided.

Radiomics extracts high-throughput quantitative data from medical images, which is helpful for disease diagnosis, staging, management, and prognostication. In the current study, we ran a 5-fold cross validation setup during the feature selection, where each fold we did feature selection and model building to provide stable features. Nine radiomics features were finally selected, mainly including first-order statistical features and Gray Level Size Zone Matrix (GLSZM). The features contain a variety of traceable image information. First-order statistics are based on histograms of the original image and describe the distribution of voxel intensities within the image region (21, 22). The GLSZM features quantify the Second-Order joint probabilities of images (21). For example, Zone Entropy (ZE) of GLSZM measures the randomness in the distribution of zone sizes and gray levels, a higher value indicates more heterogeneity in the texture patterns; High Gray Level Zone Emphasis of GLSZM measures the distribution of the higher gray-level values, with a higher value indicating a greater proportion of higher gray-level values and size zones in the image. In our study, GLSZM is a quite important set of features for differentiating SCSTs and EOCs, the intrinsic definition of these features imply that they may capture the presence of necrotic, edematous or cellular regions within the tumors (22).

To date, there is no consensus on which or how many MRI sequences should be used to establish a radiomics model. Some studies (11, 15, 23) have included several individual sequences at the same time or combined multiple sequences to establish a radiomics model for evaluation, suggesting that the multi-sequence combination model may have better performance. Only FS-T2WI sequences were used in our radiomic analysis, considering that FS-T2WI is an important and common sequence in the conventional pelvic MRI scanning protocol with the highest spatial resolution which would improve the visualization of ROIs. Sufficient predictive performance has been achieved for differentiating SCSTs and EOCs. Further research with more sequences such as DWI, ADC, and T1WI+C, would be carried out to improve our radiomics classifier.

Several limitations should be noted. First, for a radiomics study, the sample size was relatively small, so the results might be biased. Second, this study is a single-center study without external verification, so the reproducibility and generalizability of the models need to be further tested. In the future, a multicenter study with a larger dataset size should be conducted to perform an optimal radiomics analysis. Third, the majority of patients in the EOC group were in advanced stages, resulting in a demonstrated advantage of the radiomics model. Future research should include more indeterminate adnexal masses.

In conclusion, by comparing various models, we found the MR-based radiomics model achieved excellent prediction performance for differentiating SCSTs and EOCs. The mixed model which combining the radiomics features and traditional parameters achieve a performance comparable to the radiomics model. Therefore, we believe that the radiomics approach could be a more objective and accurate way for discriminating SCSTs and EOCs. Meanwhile, the mixed model developed in our study could provide a comprehensive, effective manner for clinicians to diagnose and develop appropriate management strategies.

Data availability statement

The datasets presented in this article are not readily available because of institutional restriction. Requests to access the datasets should be directed to Xin Zhao, E-mail:zhaoxinxct@vip.163.com.

Ethics statement

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

Author contributions

Conception and design: MC. Administrative support: XZ. Provision of study materials or patients: ST, ZZ, and TP. Collection and assembly of data: TR, LZ, ZY, and LM. Data analysis and interpretation: MC and XY. Writing - review and editing: MC, ST, and KW. Final approval of manuscript: All authors. All authors contributed to the article and approved the submitted version.

Funding

This research was funded by Henan Province Medical Science and Technology Research Plan Joint Construction Project (Grant No. LHGJ20190380).

Conflict of interest

KW was employed by General Electric GE Healthcare. XY was employed by Huiying Medical Technology.

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.1073983/full#supplementary-material

References

1. Horta M, Cunha TM. Sex cord-stromal tumors of the ovary: a comprehensive review and update for radiologists. *Diagn Interv Radiol* (2015) 21(4):277–86. doi: 10.5152/dir.2015.34414

2. Javadi S, Ganeshan DM, Jensen CT, Iyer RB, Bhosale PR. Comprehensive review of imaging features of sex cord-stromal tumors of the ovary. *Abdom Radiol (NY)* (2021) 46 (4):1519–29. doi: 10.1007/s00261-021-02998-w

3. McCluggage WG, Lax SF, Longacre TA, Malpica A, Soslow RA. *WHO classification of tumors: Female genital tumors. 5th edition*. Lyon: Internal Agency for Research on Cancer (IARC) (2020).

4. Zhao SH, Li HM, Qiang JW, Wang DB, Fan H. The value of MRI for differentiating benign from malignant sex cord-stromal tumors of the ovary: emphasis on diffusion-weighted MR imaging. *J Ovarian Res* (2018) 11(1):73. doi: 10.1186/s13048-018-0444-6

5. Al Harbi R, McNeish IA, El-Bahrawy M. Ovarian sex cord-stromal tumors: an update on clinical features, molecular changes, and management. *Int J Gynecol Cancer* (2021) 31(2):161–8. doi: 10.1136/ijgc-2020-002018

6. Nougaret S, McCague C, Tibermacine H, Vargas HA, Rizzo S, Sala E. Radiomics and radiogenomics in ovarian cancer: a literature review. *Abdom Radiol (NY)* (2021) 46 (6):2308–22. doi: 10.1007/s00261-020-02820-z

7. Sadowski EA, Thomassin-Naggara I, Rockall A, Maturen KE, Forstner R, Jha P, et al. O-RADS MRI risk stratification system: guide for assessing adnexal lesions from the ACR O-RADS committee. *Radiology* (2022) 303(1):35–47. doi: 10.1148/radiol.204371

8. Vargas HA, Veeraraghavan H, Micco M, Nougaret S, Lakhman Y, Meier AA, et al. A novel representation of inter-site tumor heterogeneity from pre-treatment computed tomography textures classifies ovarian cancers by clinical outcome. *Eur Radiol* (2017) 27 (9):3991–4001. doi: 10.1007/s00330-017-4779-y

9. Zhang H, Mao Y, Chen X, Wu G, Liu X, Zhang P, et al. Magnetic resonance imaging radiomics in categorizing ovarian masses and predicting clinical outcome: a preliminary study. *Eur Radiol* (2019) 29(7):3358–71. doi: 10.1007/s00330-019-06124-9

10. Song XL, Ren JL, Zhao D, Wang L, Ren H, Niu J. Radiomics derived from dynamic contrast-enhanced MRI pharmacokinetic protocol features: the value of precision diagnosis ovarian neoplasms. *Eur Radiol* (2021) 31(1):368–78. doi: 10.1007/s00330-020-07112-0

11. Jian J, Li Y, Pickhardt PJ, Xia W, He Z, Zhang R, et al. MR image-based radiomics to differentiate type iota and type IotaIota epithelial ovarian cancers. *Eur Radiol* (2021) 31 (1):403–10. doi: 10.1007/s00330-020-07091-2

12. Zhang Q, Peng Y, Liu W, Bai J, Zheng J, Yang X, et al. Radiomics based on multimodal MRI for the differential diagnosis of benign and malignant breast lesions. *J Magn Reson Imaging* (2020) 52(2):596–607. doi: 10.1002/jmri.27098

13. Lathwal A, Arora C, Raghava GPS. Prediction of risk scores for colorectal cancer patients from the concentration of proteins involved in mitochondrial apoptotic pathway. *PloS One* (2019) 14(9):e0217527. doi: 10.1371/journal

14. Imaoka I, Wada A, Kaji Y, Hayashi T, Hayashi M, Matsuo M, et al. Developing an MR imaging strategy for diagnosis of ovarian masses. *Radiographics* (2006) 26(5):1431–48. doi: 10.1148/rg.265045206

15. Qian L, Ren J, Liu A, Gao Y, Hao F, Zhao L, et al. MR imaging of epithelial ovarian cancer: a combined model to predict histologic subtypes. *Eur Radiol* (2020) 30(11):5815–25. doi: 10.1007/s00330-020-06993-5

16. Hanley KZ, Mosunjac MB. Practical review of ovarian sex cord-stromal tumors. *Surg Pathol Clin* (2019) 12(2):587–620. doi: 10.1016/j.path.2019.02.005

17. Romagnolo C, Leon AE, Fabricio ASC, Taborelli M, Polesel J, Del Pup L, et al. HE4, CA125 and risk of ovarian malignancy algorithm (ROMA) as diagnostic tools for ovarian cancer in patients with a pelvic mass: an Italian multicenter study. *Gynecol Oncol* (2016) 141(2):303–11. doi: 10.1016/j.ygyno.2016.01.016

18. Suri A, Perumal V, Ammalli P, Suryan V, Bansal SK. Diagnostic measures comparison for ovarian malignancy risk in epithelial ovarian cancer patients: a meta-analysis. *Sci Rep* (2021) 11(1):17308. doi: 10.1038/s41598-021-96552-9

19. Yin B, Li W, Cui Y, Chu C, Ding M, Chen J, et al. Value of diffusion-weighted imaging combined with conventional magnetic resonance imaging in the diagnosis of thecomas/fibrothecomas and their differential diagnosis with malignant pelvic solid tumors. *World J Surg Oncol* (2016) 14(1):5. doi: 10.1186/s12957-015-0760-x

20. Wei W, Liu Z, Rong Y, Zhou B, Bai Y, Wei W, et al. A computed tomography-based radiomic prognostic marker of advanced high-grade serous ovarian cancer recurrence: A multicenter study. *Front Oncol* (2019) 9:255. doi: 10.3389/fonc.2019.00255

21. Li W, Feng J, Zhu D, Xiao Z, Liu J, Fang Y, et al. Nomogram model based on radiomics signatures and age to assist in the diagnosis of knee osteoarthritis. *Exp Gerontol* (2022) 171:112031. doi: 10.1016/j.exger.2022.112031

22. Yang Y, Yan LF, Zhang X, Nan HY, Hu YC, Han Y, et al. Optimizing texture retrieving model for multimodal MR image-based support vector machine for classifying glioma. *J Magn Reson Imaging* (2019) 49(5):1263–74. doi: 10.1002/jmri.26524

23. Li C, Wang H, Chen Y, Zhu C, Gao Y, Wang X, et al. Nomograms of combining MRI multisequences radiomics and clinical factors for differentiating high-grade from low-grade serous ovarian carcinoma. *Front Oncol* (2022) 12:816982. doi: 10.3389/fonc.2022.816982

frontiers | Frontiers in Oncology

TYPE Review
PUBLISHED 26 January 2023
DOI 10.3389/fonc.2022.1086258

OPEN ACCESS

EDITED BY
Jasper Nijkamp,
Aarhus University, Denmark

REVIEWED BY
Marcel van Herk,
The University of Manchester,
United Kingdom
Annette Haworth,
The University of Sydney, Australia

*CORRESPONDENCE
Brigid A. McDonald
bmcdonald@mdanderson.org

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

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

RECEIVED 01 November 2022
ACCEPTED 21 December 2022
PUBLISHED 26 January 2023

CITATION
McDonald BA, Zachiu C, Christodouleas J, Naser MA, Ruschin M, Sonke J-J, Thorwarth D, Létourneau D, Tyagi N, Tadic T, Yang J, Li XA, Bernchou U, Hyer DE, Snyder JE, Bubula-Rehm E, Fuller CD and Brock KK (2023) Dose accumulation for MR-guided adaptive radiotherapy: From practical considerations to state-of-the-art clinical implementation. *Front. Oncol.* 12:1086258. doi: 10.3389/fonc.2022.1086258

Dose accumulation for MR-guided adaptive radiotherapy: From practical considerations to state-of-the-art clinical implementation

Brigid A. McDonald[1*], Cornel Zachiu[2], John Christodouleas[3], Mohamed A. Naser[1], Mark Ruschin[4], Jan-Jakob Sonke[5], Daniela Thorwarth[6], Daniel Létourneau[7,8], Neelam Tyagi[9], Tony Tadic[7,8], Jinzhong Yang[10], X. Allen Li[11], Uffe Bernchou[12,13], Daniel E. Hyer[14], Jeffrey E. Snyder[14], Edyta Bubula-Rehm[3], Clifton D. Fuller[1†] and Kristy K. Brock[10,15†]

[1]Department of Radiation Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, United States, [2]Department of Radiotherapy, University Medical Center Utrecht, Utrecht, Netherlands, [3]Elekta AB, Stockholm, Sweden, [4]Department of Radiation Oncology, University of Toronto, Sunnybrook Health Sciences Centre, Toronto, ON, Canada, [5]Department of Radiation Oncology, The Netherlands Cancer Institute, Amsterdam, Netherlands, [6]Section for Biomedical Physics, Department of Radiation Oncology, University of Tuebingen, Tuebingen, Germany, [7]Radiation Medicine Program, Princess Margaret Cancer Centre, Toronto, ON, Canada, [8]Department of Radiation Oncology, University of Toronto, Toronto, ON, Canada, [9]Department of Medical Physics, Memorial Sloan-Kettering Cancer Center, New York, NY, United States, [10]Department of Radiation Physics, The University of Texas MD Anderson Cancer Center, Houston, TX, United States, [11]Department of Radiation Oncology, Medical College of Wisconsin, Milwaukee, WI, United States, [12]Laboratory of Radiation Physics, Department of Oncology, Odense University Hospital, Odense, Denmark, [13]Department of Clinical Research, University of Southern Denmark, Odense, Denmark, [14]Department of Radiation Oncology, University of Iowa Hospitals and Clinics, Iowa City, IA, United States, [15]Department of Imaging Physics, The University of Texas MD Anderson Cancer Center, Houston, TX, United States

MRI-linear accelerator (MR-linac) devices have been introduced into clinical practice in recent years and have enabled MR-guided adaptive radiation therapy (MRgART). However, by accounting for anatomical changes throughout radiation therapy (RT) and delivering different treatment plans at each fraction, adaptive radiation therapy (ART) highlights several challenges in terms of calculating the total delivered dose. Dose accumulation strategies—which typically involve deformable image registration between planning images, deformable dose mapping, and voxel-wise dose summation—can be employed for ART to estimate the delivered dose. In MRgART, plan adaptation on MRI instead of CT necessitates additional considerations in the dose accumulation process because MRI pixel values do not contain the quantitative information used for dose calculation. In this review, we discuss considerations for dose accumulation specific to MRgART and in relation to current MR-linac clinical workflows. We present a general dose accumulation framework for MRgART and discuss relevant quality assurance criteria. Finally, we highlight the clinical importance of dose accumulation in the ART era as well as the possible ways in which dose accumulation can transform clinical practice and improve our ability to deliver personalized RT.

KEYWORDS

dose accumulation, MR-guided radiation therapy, adaptive radiation therapy, deformable image registration, MR-linac

1 Introduction

In the current era of image-guided radiation therapy (RT), many technological advances in on-board imaging systems and treatment delivery techniques have enabled the delivery of highly conformal RT (1, 2). One major development in recent years has been the integration of magnetic resonance imaging (MRI) with linear accelerators (linacs) to form hybrid systems known as MR-linacs (3–5). MRI offers enhanced visualization of both tumor and normal tissue structures compared to other on-board imaging systems such as kV or MV planar x-ray, computed tomography (CT), and cone beam CT (6). The ability to clearly visualize the anatomy during patient setup has accelerated the development of on-line adaptive RT (ART), in which a new treatment plan is created each day based on the patient's daily setup image while the patient is on the treatment table (7–9).

Daily MR-guided ART (MRgART) offers many dosimetric advantages over the traditional single-plan RT workflow, including the ability to conform the high-dose region to the tumor as the anatomy changes throughout the course of RT (9, 10). However, MRgART brings to the forefront a number of challenges in terms of calculating and interpreting the delivered dose that have largely been ignored in the past. In contrast to conventional RT, where dose estimates are calculated on a single pre-treatment simulation image, ART uses multiple plans created on longitudinal images reflecting anatomical variations throughout the treatment course. Currently, most clinical implementations of ART simply create new plans meeting the original treatment constraints and do not use advanced dose accumulation strategies to sum the doses from individual plans and account for these anatomical changes. Without using deformable image registration to truly sum the dose, the contributions of individual plans cannot be interpreted in the context of the total delivered dose and statements regarding over- or under-dosage of tissues cannot be accurately made. The ability to accurately quantify delivered dose allows clinicians to evaluate whether dosimetric criteria are being met in aggregate over multiple fractions and enables adaptation throughout RT to ensure that therapeutic goals are achieved. Dose accumulation also allows us to relate delivered dose to clinical outcomes when evaluating the clinical effectiveness of any ART intervention. In MRgART, plan adaptation on MRI instead of CT adds an additional layer of complexity to the dose accumulation process because the pixel values of MR images do not contain electron density information needed for dose calculation and are subject to signal intensity fluctuations depending on coil setup, magnetic field inhomogeneities, and other factors. Currently, a wide range of research-grade solutions are available for dose accumulation (11–16), but few have been thoroughly validated for clinical use and/or implemented into commercial systems, and none have been specifically optimized for MRgART.

Thus, the development of a robust and accurate dose accumulation solution for MRgART is a subject of active research, particularly within the MR-Linac Consortium (17). In this article, we review and discuss the current status, practical challenges, and potential role of dose accumulation for MRgART and outline a framework for quality assurance of proposed solutions.

2 Definition and clinical relevance of dose accumulation

"Dose accumulation" is a term that encompasses a range of techniques for summing multiple RT dose distributions for a single patient (18, 19). The goal is to arrive at a better estimate of the delivered dose compared to standard RT practices, in which a single plan is generated and the dose is calculated only on a static pre-treatment representation of the anatomy. Dose distributions in conventional RT are modeled on the often-flawed assumption that the tumor and surrounding anatomy remain static throughout the course of RT. While it is well understood that the pre-treatment *planned* dose is not equal to the true *delivered* dose due to both RT-induced anatomical changes over time and positional uncertainties during each RT fraction, our current treatment plan acceptability criteria and clinical outcome models are based only on the planned dose (19). Dose accumulation can help us arrive at a more realistic depiction of the delivered dose, but uncertainties in the dose accumulation process make the calculated dose distributions, at best, *estimates* of the delivered dose. Nonetheless, dose accumulation remains a vital mechanism for quantifying delivered dose and evaluating the benefit of various ART strategies used in clinical practice.

Dose accumulation is most often considered in the context of ART, where adaptive plans are generated on either on-board setup images (i.e. on-line ART) or simulation images acquired throughout RT (i.e. off-line ART). In a general dose accumulation pipeline, the dose distribution for each plan is scaled to the number of fractions delivered, the planning images are co-registered, and the dose distributions are then mapped according to the estimated displacements and added voxel-by-voxel (12, 16, 20). Dose accumulation most often utilizes deformable image registration (DIR), which creates a spatial correspondence between two images that accounts for anatomical deformations (21). Doses may be mapped backward onto the pre-treatment anatomy or forward to any time point during treatment, depending on the intended use case for the accumulated dose. Furthermore, dose accumulation techniques may be classified as either inter-fraction or intra-fraction approaches. Inter-fraction dose accumulation uses only the setup image at each fraction (12, 20, 22, 23), while intra-fraction methods account for motion during beam delivery using continuous or periodic motion monitoring imaging such as cine MRI (11, 14). Despite the broad range of methods to perform dose accumulation, we will limit our discussions in this paper to inter-fraction DIR-based methods, as, presumably, any multi-fraction treatment regimen would profit from accurate serial dose estimation. Finally, dose accumulation can be performed either as dose back projection onto the pre-treatment anatomy or forward projection onto the anatomy at any time during or after RT. Both scenarios will be discussed in this article, but the general steps of a dose accumulation workflow remain the same.

Dose accumulation is valuable from a clinical standpoint for a number of reasons, both during and after RT. MRgART allows physicians to set complex goals for treatment personalized for each individual patient, and dose accumulation helps us determine whether the intended goals are being met. These intentions may include sparing dose to specific organs at risk (OARs), escalating dose

to target structures, or modifying target volumes as the tumor shrinks to spare tumor-adjacent OARs (24). During a course of treatment, the ability to accumulate dose informs clinicians as to whether or not the daily dose distributions are representative of the cumulative dose. For example, if the dose to an OAR exceeds tolerance on one day, it is clinically relevant to know whether the cumulative dose is in excess, as this may inform the optimization strategy for subsequent fractions (25–27). Alternatively, if OAR doses are sufficiently low after a certain number of fractions, the physician may choose to increase the target dose or add an extra fraction based on individualized treatment response (28–30). When any adaptive modification is made during a course of RT without using a validated dose accumulation method to accurately quantify total delivered dose, potential risks to the patient include overdosing OARs and underdosing target volumes (thus increasing the risk of incomplete treatment response or cancer recurrence). Dose accumulation can enable us to evaluate the safety and efficacy of various ART strategies, which will allow clinicians to personalize RT for each patient while maintaining our field's commitment to safe, evidence-based treatment approaches.

Dose accumulation may also lead to opportunities to reevaluate normal tissue complication probability (NTCP) and tumor control probability (TCP) models (31–35). The existing dose-response models are largely based on doses calculated on the pre-treatment simulation anatomy, which is often assumed to be static. If we can more accurately quantify the delivered dose after the conclusion of RT in a systematic way, there is an opportunity to refine the current NTCP and TCP models and develop a novel set of planning constraints for the era of ART and personalized medicine (18, 36).

3 Considerations for dose accumulation with on-line MR-guided adaptive RT

There are currently two MR-linac systems commercially available: the Elekta Unity (Elekta AB; Stockholm, Sweden) and the ViewRay MRIdian (ViewRay, Inc.; Cleveland, OH, USA). The Unity system combines a modified 1.5 T Philips diagnostic MRI scanner (Philips Healthcare; Best, Netherlands) with a 7 MV flattening filter-free (FFF) linac, while the MRIdian system uses a 0.35 T MRI with a 6 MV FFF linac. Although the specific workflows of the two systems differ, both systems are capable of on-line ART by registering a prior reference (i.e. planning) image to the daily setup image and adapting the reference plan. The MRIdian workflow offers the choice between adapting the reference plan to the current anatomy and delivering the reference plan without modification (8). After the reference and setup images are registered, the user views the predicted dose of the reference plan on the current anatomy and decides whether to treat with the reference plan or adapt. In contrast, the Unity workflow provides two workflow options: Adapt to Position (ATP) and Adapt to Shape (ATS) (9). ATP virtually accounts for the isocenter shift by rigidly registering the reference and setup images but recalculates the dose on the reference image (which can be either a CT or MRI), while ATS involves a full plan adaptation on the setup image (MRI). In either workflow, the original multileaf collimator segments and monitor units can be kept the same to deliver the reference plan without modification.

Although the end-to-end MRgART process differs among MR-linac platforms, these workflows can be summarized as three general classes of solutions: i) treat with the initial plan (e.g. ViewRay MRIdian workflow or Elekta Unity workflows), ii) shift the reference plan to a new position (virtual isocenter shift) and recalculate the dose on the reference plan anatomy (e.g. Elekta Unity ATP workflow), and iii) perform a full re-optimization of the treatment plan on the anatomy of the day (e.g. ViewRay MRIdian workflow or Elekta Unity ATS workflow).

In the absence of commercially available dose accumulation tools that can be run in parallel with the on-line MRgART workflows, there is currently no standardized mechanism for summing and tracking delivered dose over the course of RT. In this section, we will outline a general framework for a potential dose accumulation process for on-line MRgART and discuss considerations and challenges for each step.

A possible workflow for an inter-fraction DIR-based dose accumulation for on-line MRgART would include the following four steps:

1. Autosegmentation of daily setup images for electron density mapping (optional)
2. Recalculation of adaptive plan doses on daily images (optional)
3. DIR between daily images & reference image
4. Deformable dose mapping and dose summation

For each step of the general framework (Figure 1), there may be multiple techniques that may be used to accomplish the same goal, each of which must be thoroughly evaluated for each anatomical site and application. (Depending on the on-line adaptive workflow used, steps 1 and 2 may or may not be necessary; they are most relevant in a virtual isocenter shift workflow where adaptive plan doses are calculated on the reference image rather than daily image. In this case, the delivered dose can be calculated on the daily setup image by segmenting the image to produce an electron density map and recalculating the dose.) We present many of the common approaches that are being investigated for each dose accumulation step as well as considerations for each technique in the realm of MRgART.

3.1 Step 1: Autosegmentation of daily setup images for electron density mapping (optional)

To generate an adaptive treatment plan on MRI, the target volume(s) and OARs must be segmented on the image used for dose calculation to create planning constraints and to approximate an electron density map for dose calculation on MRI. Depending on the on-line MRgART workflow used, an autosegmentation step may be necessary for electron density mapping to accurately reconstruct the delivered dose on the daily images prior to implementing a DIR-based dose accumulation approach. This step is particularly relevant in on-

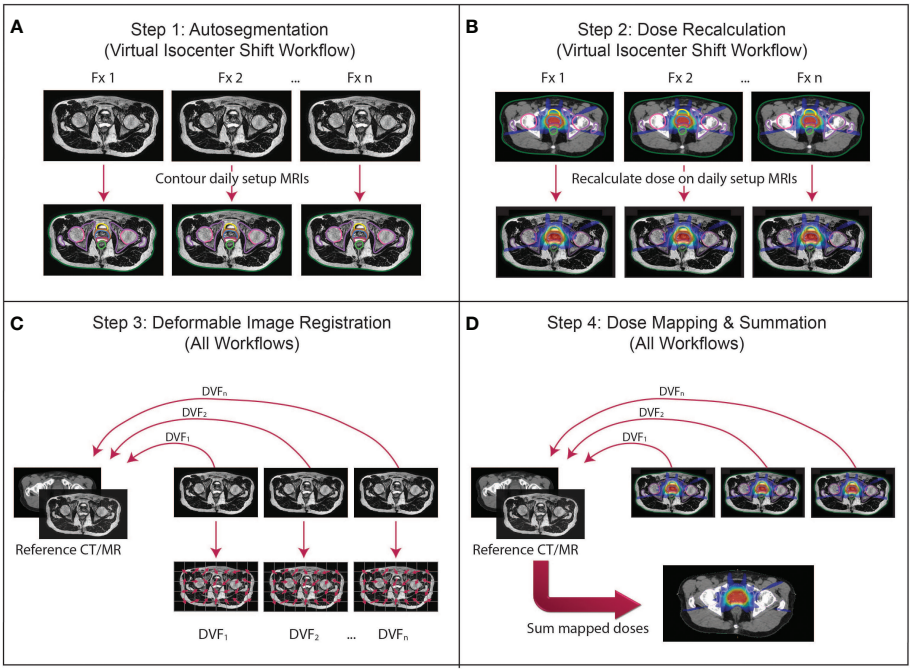


FIGURE 1
The proposed dose accumulation framework for MRgART. **(A)** Step 1 (virtual isocenter shift workflow only): Because the virtual isocenter shift workflow calculates dose on the reference image rather than the daily setup MRI, the daily images are not contoured during the treatment and must be segmented off-line. **(B)** Step 2 (virtual isocenter shift workflow only): The doses for each fraction must be recalculated on the daily setup images off-line to reconstruct the delivered dose at each fraction. **(C)** Step 3 (all workflows): The geometric correspondence between each daily image and the reference image set (i.e. simulation or any other established time point) is created *via* deformable image registration (DIR) and represented by a deformation vector field (DVF). **(D)** Step 4 (all workflows): The DVFs are applied to the corresponding dose distributions to map the doses onto the reference image set, then the doses are summed to calculate the final accumulated dose. (This figure describes a general workflow where a transformation is established that links the daily image to the reference image through a DVF for the purpose of dose mapping. The details of the implementation depend on the DIR algorithm and mechanism for mapping the dose. The DVFs in this figure are demonstrating the direction of the mapping from the daily images into the reference frame. Whether the dose is "pushed" or "pulled" and whether an inverse DVF is required depend on the implementation of the algorithm created by the user).

line workflows where the daily image is not contoured and the dose is calculated on the reference image rather than the daily image (e.g. virtual isocenter shift workflow) (9). Otherwise, if segmentation is done during the on-line workflow, this step would not be necessary.

Several autosegmentation methods are appropriate for MRgART, including DIR-based contour propagation from the reference image as well as atlas-based and deep learning approaches. In the first method, DIR is used to generate a spatial correspondence between the daily image and the reference image or some other prior image, and the contours are mapped according to the established transformation (37). The quality of the segmentations depends on the DIR algorithm performance; specific considerations for DIR with MRgART are discussed in Step 3 below.

Next, atlas-based autosegmentation uses a small collection of contoured image sets (i.e. "atlases") which serve as templates for contouring the image set of interest (i.e. "patient") (38). Each atlas is aligned with the patient, and contours are propagated *via* DIR to produce one structure set per atlas on the patient. These intermediate results are combined into a final structure set using either a method to combine all structure sets such as a STAPLE algorithm (38–40) or a voting mechanism which selects the best contour for each structure such as Majority Vote (38, 41) or Random Forest (42, 43). Atlas-based approaches have historically used atlases from other patients (i.e. "inter-patient"), but serial imaging for on-line ART introduces the possibility of using a patient's own images from prior fractions as atlases (i.e. "intra-patient") (44). Using multiple intra-patient atlases has the potential to improve performance over both inter-patient atlas-based approaches and intra-patient DIR from the reference plan image—especially for later fractions when anatomical changes from the simulation image may be quite large—but requires further validation. Intra-patient methods can also produce comparable performance to inter-patient methods with fewer atlases, which speeds up execution time considerably.

Another promising autosegmentation approach, deep learning, uses a large number of contoured image sets to train a computer algorithm called a deep neural network to contour unlabeled input images (45–47). Most deep learning autosegmentation methods use a convolutional neural network architecture such as U-Net (48), which is formed by stacking multiple hidden layers, including convolutional layers, that each learn a feature of the training data. After a model is trained and validated, it can be used to contour unseen input images at rapid speeds. While atlas-based approaches typically reach peak performance using 5-15 atlases (49–52), deep learning generally requires dozens to hundreds of images as training data for optimal performance (53, 54). As more patients are treated on MR-linac devices with standardized MR sequences, we can leverage vendor-

supported data repositories such as the MOMENTUM study (55) to aggregate curated, high-quality imaging data in a coordinated, efficient manner to train these models (56).

There are a few nuances to autosegmentation on MR rather than CT. MR demonstrates enhanced soft tissue contrast, rendering many OAR boundaries more clearly visible on MR than on CT and thus potentially improving the quality of autosegmentation (57). However, unlike CT, pixel values in T1- and T2-weighted MRIs are not inherently quantitative and are subject to variations due to coil positioning, radiofrequency and electronic noise, and magnetic field inhomogeneities (58–60). The arbitrary pixel value scaling and image intensity variations may affect the performance of autosegmentation algorithms, which rely on intensity and/or contrast similarities between training data/atlases and the image sets to be segmented. These effects can be minimized on the acquisition side by using consistent pulse sequence parameters, immobilization, and coil setup during MRgART and on the post-processing side by performing image intensity standardization on the images (61, 62). Furthermore, while one advantage of MR is the wide array of image contrast mechanisms obtained by using different pulse sequences, each of the aforementioned autosegmentation techniques is pulse sequence-specific, meaning that they should be trained and executed on images with identical pulse sequence parameters for optimal performance (63, 64). However, aggregating data across different patients and/or institutions for deep learning models will require adoption of consistent protocols across sites or the implementation of data augmentation techniques to generalize trained models to multiple sequences (65, 66).

3.2 Step 2: Recalculation of adaptive plan doses on daily images (optional)

Depending on the workflow used, the dose distributions for adaptive plans may need to be recalculated on the daily setup image off-line prior to accumulating the doses. Like Step 1, this step would be required if the dose is calculated on the reference image rather than the daily setup image during on-line plan adaptation (e.g. virtual isocenter shift workflow).

Unlike CT, where the pixel values represent physical measurements of photon attenuation in Hounsfield units (HU) and are easily converted to relative electron density (ED) maps to calculate dose, dose calculation on MRI requires approximation of the ED values for each voxel. Currently, the most common ED approximation method is called bulk density assignment: for each structure contoured on the planning CT, the mean ED value of the structure on CT is assigned to all voxels in the structure on MRI (8, 9). These values may be overridden with user-defined values when corrections are needed or if no CT is available in an MR-only RT workflow. Bulk density assignment has been shown to be accurate across disease sites, resulting in minimal deviations in dose volume histogram parameters compared to doses calculated on CT (37–39), but may not adequately handle largely heterogeneous volumes such as spinal vertebrae (40), femoral heads (39), or lung (41). An alternative approach is to deformably register the planning CT into the MR frame of reference to create an ED map (42, 43). While this method preserves the heterogeneity of ED values throughout each structure, CT-to-MR DIR may have limited accuracy due to the different signal and contrast characteristics of the two modalities (44, 45), and it will likely fail in situations with large deformations such as lung collapse.

As researchers move toward MR-only treatment planning, alternative ED approximation methods are also being explored that do not require any CT input and would eliminate the need for segmentation of daily MRIs for ED mapping (46). Synthetic CTs can be generated directly from MRIs using specialized MR sequences such as Dixon MRI, which separates signal from fat and water and enables clusters of voxels to be assigned discrete tissue classes with associated ED values, similar to the bulk density assignment (47, 48). Another option is to apply a calibration curve relating MRI signal intensities to HU values on a voxel-wise basis to generate a synthetic CT that preserves tissue inhomogeneities (49, 50). This method can be used directly on the T1- or T2-weighted images used for setup and plan reoptimization rather than requiring an additional MR sequence. However, MR pixel values are subject to intensity variations, and depending on the contrast mechanism of the pulse sequence, the calibration curve is unlikely to follow a simple linear or logarithmic fit. Deep learning models may also be trained to generate heterogeneous synthetic CTs from Dixon or T1- or T2-weighted MRIs. Several studies have demonstrated excellent performance of such models (51–54).

It should also be noted that a shift invariance approach, which involves a simple shift of the dose distribution to account for the isocenter shift, may be used instead of a full dose recalculation. This approach has been shown to be a good approximation of the delivered dose for many deep lying tumors but fails for shallow tumors in the build-up region and when anatomical changes are substantial (55, 56).

3.3 Step 3: DIR between daily images & reference image

Once the dose distributions for each adaptive plan have been accurately calculated on each daily setup image, the next step is to establish a geometric transformation between each daily image and the reference image *via* DIR. This step of dose accumulation is required regardless of the on-line MRgART workflow used.

There are several DIR approaches with varying degrees of complexity, which are summarized in the literature (57–59). Most implementations of DIR share three main components: 1) a transformation, or a mathematical model, that establishes the geometric correspondence between the source and target images; 2) an objective function, which typically includes a similarity metric used for evaluating the alignment between the images and a regularization term to impose constraints on the deformation field; and 3) an optimization method that optimizes the parameters of the transformation model to maximize the similarity between the source and target images under the imposed constraints (57–59). In this section, we will focus on the first two elements and their implications for dose accumulation in MR-guided adaptive RT.

A number of transformation models are available for DIR, including B-spline and several non-parametric models. The B-spline transformation is a commonly used non-linear parametric model generated using a weighted sum of a set of spline functions

defined at a set of control points spaced evenly throughout the source and target images (21, 57, 60). In contrast, non-parametric models such as elastic, fluid and optical flow and finite element methods can generate much more complex transformations to model anatomical changes (58, 59, 61–64). Unlike B-spline models, which represent the image deformation using parameters defined at each control point, non-parametric transformation models are usually represented by more complex deformation vector fields (DVFs) where the displacement in all three directions is defined for each individual image voxel. When selecting a transformation model to use for dose accumulation, one must consider the expected degree of deformation in the anatomical site of interest as well as the complexity and underlying assumptions of the model. This choice is particularly important for dose accumulation, as the transformation model impacts the registration accuracy within contrast-devoid regions and, implicitly, the accuracy of the accumulated dose.

The second component of any registration algorithm is the similarity metric, which is used to evaluate the alignment between the registered images throughout successive iterations in the optimization process. Similarity metrics are classified as either intensity-based or feature-based (58, 59). Intensity-based metrics evaluate the alignment of intensity patterns (i.e. gray-scale information) between the source and target images. Feature-based metrics use anatomical landmarks such as points, lines, and surfaces to obtain the correspondence between the source and target images.

The choice of similarity metric depends on the intensity ranges and modalities of the source and target images. In the context of MR-guided adaptive RT, intensity-based metrics such as sum of squared intensity differences (65) and/or cross correlation (66) typically work well for images of the same modality and intensity range, such as the daily MRIs, as long as the same MRI pulse sequence is used for setup at each fraction of a patient's treatment. Registration across modalities such as CT-to-MR and registration across different MR sequences such as T1-to-T2 present a more complex problem due to intensity inconsistencies between the images. Despite the different contrast mechanisms and intensity ranges, most of the recently developed multi-modality registration approaches still use certain intensity-based metrics such as normalized mutual information (67) over feature-based metrics. Normalized mutual information is based on global histogram matching (i.e. the distribution of intensity values across the entire image). Several alternative approaches have been proposed for multi-modality registration, including normalized gradient fields and modality independent descriptors. The normalized gradient fields metric uses the gradient (i.e. derivative) of the intensity in each image rather than the image intensities themselves (68, 69). In the modality independent descriptors approach, the images are pre-processed into a modality-independent format that preserves local image feature information and can be directly compared using established similarity metrics (70, 71).

While the field of multi-modality DIR has made great progress in recent years, many of these newer techniques have yet to be implemented into commercial treatment planning systems for MRgART. Recent studies have shown that the CT-to-MR registration currently used for the 1.5T MR-linac workflow underperforms compared to same-sequence MR-to-MR registration in both prostate (45) and head and neck cancers (72). This discrepancy has implications for daily plan adaptation and dose accumulation. Many clinics create the reference plan on the CT simulation, while others acquire a CT for electron density information but create the reference plan on the MR simulation. The latter method may improve the quality of the DIR in the on-line workflow, which will likely reduce the time spent manually editing contours. For dose accumulation, if doses are mapped back to the pre-treatment time point for comparison to the reference plan, the DIR quality may improve if the daily MRIs and daily fraction doses are registered to the MR simulation image or first fraction MRI rather than the CT. However, the implementation of state-of-the-art algorithms may improve the performance of multi-modality image registration, and rigorous evaluation of both CT-to-MR and MR-to-MR DIR quality is needed for all organ sites.

In addition to the similarity metric, DIR algorithms must impose constraints on the deformation field by adding a regularization term to the objective function. When a regularization term is used, the final solution (i.e. the estimated deformation field) will be a tradeoff between maximizing the similarity between the source and target images and satisfying the constraints. Some examples of constraints include preserving topology, ensuring a smooth deformation field, and penalizing non-physical deformations given prior knowledge of the underlying anatomy (e.g. preventing the warping of rigid structures such as bones) (59, 73). In anatomical regions with sliding tissues such as the lung and chest wall, regularization terms that allow a discontinuity in the DVF can be used (74, 75). Another constraint with particular relevance to dose accumulation is inverse consistency, which ensures that the forward and backward transformations, computed simultaneously, are direct inverse mappings of one another. Inverse consistency would be an important consideration if clinicians are interested in evaluating accumulated dose in both the forward and backward directions. (See "Interpretation of Accumulated Dose" section for a more detailed discussion on forward and backward mapping.)

While the myriad of deformable image registration algorithms enable us to model anatomical changes in a wide range of clinical scenarios, these algorithms are all based on fundamental assumptions about the anatomy that do not always hold true throughout a course of RT. For example, assumptions that the deformations are smooth/continuous and invertible are violated in scenarios such as organ sliding and tissue gain/loss. While these assumptions are necessary for estimating deformations in an anatomically plausible matter, they also fundamentally limit the ability of DIR to accurately characterize the true anatomical changes.

An additional consideration for DIR in the context of dose accumulation for daily MRgART is that image sets from up to dozens of fractions will need to be registered, but image registration occurs as a separate operation between only two sets of images. If all doses are being mapped back to the simulation image, the simplest options for composing registrations are 1) registering each daily image to the reference image, or 2) registering each image onto the previous fraction's image in a sequential fashion to create a "DVF chain" (76). The second option is particularly advantageous when anatomical changes between the beginning and end of the course of RT are substantial because it minimizes the change between each image set being registered. However, if the performance of the DIR algorithm is poor despite the minimal anatomical change from day to

day, this approach compounds the error throughout the chain of registrations and dose deformations (76).

3.4 Step 4: Deformable dose mapping and dose summation

Once the daily image sets are registered, the doses are mapped by applying the transformations defined by the image registrations to the dose grids. If the doses of each adaptive plan are scaled to the prescription dose throughout the entire course of RT, then they must be scaled down to the dose delivered at each fraction. Finally, the mapped doses are summed voxel-by-voxel to calculate the delivered dose distribution. (If large variations exist in the dose delivered to each structure from day to day, fractionation effects should also be taken into account using linear quadradic models.)

While this process is, in effect, a simple computational task after the DVFs between each corresponding image set have been calculated, the discrete nature of voxels/dose grids and the deformations occurring between each time point make it infeasible to assume that the same individual cells are contained within the same matched voxels from day to day. For this reason, tri-linear interpolation is often employed: each voxel is divided into sub-volumes before dose mapping, and the values of the mapped sub-volumes corresponding to the same voxel on the reference dose grid are averaged and assigned to that voxel (21, 77, 78). The interpolation method is fast, which may be highly advantageous in the context of MRgART when there is a different treatment plan for each fraction. However, this method is less accurate in steep dose gradients and treats dose as an imaging voxel intensity rather than a physical quantity (i.e. energy per unit mass) (79, 80).

It's important to note that the details of the algorithm that performs the DIR and dose mapping will specify the direction of the DVF and whether the dose mapping is done by "pushing" the dose from the image it's calculated on to the summed image or "pulling" it. In many scenarios, one direction of the DVF is sufficient for DIR and dose mapping, but how that is implemented depends on the specifics of the algorithm. The combination of the DIR algorithm and dose mapping mechanism will dictate whether an inverse DVF is required.

While this issue is not unique to MR-guided therapy, no discussion on dose accumulation would be complete without mentioning the difficulties in accurately accumulating dose when tumors or OARs exhibit volumetric changes throughout RT. Conceptually, one must consider what happens to the dose delivered to a small volume of tissue if that tissue disappears before the end of treatment, which routinely occurs for certain tumor types such as human papillomavirus-positive oropharyngeal cancers (16). This dilemma is illustrated mathematically by Zhong & Chetty (79), who demonstrate that simply deformably mapping dose violates the fundamental physics principle of conservation of energy. Energy/mass transfer methods have been proposed to account for conservation of energy, whereby the energy deposited in each voxel and the mass of each voxel are mapped separately onto the reference dose grid then divided to calculate the dose (21). The initial implementations of energy/mass transfer-based dose accumulation used Monte Carlo methods to simulate the energy deposition in each voxel (81, 82), which required significant computational power. In more recent years, these approaches have been extended to non-Monte Carlo techniques that can be interfaced with commercial treatment planning systems (83) and run in real-time (84).

Finally, it is important to remember when designing a dose accumulation workflow that translation of the couch after the image acquisition must be incorporated into the dose accumulation process. The MR-linac on-line treatment adaptation workflows already account for this through either a physical couch shift (ViewRay MRIdian) or a virtual isocenter shift (Elekta Unity). However, if images and doses are exported outside of the closed system of the MR-linac device and its associated treatment planning system, one must ensure that the dose accumulation algorithm accounts for the isocenter shift prior to dose mapping.

4 Validation and quality assurance of DIR for dose accumulation

4.1 General considerations

To use DIR within the dose accumulation workflow, the employed DIR algorithms must fulfill particular validation benchmarks and must be subjected to stringent quality assurance (QA) criteria to ensure patient safety and the attainment of the therapeutic endpoint. A distinction has to be made, however, whether the aim is the commissioning of a DIR solution prior to clinical use or whether the QA of the estimated deformations at the time of treatment needs to be ensured. Depending on the situation, different criteria may be considered in favor of others. For example, criteria which require known inputs such as expert contours, landmarks and/or deformations are more suitable for commissioning, whereas QA at the time of treatment preferably relies on criteria with a higher degree of automation. In this section, we will briefly describe several classes of criteria which can be used for this purpose, while, where applicable, also indicating value ranges for these criteria where DIR algorithms may be considered reliable. It is important to note, however, that while the discussed criteria provide a practical starting point for evaluation of DIR, acceptability criteria for any metric are necessarily driven by specific application considerations. A summary of the criteria discussed below and recommended tolerances are provided in Table 1.

4.2 Qualitative criteria

A simple approach for both validation and QA of DIR is visual inspection of the post-registration alignment of organ boundaries and/or high-contrast anatomical landmarks (73, 85, 98). This can typically be performed by the radiation oncologist, physicist, and/or radiation therapy technologist as soon as the registration step is completed and can help identify gross potential mis-registrations. For same-contrast images, this may be complemented by a visualization of intensity-based maps such as the absolute image difference or structural similarity (99–101), which contain brighter or darker voxels, depending on the degree of alignment between the registered images. However, visual inspection is typically subject to interpretation, and two registrations can be visually identical while

TABLE 1 Summary of quality assurance metrics for deformable image registration (DIR) and recommended tolerances.

Metric	Tolerance/Ranges	Reference
Dice Similarity Coefficient	Structure size-dependent (~ 0.8 – 0.9)	(73, 85)
Jaccard Index	Structure size-dependent (~ 0.8 – 0.9)	(86–88)
Hausdorff Distance, Mean Distance to Agreement	Maximum voxel size (~2 – 3 mm)	(73, 85)
Target Registration Error	Maximum voxel size (~2 – 3 mm)	(73, 85)
Distance to Discordance Metric	Maximum voxel size (~2 – 3 mm)	(89, 90)
Jacobian Determinant	Tissue-dependent (~0.8 – 1.2 for biological soft tissues)	(73, 91, 92)
Curl Magnitude	Tissue-dependent (~0 – 0.2 for biological soft tissues)	(91, 93)
Biomechanical Criteria with Thresholds on Mechanical Stresses	Tissue-dependent	(91)
Dosimetric criteria	Application-dependent	(94–97)

having completely different anatomical mappings. While visual assessment of the DIR result is necessary, it alone is not sufficient, and therefore additional objective complementary criteria are required.

4.3 Contour-based criteria

A feasible solution towards the commissioning of DIR is the evaluation of the algorithm's capability for aligning organ boundaries. If a DIR algorithm consistently fails to provide a satisfactory boundary alignment, then that is a good indicator that its performance may be insufficient for an accurate dose accumulation. In this sense, the Dice similarity coefficient (DSC) (73, 102, 103) and the Jaccard index (103, 104) provide an objective manner of evaluating an algorithm's capability for organ boundary alignment. Using the DSC and the Jaccard index as a DIR validation criterion requires expert-drawn contours of the same anatomical structure(s) on the images to be registered for ground truth comparison. The DIR-estimated deformations are used to map the contour(s), and the two criteria can be used to evaluate the overlap between the expert-drawn and the DIR-mapped contour(s). The values of both the DSC and the Jaccard index range from zero to one, with zero indicating no overlap and one corresponding with perfect overlap. In the scope of image-guided radiotherapy, a DIR algorithm that provides consistent DSC and Jaccard index values of 0.8 – 0.9 is generally considered to be reliable (73, 85–88). However, the values of the DSC and Jaccard index depend heavily on the volume of the structure; a large structure such as brain could have a DSC or Jaccard index close to 1 and a small structure such as the optic chiasm close to 0 for the same geometrical distance. Therefore, tolerance values for DSC and Jaccard index should be based on structure size and cannot be generalized.

In addition to the DSC and the Jaccard index, complementary criteria such as the Hausdorff distance (HD) (103, 105) and the mean distance to agreement (MDA) (73, 86, 103) are also recommended for inclusion. Instead of evaluating the post-registration contour overlap, the HD and the MDA are used to compute an actual distance between the mapped and the expert-drawn contours, providing further information on the validity of the DIR-estimated deformations in the vicinity of organ boundaries. Acceptable values for both the HD and MDA should be within the uncertainty of the contouring process, typically in the 2 – 3 mm range (73, 85). However, it should be noted that the HD represents the maximum distance between associated boundary points in the contours and is therefore more prone to outliers than the MDA; alternatively, the 95% HD, which reports the 95^{th} percentile of distances between boundary points, can be used instead to limit the effects of outliers.

As previously stated, criteria such as the DSC, Jaccard index, HD and/or MDA provide an evaluation of the organ boundary alignment and/or volume overlap capabilities of DIR algorithms. However, they are limited in their ability to provide a comprehensive evaluation of the estimated deformations because they provide no information about the accuracy within organ boundaries. Moreover, criteria such as the DSC and the Jaccard index are strongly dependent on the size of the evaluated contours and may thus lead to an interpretation bias. On the other hand, it has to be taken into consideration that the MR-Linac allows for a more accurate definition and delineation of anatomical structures due to the high soft tissue contrast present in the MR images. This is a considerable advantage in areas containing a large number of small anatomical structures (e.g. head and neck), thus allowing a more consistent evaluation of the DIR algorithm performance *via* contour-based criteria.

4.4 Criteria employing known displacements/deformations

The target registration error (TRE) (73, 106) allows a quantitative evaluation of a DIR method's accuracy and precision in any anatomical region showcasing identifiable anatomical landmarks

and is thus not limited to organ boundaries. The calculation of the TRE requires an expert to manually indicate the location of the same anatomical landmarks on the images to be registered. The DIR-estimated deformations are then used to map the landmarks annotated on one of the images onto the second image. The distance between the mapped and the manually annotated landmarks on the second image is then calculated to evaluate the registration errors. For a DIR algorithm to be considered reliable for clinical use, the average TRE calculated for the annotated landmarks should consistently reside under the maximum image voxel size, typically in the 2-3 mm range (73, 85). Similar to defining anatomical boundaries, the manual annotation of anatomical landmarks is also facilitated by the high soft tissue contrast provided by MR images. In turn, this may lead to an improved landmark-based evaluation of a DIR solution compared to imaging modalities such as CT/CBCT, in the absence of contrast administration. Still, while such an approach can aid in evaluating the typical accuracy and precision of DIR methods in the vicinity of high-contrast anatomical landmarks, it has limited validation capabilities for homogeneous image areas due to the intrinsic difficulty of defining and annotating landmarks in such regions. Also, many landmarks are defined at extreme positions of an organ and can therefore represent different anatomy if the organ slides or rotates (107, 108). Moreover, the landmarks themselves are often used by the data similarity term of the DIR algorithm and therefore may not be a completely independent measure.

These limitations can be addressed, for example, by the use of physical (96, 109–111) and/or digital phantoms (92, 111, 112). Physical phantoms typically consist of tissue mimicking materials which are displaced/deformed in a controlled manner under the effect of a mechanical actuator (e.g. a piston) such that the displacements/deformations of the phantom are (partially) known (potentially by the use of implanted fiducials). While physical phantoms which can be effectively and safely operated within an MR-Linac are commercially available, the design, development and optimization of MR-visible phantoms is an ongoing area of investigation.

4.5 Criteria based on tissue biomechanics

The QA of DIR algorithms can also be performed by employing criteria based on the mechanical properties of the observed anatomy (87, 91). Depending on individual physical properties, the deformations of the various anatomical tissues have a limited number of degrees of freedom. For example, elastic biological soft tissues are near-incompressible due to their high water content, and therefore, strong compressions and expansions within such regions are anatomically implausible. In effect, if such implausible deformation patterns are estimated by the employed DIR algorithm, a misregistration has most likely occurred.

To determine the amount of compression or expansion present in the estimated deformations, a voxel-wise evaluation of the Jacobian determinant of the deformations can be performed (73, 92). It is known from continuum mechanics that the Jacobian determinant of incompressible materials is equal to 1. Thus, large deviations from 1 within the deformations estimated for elastic biological soft tissues are indicative of misregistration. Similarly, during typical anatomical motion, strong local torsions are not expected to occur deep within the boundaries of elastic soft tissues and can also indicate the occurrence of misregistration. Such torsions can be quantified, for example, by evaluating the curl magnitude of the deformations, with large local deviations from zero being anatomically implausible (93). Typical values of these metrics for the liver and kidneys have been determined to be between 0.8-1.2 for the Jacobian and 0-0.2 for the curl magnitude (91). We do not expect the deformations within other elastic soft tissues to deviate significantly from these values.

Alternatively, QA criteria relying on the biomechanical properties of the observed anatomy can be even further individualized for specific tissues. By providing the elastic modulus and Poisson ratio of the structures of interest as input during the planning/re-planning phase of treatment, the mechanical stress occurring as a result of the estimated deformations can be evaluated within these regions (91). The two parameters can be extracted either from look-up tables, following mechanical tests performed in previous studies, or from quantitative functional imaging such as MR elastography (113). During typical anatomical motion, the mechanical stress within the observed tissues is not expected to lead to tissue rupture or occlusion of blood circulation. Therefore, if such occurrences are detected within the DIR-estimated deformations, they are most likely indicative of misregistration and have been shown to be correlated with errors within the accumulated dose map (91). Such tissue-damaging mechanical stress limits are tissue-specific and can again be extracted from look-up tables generated on the basis of previous studies which have performed the required mechanical tests.

4.6 Dosimetric criteria

While it is generally agreed that geometric DIR uncertainties play a determining role in the accuracy of deformable dose accumulation, the precise manner in which such uncertainties relate to dose accumulation errors is the topic of ongoing research. For example, DIR errors within isodose areas will most likely have less of an impact on the overall accumulated dose compared to registration errors occurring within regions containing steep dose gradients. In this sense, previous studies propose establishing a non-linear relationship between the DIR and the warped/accumulated dose uncertainties (94, 95). The challenge is hereby the selection of the criteria used to evaluate the DIR accuracy as well as the dosimetric parameters to be used as an input for the non-linear model. An additional challenge is determining the model itself, which could, for example, imply an empirical selection of a composition of a set of mathematical functions or an estimation of the model by means of machine learning. An alternative approach, which bypasses the construction of such a model, consists of the use of deformable phantoms made of radiosensitive gels (96, 97). The design paradigm for such phantoms is similar to the one used for phantoms evaluating DIR uncertainties (as described above) with, for example, the additional inclusion of radiosensitive materials whose MR signal is dependent on the absorbed radiation dose. Irradiating such phantoms while undergoing deformations, followed by an MR-based readout of the delivered dose, provides a theoretical gold standard for the warped dose, which can be subsequently compared to the one estimated by the DIR algorithm. For the purpose of commissioning DIR algorithms for dose accumulation, this provides a direct estimation of the

expected dose accumulation errors of the algorithm under evaluation. However, it is important to take into consideration the high sensitivity of existing gels to environmental factors such as temperature, as well as phantom deformations in the absence of irradiation. This can in turn introduce uncertainties within the readout process and therefore bias the evaluation of the DIR algorithm. Consequently, the construction of robust radiosensitive phantoms is the subject of ongoing research.

5 Interpretation of accumulated dose

For simplicity, we have assumed until this point that the dose distributions from each adaptive plan are always mapped back onto the reference (i.e. simulation) image for comparison between the planned and delivered doses. In clinical practice, doses may be mapped either forward or backward to any time point (i.e. simulation, any fraction during RT, or post-RT), as illustrated in Figure 2. However, the aforementioned issues in dealing with volumetric changes and tissue gain/loss will inevitably cause differences in the accumulated dose depending on the direction of dose mapping (79). Rather than considering which direction is "more accurate," we can regard forward- and backward-mapping as two distinct perspectives for understanding accumulated dose, each designed to answer different clinical questions.

When doses are mapped backward and accumulated in the pre-treatment simulation image frame of reference, it is straightforward to compare the planned dose to the accumulated dose over the entire course of RT. For an individual patient, this type of analysis primarily serves to determine whether the intended goals of therapy were met in the aggregate of all adaptive plans. In cases where the intent of adaptive RT was to reduce dose to OARs or escalate dose to the tumor, comparison to the planned dose on the pre-treatment anatomy can also determine whether OAR doses were indeed lowered or whether tumor doses were indeed increased, respectively. One drawback of such a comparison, performed for an individual patient post-RT, is that it is too late to modify the treatment if the intended goals were not met. Still, such comparisons may inform clinicians as to how to approach future patients by establishing a relationship between dose of the day and accumulated dose.

Instead, the post-RT backward-mapped accumulated dose is perhaps a more useful quantity when analyzing side effects and treatment response on large cohorts of patients. Although we have already entered the era of daily MR-guided adaptive replanning, many questions remain regarding the degree of OAR dose sparing and reduction in clinical side effects that can be achieved with this technique, as well as how we can identify the individual patients who stand to benefit the most from treatment on an MR-linac. To answer these questions, we may analyze cumulative delivered dose in the same perspective in which we have historically conceptualized RT: in the frame of reference of the pre-treatment simulation scan. Furthermore, as previously mentioned, the OAR dose constraints and prescription doses that we use for planning are derived from NTCP and TCP models, respectively, which are based on planned doses on the pre-treatment simulation anatomy. In the era of adaptive RT and dose accumulation, more research is needed to determine whether these models remain accurate for delivered dose and whether dose constraints should be redefined in the context of adaptive RT (18, 36).

While this type of post-treatment analysis may be advantageous for answering many research questions, there may still be instances when we want to use a patient's cumulative delivered dose at some point during RT to adapt the remaining fractions. Backward mapping may be used in this scenario, whereby the accumulated dose over a certain number of delivered fractions can be compared to the proportion of the total planned dose for the same number of fractions. However, if the ultimate intention is to use the accumulated dose to modify the treatment plans for subsequent fractions, it may be more appropriate to map the dose forward into the frame of reference in which the next fraction will be planned.

If forward dose mapping can be automated and integrated into the MR-linac on-line clinical workflow, it may radically change how we approach daily MRgART. In current commercial MR-linac systems, adaptive plans are currently generated *via* warm start optimization using the reference plan as a starting point (9, 13, 114). A previous adaptive plan may be used as a new reference plan rather than the pre-treatment reference plan, but the IMRT objectives and dose constraints remain the same for each adaptive plan unless they are manually modified. We have seen with this process that there is a reasonable degree of dosimetric variability between plans from day to day despite the same objectives being used for planning. For example, even if an OAR dose constraint is violated in one or more fractions, that constraint may still be met in the cumulative delivered dose if doses to that OAR fall far enough below the constraint threshold for all other fractions (72). If we could accumulate the dose at each fraction in the frame of reference of the daily setup image, then we could adapt the plan using modified and/or re-prioritized planning objectives based on knowledge of the cumulative delivered dose. In other words, a dose constraint that is routinely met may be de-prioritized in the set of IMRT planning objectives in favor of a dose constraint that is routinely violated. This approach may also be useful if the clinical intent is to escalate dose to the tumor because the physician can make an informed decision about how much the tumor dose can be increased without exceeding the OAR constraints.

In summary, we propose that dose accumulation for purposes of toxicity assessment should be reported on the planning image, as that represents the anatomy at the time of planning and is necessary for useful implementation of toxicity models. However, for ART purposes, we propose that the accumulated dose should be represented on the most recent anatomy so that replanning can be assessed and applied.

As a final note, as with every step of the radiation therapy process, cumulative dose has an associated uncertainty that is a combination of the uncertainties in each step of the dose accumulation process. Understanding these uncertainties and how they impact the use of the cumulative dose is critical to clinical decision making. Thus, prior to clinical implementation of any dose accumulation workflow, it is essential for clinicians to clinicians to fully understand the process and the inherent uncertainties so that they can make the best decisions for their patients.

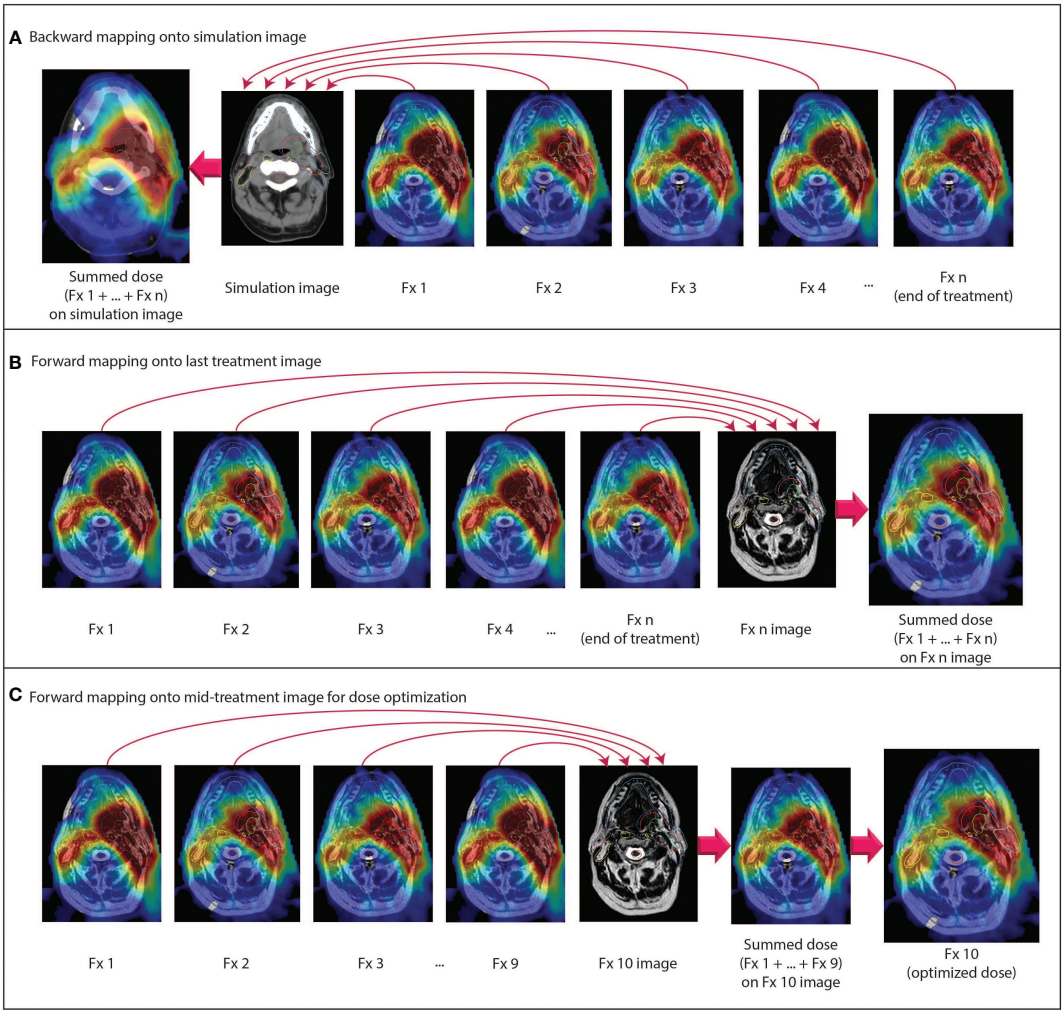


FIGURE 2
Examples of different dose accumulation perspectives. **(A)** Doses from all fractions are mapped backward onto the pre-treatment simulation anatomy. **(B)** Doses from all fractions are mapped forward onto the anatomy from the last fraction. **(C)** At a given time point during RT, all previous doses are mapped forward onto the anatomy for the current fraction. The summed dose is used as an input for dose-optimized adaptive replanning. All doses are scaled to the total prescription dose.

6 Discussion

High-frequency on-treatment imaging and target volume serial assessment with MR-linac devices now affords clinicians the opportunity to move past the historical concept of *planned dose* into an era where *delivered dose* can be used for individual response assessment and dose-optimized adaptive RT. However, the implementation of dose accumulation requires a basic understanding of key considerations and careful validation of dose accumulation solutions tailored to distinct clinical scenarios. In this review article, we have outlined four general steps for dose accumulation—autosegmentation, dose calculation, deformable image registration, and dose mapping/summation—and discussed considerations specific to MR-linac/MRgART workflows. Additional efforts to standardize best practices are imperative to ensure that we move towards a future of adaptively optimized dose as patient-specific precision radiotherapy evolves.

Author contributions

All authors participated in discussions regarding the concept/design and topics reviewed in this manuscript. The manuscript was written by BM, CZ, and MN. All authors contributed to the article and approved the submitted version.

Funding

BM has received research and training support through a Ruth L. Kirschstein National Research Service Award Fellowship from the National Institutes of Health (NIH)/National Institute of Dental and Craniofacial Research (NIDCR) (F31DE029093), a Dr. John J. Kopchick Fellowship from the University of Texas MD Anderson Cancer UTHealth Graduate School of Biomedical Sciences, and an Image-Guided Cancer Therapy T32 Fellowship from MD Anderson Cancer

Center and NIH/National Cancer Institute (NCI). CF received funding from the NIH/NIDCR (1R01DE025248–01/R56DE025248); an NIH/NIDCR Academic-Industrial Partnership Award (R01DE028290); the National Science Foundation (NSF), Division of Mathematical Sciences, Joint NIH/NSF Initiative on Quantitative Approaches to Biomedical Big Data (QuBBD) Grant (NSF 1557679); the NIH Big Data to Knowledge (BD2K) Program of the NCI Early Stage Development of Technologies in Biomedical Computing, Informatics, and Big Data Science Award (1R01CA214825); the NCI Early Phase Clinical Trials in Imaging and Image-Guided Interventions Program (1R01CA218148); an NIH/NCI Pilot Research Program Award from the UT MD Anderson CCSG Radiation Oncology and Cancer Imaging Program (P30CA016672); an NIH/NCI Head and Neck Specialized Programs of Research Excellence (SPORE) Developmental Research Program Award (P50CA097007); and the National Institute of Biomedical Imaging and Bioengineering (NIBIB) Research Education Program (R25EB025787). KB received funding from RaySearch Laboratories AB through a Co-Development and Collaboration Agreement. KB has a licensing agreement with RaySearch Laboratories AB. DT reports institutional collaborations including financial and non-financial support by Elekta, Philips, Kaiku, Therapanacea and PTW Freiburg. JC and EB-R are employees of Elekta AB.

Acknowledgments

The authors would like to thank Amanda Pan and Nicole O'Connell for their input to this manuscript regarding deformable image registration and Brett Sloman, Percy Lee, Kathryn Mittauer, and Anthony Doemer for clarification of vendor-specific MR-linac workflows.

Conflict of interest

MD Anderson Cancer Center and Memorial Sloan Kettering Cancer Centers both have master research agreements with Elekta AB and Philips Healthcare. CF has received direct industry grant support, speaking honoraria, and travel funding from Elekta AB. KB received funding from RaySearch Laboratories AB through a Co-Development and Collaboration Agreement. KB has a licensing agreement with RaySearch Laboratories AB. NT has received travel funding and speaking honoraria from Elekta and Philips. DT reports institutional collaborations including financial and non-financial support by Elekta, Philips, Kaiku, Therapanacea and PTW Freiburg. JC and EB-R are employees of Elekta AB.

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

References

1. Gardner SJ, Kim J, Chetty IJ. Modern radiation therapy planning and delivery. *Hematol Oncol Clin North Am* (2019) 33:947–62. doi: 10.1016/j.hoc.2019.08.005

2. Herrmann H, Seppenwoolde Y, Georg D, Widder J. Image guidance: past and future of radiotherapy. *Radiologe* (2019) 59:21–7. doi: 10.1007/s00117-019-0573-y

3. Raaymakers BW, Jürgenliemk-Schulz IM, Bol GH, Glitzner M, Kotte ANTJ, Van Asselen B, et al. First patients treated with a 1.5 T MRI-linac: Clinical proof of concept of a high-precision, high-field MRI guided radiotherapy treatment. *Phys Med Biol* (2017) 62: L41–50. doi: 10.1088/1361-6560/aa9517

4. Mutic S, Dempsey JF. The ViewRay system: Magnetic resonance-guided and controlled radiotherapy. *Semin Radiat Oncol* (2014) 24:196–9. doi: 10.1016/j.semradonc.2014.02.008

5. Raaymakers BW, Lagendijk JJW, Overweg J, Kok JGM, Raaijmakers AJE, Kerkhof EM, et al. Integrating a 1.5 T MRI scanner with a 6 MV accelerator: Proof of concept. *Phys Med Biol* (2009) 54:N229–37. doi: 10.1088/0031-9155/54/12/N01

6. Kashani R, Olsen JR. Magnetic resonance imaging for target delineation and daily treatment modification. *Semin Radiat Oncol* (2018) 28:178–84. doi: 10.1016/j.semradonc.2018.02.002

7. Lim-Reinders S, Keller BM, Al-Ward S, Sahgal A, Kim A. Online adaptive radiation therapy. *Int J Radiat Oncol Biol Phys* (2017) 99:994–1003. doi: 10.1016/j.ijrobp.2017.04.023

8. Klüter S. Technical design and concept of a 0.35 T MR-linac. *Clin Transl Radiat Oncol* (2019) 18:98–101. doi: 10.1016/j.ctro.2019.04.007

9. Winkel D, Bol GH, Kroon PS, van Asselen B, Hackett SS, Werensteijn-Honingh AM, et al. Adaptive radiotherapy: The elekta unity MR-linac concept. *Clin Transl Radiat Oncol* (2019) 18:54–9. doi: 10.1016/j.ctro.2019.04.001

10. Hall WA, Paulson ES, van der Heide UA, Fuller CD, Raaymakers BW, Lagendijk JJW, et al. The transformation of radiation oncology using real-time magnetic resonance guidance: A review. *Eur J Cancer* (2019) 122:42–52. doi: 10.1016/j.ejca.2019.07.021

11. Menten MJ, Mohajer JK, Nilawar R, Bertholet J, Dunlop A, Pathmanathan AU, et al. Automatic reconstruction of the delivered dose of the day using MR-linac treatment log files and online MR imaging. *Radiother Oncol* (2020) 145:88–94. doi: 10.1016/j.radonc.2019.12.010

12. Heukelom J, Kantor ME, Mohamed ASR, Elhalawani H, Kocak-Uzel E, Lin T, et al. Differences between planned and delivered dose for head and neck cancer, and their consequences for normal tissue complication probability and treatment adaptation. *Radiother Oncol* (2020) 142:100–6. doi: 10.1016/j.radonc.2019.07.034

13. Kontaxis C, Bol GH, Lagendijk JJW, Raaymakers BW. A new methodology for inter- and intrafraction plan adaptation for the MR-linac. *Phys Med Biol* (2015) 60:7485–97. doi: 10.1088/0031-9155/60/19/7485

14. Kontaxis C, de Muinck Keizer DM, Kerkmeijer LGW, Willigenburg T, den Hartogh MD, van der Voort van Zyp JRN, et al. Delivered dose quantification in prostate radiotherapy using online 3D cine imaging and treatment log files on a combined 1.5T magnetic resonance imaging and linear accelerator system. *Phys Imaging Radiat Oncol* (2020) 15:23–9. doi: 10.1016/j.phro.2020.06.005

15. Lowther NJ, Marsh SH, Louwe RJW. Quantifying the dose accumulation uncertainty after deformable image registration in head-and-neck radiotherapy. *Radiother Oncol* (2020) 143:117–25. doi: 10.1016/j.radonc.2019.12.009

16. Mohamed ASR, Bahig H, Aristophanous M, Blanchard P, Kamal M, Ding Y, et al. Prospective in silico study of the feasibility and dosimetric advantages of MRI-guided dose adaptation for human papillomavirus positive oropharyngeal cancer patients compared with standard IMRT. *Clin Transl Radiat Oncol* (2018) 11:11–8. doi: 10.1016/j.ctro.2018.04.005

17. Kerkmeijer LGW, Fuller CD, Verkooijen HM, Verheij M, Choudhury A, Harrington KJ, et al. The MRI-linear accelerator consortium: Evidence-based clinical introduction of an innovation in radiation oncology connecting researchers, methodology, data collection, quality assurance, and technical development. *Front Oncol* (2016) 6:215. doi: 10.3389/fonc.2016.00215

18. Sonke JJ, Aznar M, Rasch C. Adaptive radiotherapy for anatomical changes. *Semin Radiat Oncol* (2019) 29:245–57. doi: 10.1016/j.semradonc.2019.02.007

19. Jaffray DA, Lindsay PE, Brock KK, Deasy JO, Tomé WA. Accurate accumulation of dose for improved understanding of radiation effects in normal tissue. *Int J Radiat Oncol Biol Phys* (2010) 76:S135–9. doi: 10.1016/j.ijrobp.2009.06.093

20. Schwartz DL, Garden AS, Shah SJ, Chronowski G, Sejpal S, Rosenthal DI, et al. Adaptive radiotherapy for head and neck cancer - dosimetric results from a prospective clinical trial. *Radiother Oncol* (2013) 106:80–4. doi: 10.1016/j.radonc.2012.10.010

21. Chetty IJ, Rosu-Bubulac M. Deformable registration for dose accumulation. *Semin Radiat Oncol* (2019) 29:198–208. doi: 10.1016/j.semradonc.2019.02.002

22. Veiga C, Lourenço AM, Mouinuddin S, Van Herk M, Modat M, Ourselin S, et al. Toward adaptive radiotherapy for head and neck patients: Uncertainties in dose warping due to the choice of deformable registration algorithm. *Med Phys* (2015) 42:760–9. doi: 10.1118/1.4905050

23. Rigaud B, Simon A, Castelli J, Gobeli M, Ospina Arango J-D, Cazoulat G, et al. Evaluation of deformable image registration methods for dose monitoring in head and neck radiotherapy. *BioMed Res Int* (2015) 2015:726268. doi: 10.1155/2015/726268

24. Heukelom J, Fuller CD. Head and neck cancer adaptive radiation therapy (ART): Conceptual considerations for the informed clinician. *Semin Radiat Oncol* (2019) 29:258–73. doi: 10.1016/j.semradonc.2019.02.008

25. De La Zerda A, Armbruster B, Xing L. Formulating adaptive radiation therapy (ART) treatment planning into a closed-loop control framework. *Phys Med Biol* (2007) 52:4137–53. doi: 10.1088/0031-9155/52/14/008

26. Liu H, Wu Q. Evaluations of an adaptive planning technique incorporating dose feedback in image-guided radiotherapy of prostate cancer. *Med Phys* (2011) 38:6362–70. doi: 10.1118/1.3658567

27. Kontaxis C, Bol GH, Stemkens B, Glitzner M, Prins FM, Kerkmeijer LGW, et al. Towards fast online intrafraction replanning for free-breathing stereotactic body radiation therapy with the MR-linac. *Phys Med Biol* (2017) 62:7233–48. doi: 10.1088/1361-6560/aa82ae

28. Henke L, Kashani R, Robinson C, Curcuru A, DeWees T, Bradley J, et al. Phase I trial of stereotactic MR-guided online adaptive radiation therapy (SMART) for the treatment of oligometastatic or unresectable primary malignancies of the abdomen. *Radiother Oncol* (2018) 126:519–26. doi: 10.1016/j.radonc.2017.11.032

29. Henke LE, Olsen JR, Contreras JA, Curcuru A, DeWees TA, Green OL, et al. Stereotactic MR-guided online adaptive radiation therapy (SMART) for ultracentral thorax malignancies: Results of a phase 1 trial. *Adv Radiat Oncol* (2019) 4:201–9. doi: 10.1016/j.adro.2018.10.003

30. Koay EJ, Hanania AN, Hall WA, Taniguchi CM, Rebueno N, Myrehaug S, et al. Dose-escalated radiation therapy for pancreatic cancer: A simultaneous integrated boost approach. *Pract Radiat Oncol* (2020) 10:e495–507. doi: 10.1016/j.prro.2020.01.012

31. Bentzen SM, Constine LS, Deasy JO, Eisbruch A, Jackson A, Marks LB, et al. Quantitative analyses of normal tissue effects in the clinic (QUANTEC): An introduction to the scientific issues. *Int J Radiat Oncol Biol Phys* (2010) 76:3–9. doi: 10.1016/j.ijrobp.2009.09.040

32. Marks LB, Yorke ED, Jackson A, Ten Haken RK, Constine LS, Eisbruch A, et al. Use of normal tissue complication probability models in the clinic. *Int J Radiat Oncol Biol Phys* (2010) 76:S10–19. doi: 10.1016/j.ijrobp.2009.07.1754

33. Grimm J, Marks LB, Jackson A, Kavanagh BD, Xue J, Yorke E. High dose per fraction, hypofractionated treatment effects in the clinic (HyTEC): An overview. *Int J Radiat Oncol Biol Phys* (2021) 110:1–10. doi: 10.1016/j.ijrobp.2020.10.039

34. Martel MK, Ten Haken RK, Hazuka MB, Kessler ML, Strawderman M, Turrisi AT, et al. Estimation of tumor control probability model parameters from 3-d dose distributions of non-small cell lung cancer patients. *Lung Cancer* (1999) 24:31–7. doi: 10.1016/S0169-5002(99)00019-7

35. Sanchez-Nieto B, Nahum AE. The delta-TCP concept: A clinically useful measure of tumor control probability. *Int J Radiat Oncol Biol Phys* (1999) 44:369–80. doi: 10.1016/S0360-3016(99)00029-2

36. 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) 3:662–72. doi: 10.1016/j.adro.2018.05.003

37. Hsu S-H, Zawisza I, O'Grady K, Peng Q, Tomé WA. Towards abdominal MRI-based treatment planning using population-based hounsfield units for bulk density assignment. *Phys Med Biol* (2018) 63:155003. doi: 10.1088/1361-6560/aacfb1

38. Jonsson JH, Karlsson MG, Karlsson M, Nyholm T. Treatment planning using MRI data: An analysis of the dose calculation accuracy for different treatment regions. *Radiat Oncol* (2010) 5:1–8. doi: 10.1186/1748-717X-5-62

39. Kim J, Garbarino K, Schultz L, Levin K, Movsas B, Siddiqui MS, et al. Dosimetric evaluation of synthetic CT relative to bulk density assignment-based magnetic resonance-only approaches for prostate radiotherapy. *Radiat Oncol* (2015) 10:1–9. doi: 10.1186/s13014-015-0549-7

40. Hoogcarspel SJ, van der Velden JM, Lagendijk JJW, Van Vulpen M, Raaymakers BW. The feasibility of utilizing pseudo CT-data for online MRI based treatment plan adaptation for a stereotactic radiotherapy treatment of spinal bone metastases. *Phys Med Biol* (2014) 59:7383–91. doi: 10.1088/0031-9155/59/23/7383

41. Prior P, Chen X, Gore E, Johnstone C, Li XA. Technical note: Is bulk electron density assignment appropriate for MRI-only based treatment planning for lung cancer. *Med Phys* (2017) 44:3437–43. doi: 10.1002/mp.12267

42. Ahunbay EE, Thapa R, Chen X, Paulson E, Li XA. A technique to rapidly generate synthetic computed tomography for magnetic resonance imaging–guided online adaptive replanning: An exploratory study. *Int J Radiat Oncol Biol Phys* (2019) 103:1261–70. doi: 10.1016/j.ijrobp.2018.12.008

43. Acharya S, Fischer-Valuck BW, Kashani R, Parikh P, Yang D, Zhao T, et al. Online magnetic resonance image guided adaptive radiation therapy: First clinical applications. *Int J Radiat Oncol Biol Phys* (2016) 94:394–403. doi: 10.1016/j.ijrobp.2015.10.015

44. Kiser K, Meheissen MAM, Mohamed ASR, Kamal M, Ng SP, Elhalawani H, et al. Prospective quantitative quality assurance and deformation estimation of MRI-CT image registration in simulation of head and neck radiotherapy patients. *Clin Transl Radiat Oncol* (2019) 18:120–7. doi: 10.1016/j.ctro.2019.04.018

45. Christiansen RL, Dysager L, Bertelsen AS, Hansen O, Brink C, Bernchou U. Accuracy of automatic deformable structure propagation for high-field MRI guided prostate radiotherapy. *Radiat Oncol* (2020) 15:1–11. doi: 10.1186/s13014-020-1482-y

46. Owrangi AM, Greer PB, Glide-Hurst CK. MRI-Only treatment planning: Benefits and challenges. *Phys Med Biol* (2018) 63:05TR01. doi: 10.1088/1361-6560/aaaca4

47. Wang H, Chandarana H, Block KT, Vahle T, Fenchel M, Das IJ. Dosimetric evaluation of synthetic CT for magnetic resonance-only based radiotherapy planning of lung cancer. *Radiat Oncol* (2017) 12:1–9. doi: 10.1186/s13014-017-0845-5

48. Thorwarth D, Warschburger C, Mönnich D, Grosse U, Kündel M, Nikolaou K, et al. Synthetic CT generation for the pelvic region based on Dixon-MR sequences: Workflow, dosimetric quality and daily patient positioning. *MAGNETOM Flash* (2019) 79:23–7. Available at: https://cdn0.scrvt.com/39b415fb07de4d9656c7b516d8e2d907/1800000006237864/6d626d0a256c/siemens-healthineers-mreadings-edition-5-estro-synthetic-ct-generation_1800000006237864.pdf).

49. Korhonen J, Kapanen M, Keyriläinen J, Seppälä 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) 41:11704. doi: 10.1118/1.4842575

50. 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 (Madr)* (2013) 52:1369–73. doi: 10.3109/0284186X.2013.819119

51. Liu Y, Lei Y, Wang Y, Shafai-Erfani G, Wang T, Tian S, et al. Evaluation of a deep learning-based pelvic synthetic CT generation technique for MRI-based prostate proton treatment planning. *Phys Med Biol* (2019) 64:205022. doi: 10.1088/1361-6560/ab41af

52. Han X. MR-based synthetic CT generation using a deep convolutional neural network method. *Med Phys* (2017) 44:1408–19. doi: 10.1002/mp.12155

53. 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

54. Baydoun A, Xu K, Yang H, Zhou F, Heo JU, Jones RS, et al. Dixon-Based thorax synthetic CT generation using generative adversarial network. *Intell Med* (2020) 3–4:100010. doi: 10.1016/j.ibmed.2020.100010

55. Cho BCJ, van Herk M, Mijnheer BJ, Bartelink H. The effect of set-up uncertainties, contour changes, and tissue inhomogeneities on target dose-volume histograms. *Med Phys* (2002) 29:2305–18. doi: 10.1118/1.1508800

56. Nguyen T, Hoole A, Burnet N, Thomas S. A new method to calculate the dose distribution from an isocenter shift without recalculating dose distribution to evaluate plan with geometric uncertainties. *Med Phys* (2009) 36:2663–3. doi: 10.1118/1.3182100

57. Kessler ML. Image registration and data fusion in radiation therapy. *Br J Radiol* (2006) 79:99–108. doi: 10.1259/bjr/70617164

58. Oliveira FPM, Tavares JMRS. Medical image registration: A review. *Comput Methods Biomech BioMed Engin* (2014) 17:73–93. doi: 10.1080/10255842.2012.670855

59. Sotiras A, Davatzikos C, Paragios N. Deformable medical image registration: A survey. *IEEE Trans Med Imaging* (2013) 32:1153–90. doi: 10.1109/TMI.2013.2265603

60. Unser M, Aldroubi A, Eden M. B-spline signal processing: Part I–theory. *IEEE Trans Signal Process* (1993) 41:821. doi: 10.1109/78.193220

61. Thirion J-P. Image matching as a diffusion process: an analogy with maxwell's demons. *Med Image Anal* (1998) 2:243–60. doi: 10.1016/S1361-8415(98)80022-4

62. Davatzikos C. Spatial transformation and registration of brain images using elastically deformable models. *Comput Vis Image Underst* (1997) 66:207–22. doi: 10.1006/cviu.1997.0605

63. Christensen GE, Rabbitt RD, Miller MI. Deformable templates using large deformation kinematics. *IEEE Trans Image Process* (1996) 5:1435–47. doi: 10.1109/83.536892

64. Brock KK, Sharpe MB, Dawson LA, Kim SM, Jaffray DA. Accuracy of finite element model-based multi-organ deformable image registration. *Med Phys* (2005) 32:1647–59. doi: 10.1118/1.1915012

65. Ashburner J, Friston KJ. Nonlinear spatial normalization using basis functions. *Hum Brain Mapp* (1999) 7:254–66. doi: 10.1002/(SICI)1097-0193(1999)7:4<254::AID-HBM4>3.0.CO;2-G

66. Cideciyan AV. Registration of ocular fundus images: an algorithm using cross-correlation of triple invariant image descriptors. *IEEE Eng Med Biol Mag* (1995) 14:52–8. doi: 10.1109/51.340749

67. Maes F, Collignon A, Vandermeulen D, Marchal G, Suetens P. Multimodality image registration by maximization of mutual information. *IEEE Trans Med Imaging* (1997) 16:187–98. doi: 10.1109/42.563664

68. Haber E, Modersitzki J. Intensity gradient based registration and fusion of multi-modal images. *Methods Inf Med* (2007) 46:292–9. doi: 10.1160/ME9046

69. Spahr N, Thoduka S, Abolmaali N, Kikinis R, Schenk A. Multimodal image registration for liver radioembolization planning and patient assessment. *Int J Comput Assist Radiol Surg* (2019) 14:215–25. doi: 10.1007/s11548-018-1877-5

70. Heinrich MP, Jenkinson M, Bhushan M, Matin T, Gleeson FV, Brady SM, et al. MIND: Modality independent neighbourhood descriptor for multi-modal deformable registration. *Med Image Anal* (2012) 16:1423–35. doi: 10.1016/j.media.2012.05.008

71. Reaungamornrat S, De Silva T, Uneri A, Wolinsky J-P, Khanna AJ, Kleinszig G, et al. MIND demons for MR-to-CT deformable image registration in image-guided spine surgery. *Medical imaging 2016: Image-guided procedures, robotic interventions, and modeling*. RJ W, ZR Y, editors (2016) San Diego, CA, p. 97860H. doi: 10.1117/12.2208621.

72. McDonald BA, Vedam S, Yang J, Wang J, Castillo P, Lee B, et al. Initial feasibility and clinical implementation of daily MR-guided adaptive head and neck cancer radiation therapy on a 1.5T MR-linac system: Prospective r-IDEAL 2a/2b systematic clinical evaluation of technical innovation. *Int J Radiat Oncol* (2021) 109:1606–18. doi: 10.1016/j.ijrobp.2020.12.015

73. Brock KK, Mutic S, McNutt TR, Li H, Kessler ML. Use of image registration and fusion algorithms and techniques in radiotherapy: Report of the AAPM radiation therapy committee task group no. 132: Report. *Med Phys* (2017) 44:e43–76. doi: 10.1002/mp.12256

74. Pace DF, Enquobahrie A, Yang H, Aylward SR, Niethammer M. Deformable image registration of sliding organs using anisotropic diffusive regularization. *Proc - Int Symp BioMed Imaging* (2011), 407–13. doi: 10.1109/ISBI.2011.5872434

75. Zhang Y, Zhang L, Court LE, Balter P, Dong L, Yang J. Tissue-specific deformable image registration using a spatial-contextual filter. *Comput Med Imaging Graph* (2021) 88:101849. doi: 10.1016/j.compmedimag.2020.101849

76. Christiansen RL, Johansen J, Zukauskaite R, Hansen CR, Bertelsen AS, Hansen O, et al. Accuracy of automatic structure propagation for daily magnetic resonance image-guided head and neck radiotherapy. *Acta Oncol (Madr)* (2021) 60:589–97. doi: 10.1080/0284186X.2021.1891282

77. Rosu M, Chetty IJ, Balter JM, Kessler ML, McShan DL, Ten Haken RK. Dose reconstruction in deforming lung anatomy: Dose grid size effects and clinical implications. *Med Phys* (2005) 32:2487–95. doi: 10.1118/1.1949749

78. Schaly B, Kempe JA, Bauman GS, Battista JJ, Van Dyk J. Tracking the dose distribution in radiation therapy by accounting for variable anatomy. *Phys Med Biol* (2004) 49:791–805. doi: 10.1088/0031-9155/49/5/010

79. Zhong H, Chetty IJ. Caution must be exercised when performing deformable dose accumulation for tumors undergoing mass changes during fractionated radiation therapy. *Int J Radiat Oncol Biol Phys* (2017) 97:182–3. doi: 10.1016/j.ijrobp.2016.09.012

80. Schultheiss TE, Tome WA, Orton CG. Point/Counterpoint: It is not appropriate to "deform" dose along with deformable image registration in adaptive radiotherapy. *Med Phys* (2012) 39:6531–3. doi: 10.1118/1.4722968

81. Siebers JV, Zhong H. An energy transfer method for 4D Monte Carlo dose calculation. *Med Phys* (2008) 35:4096–105. doi: 10.1118/1.2968215

82. Zhong H, Siebers JV. Monte Carlo Dose mapping on deforming anatomy. *Phys Med Biol* (2009) 54:5815–30. doi: 10.1088/0031-9155/54/19/010

83. Li HS, Zhong H, Kim J, Glide-Hurst C, Gulam M, Nurushev TS, et al. Direct dose mapping versus energy/mass transfer mapping for 4D dose accumulation: Fundamental differences and dosimetric consequences. *Phys Med Biol* (2014) 59:173–88. doi: 10.1088/0031-9155/59/1/173

84. Ziegenhein P, Kamerling CP, Fast MF, Oelfke U. Real-time energy/mass transfer mapping for online 4D dose reconstruction. *Sci Rep* (2018) 8:1–10. doi: 10.1038/s41598-018-21966-x

85. Paganelli C, Meschini G, Molinelli S, Riboldi M, Baroni G. "Patient-specific validation of deformable image registration in radiation therapy: Overview and caveats." *Med Phys* (2018) 45:e908–22. doi: 10.1002/mp.13162

86. Ahmad S, Khan MF. (2014). Non rigid image registration by modeling deformations as elastic waves, in: 2014 IEEE International Conference on Image Processing (ICIP). pp. 3577–81. Paris, France: IEEE. doi: 10.1109/ICIP.2014.7025726

87. Zachiu C, De Senneville BD, Moonen CTW, Raaymakers BW, Ries M. Anatomically plausible models and quality assurance criteria for online mono- and multi-modal medical image registration. *Phys Med Biol* (2018) 63:155016. doi: 10.1088/1361-6560/aad109

88. Mohammadi R, Mahdavi SR, Jaberi R, Siavashpour Z, Janani L, Meigooni AS, et al. Evaluation of deformable image registration algorithm for determination of accumulated dose for brachytherapy of cervical cancer patients. *J Contemp Brachytherapy* (2019) 11:469–78. doi: 10.5114/jcb.2019.88762

89. Saleh ZH, Apte AP, Sharp GC, Shusharina NP, Wang Y, Veeraraghavan H, et al. The distance discordance metric - a novel approach to quantifying spatial uncertainties in intra- and inter-patient deformable image registration. *Phys Med Biol* (2014) 59:733–46. doi: 10.1088/0031-9155/59/3/733

90. van de Lindt TN, Fast MF, van Kranen SR, Nowee ME, Jansen EM, van der Heide UA, et al. MRI-Guided mid-position liver radiotherapy: Validation of image processing and registration steps. *Radiother Oncol* (2019) 138:132–40. doi: 10.1016/j.radonc.2019.06.007

91. Zachiu C, De Senneville BD, Raaymakers BW, Ries M. Biomechanical quality assurance criteria for deformable image registration algorithms used in radiotherapy guidance. *Phys Med Biol* (2020) 65:015006. doi: 10.1088/1361-6560/ab501d

92. Varadhan R, Karangelis G, Krishnan K, Hui S. A framework for deformable image registration validation in radiotherapy clinical applications. *J Appl Clin Med Phys* (2013) 14:192–213. doi: 10.1120/jacmp.v14i1.4066

93. Schreibmann E, Pantalone P, Waller A, Fox T. A measure to evaluate deformable registration fields in clinical settings. *J Appl Clin Med Phys* (2012) 13:126–39. doi: 10.1120/jacmp.v13i5.3829

94. Cunliffe AR, Contee C, Armato SG, White B, Justusson J, Malik R, et al. Effect of deformable registration on the dose calculated in radiation therapy planning CT scans of lung cancer patients. *Med Phys* (2015) 42:391–9. doi: 10.1118/1.4903267

95. Vickress J, Battista J, Barnett R, Yartsev S. Representing the dosimetric impact of deformable image registration errors. *Phys Med Biol* (2017) 62:N391–403. doi: 10.1088/1361-6560/aa8133

96. Bohoudi O, Lagerwaard FJ, Bruynzeel AME, Niebuhr NI, Johnen W, Senan S, et al. End-to-end empirical validation of dose accumulation in MRI-guided adaptive radiotherapy for prostate cancer using an anthropomorphic deformable pelvis phantom. *Radiother Oncol* (2019) 141:200–7. doi: 10.1016/j.radonc.2019.09.014

97. Niu CJ, Foltz WD, Velec M, Moseley JL, Al-Mayah A, Brock KK. A novel technique to enable experimental validation of deformable dose accumulation. *Med Phys* (2012) 39:765–76. doi: 10.1118/1.3676185

98. Bissonnette JP, Balter PA, Dong L, Langen KM, Lovelock DM, Miften M, et al. Quality assurance for image-guided radiation therapy utilizing CT-based technologies: A report of the AAPM TG-179. *Med Phys* (2012) 39:1946–63. doi: 10.1118/1.3690466

99. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: From error visibility to structural similarity. *IEEE Trans Image Process* (2004) 13:600–12. doi: 10.1109/TIP.2003.819861

100. Zachiu C, De Senneville BD, Tijssen RHN, Kotte ANTJ, Houweling AC, Kerkmeijer LGW, et al. Non-rigid CT/CBCT to CBCT registration for online external beam radiotherapy guidance. *Phys Med Biol* (2018) 63:015027. doi: 10.1088/1361-6560/aa990e

101. Park S, Plishker W, Quon H, Wong J, Shekhar R, Lee J. Deformable registration of CT and cone-beam CT with local intensity matching. *Phys Med Biol* (2017) 62:927–47. doi: 10.1088/1361-6560/aa4f6d

102. Dice LR. Measures of the amount of ecologic association between species. *Ecology* (1945) 26:297–302. doi: 10.2307/1932409

103. Taha AA, Hanbury A. Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool. *BMC Med Imaging* (2015) 15:29. doi: 10.1186/s12880-015-0068-x

104. Jaccard P. The distribution of the flora in the apline zone. *New Phytol* (1912) 11:37–50. doi: 10.1111/j.1469-8137.1912.tb05611.x

105. Cárdenes R, de Luis-García R, Bach-Cuadra M. A multidimensional segmentation evaluation for medical image data. *Comput Methods Programs BioMed* (2009) 96:108–24. doi: 10.1016/j.cmpb.2009.04.009

106. Maurer CR, Michael Fitzpatrick J, Wang MY, Galloway RL, Maciunas RJ, Allen GS. Registration of head volume images using implantable fiducial markers. *IEEE Trans Med Imaging* (1997) 16:447–62. doi: 10.1109/42.611354

107. Bookstein FL, Green WDK. A feature space for edgels in images with landmarks. *J Math Imaging Vis* (1993) 3:231–61. doi: 10.1007/BF01248355

108. Wärmländer SKTS, Garvin H, Guyomarc'h P, Petaros A, Sholts SB. Landmark typology in applied morphometrics studies: What's the point? *Anat Rec* (2019) 302:1144–53. doi: 10.1002/ar.24005

109. Mohatt DJ, Ma T, Wiant DB, Islam NM, Gomez J, Singh AK, et al. Technical and dosimetric implications of respiratory induced density variations in a heterogeneous lung phantom. *Radiat Oncol* (2018) 13:1–14. doi: 10.1186/s13014-018-1110-2

110. de Muinck Keizer DM, Kerkmeijer LGW, Willigenburg T, van Lier ALHMW, Hartogh MDde., JRN vdVv, et al. Prostate intrafraction motion during the preparation and delivery of MR-guided radiotherapy sessions on a 1.5T MR-linac. *Radiother Oncol* (2020) 151:88–94. doi: 10.1016/j.radonc.2020.06.044

111. Stanley N, Glide-Hurst C, Kim J, Adams J, Li S, Wen N, et al. Using patient-specific phantoms to evaluate deformable image registration algorithms for adaptive radiation therapy. *J Appl Clin Med Phys* (2013) 14:177–94. doi: 10.1120/jacmp.v14i6.4363

112. Schnabel JA, Tanner C, Castellano-Smith AD, Degenhard A, Leach MO, Hose DR, et al. Validation of nonrigid image registration using finite-element methods: Application to breast MR images. *IEEE Trans Med Imaging* (2003) 22:238–47. doi: 10.1109/TMI.2002.808367

113. Mariappan YK, Glaser KJ, Ehman RL. Magnetic resonance elastography: A review. *Clin Anat* (2010) 23:497–511. doi: 10.1002/ca.21006

114. Kontaxis C, Bol GH, Lagendijk JJW, Raaymakers BW. Towards adaptive IMRT sequencing for the MR-linac. *Phys Med Biol* (2015) 60:2493–509. doi: 10.1088/0031-9155/60/6/2493

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 31 January 2023
DOI 10.3389/fonc.2023.1107026

OPEN ACCESS

EDITED BY
Jasper Nijkamp,
Aarhus University, Denmark

REVIEWED BY
Jacob Peoples,
Queen's University, Canada
Chao An,
Sun Yat-sen University Cancer Center (SYSUCC), China

*CORRESPONDENCE
Wentao Li
wentaoli@fudan.edu.cn
Yuanyuan Yang
yangyuanyuan@mail.sitp.ac.cn

†These authors have contributed equally to this work

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

RECEIVED 24 November 2022
ACCEPTED 16 January 2023
PUBLISHED 31 January 2023

CITATION
Huang H, Zheng D, Chen H, Chen C, Wang Y, Xu L, Wang Y, He X, Yang Y and Li W (2023) A CT-based radiomics approach to predict immediate response of radiofrequency ablation in colorectal cancer lung metastases.
Front. Oncol. 13:1107026.
doi: 10.3389/fonc.2023.1107026

A CT-based radiomics approach to predict immediate response of radiofrequency ablation in colorectal cancer lung metastases

Haozhe Huang[1,2†], Dezhong Zheng[3,4†], Hong Chen[5†], Chao Chen[1,2], Ying Wang[1,2], Lichao Xu[1,2], Yaohui Wang[1,2], Xinhong He[1,2], Yuanyuan Yang[3,4*] and Wentao Li[1,2*]

[1]Department of Interventional Radiology, Fudan University Shanghai Cancer Center, Shanghai, China, [2]Department of Oncology, Shanghai Medical College, Fudan University, Shanghai, China, [3]Laboratory for Medical Imaging Informatics, Shanghai Institute of Technical Physics, Shanghai, China, [4]Department of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing, China, [5]Department of Medical Imaging, Shanghai Mental Health Center, Shanghai Jiao Tong University School of Medicine, Shanghai, China

Objectives: To objectively and accurately assess the immediate efficacy of radiofrequency ablation (RFA) on colorectal cancer (CRC) lung metastases, the novel multimodal data fusion model based on radiomics features and clinical variables was developed.

Methods: This case-control single-center retrospective study included 479 lung metastases treated with RFA in 198 CRC patients. Clinical and radiological data before and intraoperative computed tomography (CT) scans were retrieved. The relative radiomics features were extracted from pre- and immediate post-RFA CT scans by maximum relevance and minimum redundancy algorithm (MRMRA). The Gaussian mixture model (GMM) was used to divide the data of the training dataset and testing dataset. In the process of modeling in the training set, radiomics model, clinical model and fusion model were built based on a random forest classifier. Finally, verification was carried out on an independent test dataset. The receiver operating characteristic curves (ROC) were drawn based on the obtained predicted scores, and the corresponding area under ROC curve (AUC), accuracy, sensitivity, and specificity were calculated and compared.

Results: Among the 479 pulmonary metastases, 379 had complete response (CR) ablation and 100 had incomplete response ablation. Three hundred eighty-six lesions were selected to construct a training dataset and 93 lesions to construct a testing dataset. The multivariate logistic regression analysis revealed cancer antigen 19-9 (CA19-9, $p<0.001$) and the location of the metastases ($p< 0.05$) as independent risk factors. Significant correlations were observed between complete ablation and 9 radiomics features. The best prediction performance was achieved with the proposed multimodal data fusion model integrating radiomic features and clinical variables with the highest accuracy (82.6%), AUC value (0.921), sensitivity (80.3%), and specificity (81.4%).

Conclusion: This novel multimodal data fusion model was demonstrated efficient for immediate efficacy evaluation after RFA for CRC lung metastases, which could benefit necessary complementary treatment.

KEYWORDS

computed tomography (CT), radiomics, clinical variables, colorectal cancer, lung metastasis, radiofrequency ablation (RFA), efficacy evaluation

Introduction

Colorectal cancer (CRC) is one of the most common malignant tumors and a leading cause of cancer-related mortality worldwide (1). About 25% of CRC patients present with distant metastases at the time of initial diagnosis, with the most common sites including liver and lung (2, 3). In addition, patients with rectal cancer are more likely to have lung metastases because of anatomical differences (4, 5). However, not all patients meet the criteria for surgical resection due to lesion location, tumor burden, comorbidity, or the presence of extra-pulmonary disease. For this group, thermal ablation, including radiofrequency (RFA) or microwave (MWA), is considered a safe alternative (6).

RFA has been proven safety and efficacy in lung metastases from CRC (7–9). However, there is no pathological histological evidence of complete ablation after RFA, and recent studies demonstrated that the incomplete RFA promoted increased tumorigenesis (10) and hindered the efficacy of anti-programmed cell death protein-1 immunotherapy (11). In addition, the existence of remnant tumor masses was associated with earlier new metastases and poor survival (11). Therefore, it is crucial to clarify the local recurrence factors and assess the early-stage efficacy. To achieve complete ablation of lung cancer, any peritumoral lung parenchyma within 5 to 10 mm needs to be ablated (12–14). This area presents as necrosis, effusion and congestion from the inner zone to the outer zone on histopathology, accordingly (15), and manifests as ground-glass opacity (GGO) on CT, which is the typical post-ablation presentation and the crucial area in the early assessment after RFA (16). Previous studies based on the morphological changes of unenhanced CT found that the size of GGO was associated with residual tumor and recurrence (17, 18). However, intraoperative complications such as intra-alveolar hemorrhage (IAH) or atelectasis, make it impossible to determine the extent of ablation (9, 12, 19, 20). Therefore, the observation and measurement of the intraoperative GGO range to ascertain whether ablation is complete is subjective and uncertain as such an approach is easily influenced by doctors with differences experience.

The modified response evaluation criteria in solid tumors (mRECIST) are used to evaluate the efficacy of lung tumor ablation (21–23). However, the inflammatory response surrounding the lesion make it difficult to clearly evaluate the early efficacy. The lesions do not stabilize or shrink until at least six months after ablation, eventually manifesting in the form of disappearance, fibrosis, nodules, and cavities (24, 25). This time-lapse evaluation method may also result in a missed opportunity for the optimal complementary therapy for patients, thus affecting their survival benefits. Therefore, there is an urgent need for objective and reliable characteristic metrics or models to evaluate the immediate ablative efficacy of RFA for pulmonary metastases.

Radiomics can mine high-dimensional quantitative imaging features of medical images, which contain information related to tumor heterogeneity and microenvironment (26–29), allowing for more accurate quantification of phenotypic features and assessment of treatment response (30–33). Radiomics analysis includes target lesion segmentation, feature extraction, machine learning classifier training, and performance evaluation (34–36). However, the radiomics feature analysis approach just takes full advantage of a single mode of radiological data which is incomplete and noisy whilst ignoring other modalities data, such as histopathology, genomics, or clinical information, leaving multimodal data integration relatively underdeveloped (37).

In this study, we developed novel multimodal data fusion models integrating radiomics features based on radiological data with clinical variables originating from textual data to assist interventional physicians in evaluating the immediate efficacy of RFA for CRC lung metastases, so as to make necessary supplementary treatment during operation.

Materials and methods

Data collection

CRC patients with lung metastases who underwent percutaneous RFA under CT guidance between August 2016 and January 2019 were enrolled in this study. Patients were recruited based on the following eligibility criteria: (1) histologically confirmed CRC; (2) ablated lung metastases with maximum diameter ≤3 cm; (3) chest enhanced CT examination within 4-6 weeks before RFA; (4) complete CT images during the procedure; (5) re-examination by chest enhanced CT at least 6 months after RFA; (6) technically successful ablation; (7) adequate normal organ function. Exclusion criteria, based on the European Society for Medical Oncology (ESMO) guidelines (38) were: (1) > 5 lung metastases; (2) maximum diameter > 3 cm; (3) other local or regional treatments such as radiotherapy before or after RFA; (4) incomplete clinical data; (5) second ablation (i.e., re-ablation). We allowed the inclusion of patients with multiple nodules and analyzed each nodule individually. A cohort of 198

patients with 479 lung metastases who received RFA was retrospectively selected (Figure 1).

The following clinical data were retrieved: age at diagnosis, gender, serum tumor markers, including carcinoembryonic antigen (CEA) and cancer antigen 19-9 (CA19-9). The radiological data were recorded as follows: the location of pulmonary metastases, proximity to the heart, great blood vessels (diameter > 3 mm), pleura or diaphragm (within 1 cm) through the preoperative CT images; the IAH or pneumothorax were acquired.

All CT examinations (United Imaging uCT 760, Shanghai United Imaging Medical Technology Inc., China and Philips Brilliance 64 slice, Philips Medical Systems Inc., USA) were performed with a fixed tube current of 200 mA and a tube voltage of 120 kVp. The pixel spacing ranged from 0.684 to 0.748 mm, and the slice thickness was 1 mm or 1.5 mm. The intraoperative CT images were of fixed tube current of 200 mA, tube voltage of 120 kVp, and slice thickness of 1 mm or 3 mm. The image reconstruction method of both CT scanners is iterative reconstruction. Radiological follow-up consisted of chest-enhanced CT scans performed at 1, 3, 6, 12 months, and every 6 months after that. The shortest follow-up time was over 6 months.

This study was approved by the Institutional Review Board of the Ethics Committee of Fudan University Shanghai Cancer Center. Written informed consent was obtained from all patients.

RFA procedures and local efficacy assessment

RFA mainly utilizes 460 ~ 480 kHz high-frequency current to heat a tissue volume around a needle electrode and induce focal coagulative necrosis with minimal injury to surrounding tissues (39, 40). Here, RFA was performed using a radiofrequency applicator (MedSphere International), with the mode of temperature control or impedance control for choice. The power settings were adjusted according to the manufacturer's protocols: 5 min for a 2.0 ~ 2.5 cm active tip at 30 W, 8 min for a 3.0 ~ 3.7 cm active tip at 50 W, and 10 min for a 4.0 ~ 4.7 cm active tip at 60 W, respectively.

All the operations were performed by three senior interventional radiologists (L.X., Y.W. and X.H. with over 10 years of experience in thoracic interventions under CT guidance). Depending on the location of the target nodules, patients were placed in a prone position, lateral position or supine position to ensure the best puncture site and entry route and avoid important structures, including ribs, interlobular fissures, and blood vessels. Lidocaine was administered at the puncture site to induce local anesthesia of the pleura. With CT monitoring, the radiofrequency electrode was punctured according to the predetermined direction and angle. The ablation was not performed until the CT scan confirmed that the electrode hooked the lesion. Considering tumor shape and size, one or two needle ablations with a constant antenna position were usually acceptable to achieve complete ablation. The operators strived to achieve ablation range greater than the lesions by at least 5 mm. If the intraoperative complications such as intra-alveolar hemorrhage (IAH) or atelectasis, made it impossible to determine the extent of ablation, at least 2 cycles of ablation would be performed to raise the impedance until ablation stopped. After completion of the RFA session, the ablation electrode was withdrawn, and a repeat CT (same parameters) scan was performed to evaluate whether the ablation zones covered the tumor and the occurrence of ablation-related complications, mainly including pneumothorax and hemorrhage.

Local efficacy was assessed by two radiologists who were blind to clinical data (H.C. and H.H. with over 5 years of experience) through chest enhanced CT examination at least 6 months after RFA according to mRECIST criteria (24, 41). If they had disagreements, it would be determined in consultation with the senior expert (W.L. with over 20 years of experience). The follow-up CT examination one month after ablation was taken as the baseline (42). Based on the mRECIST criteria, CR was defined if any of the following manifestations on CT were seen: the disappearance of the lesion, cavity, fibrosis or nodule without enhancement. If two consecutive CT

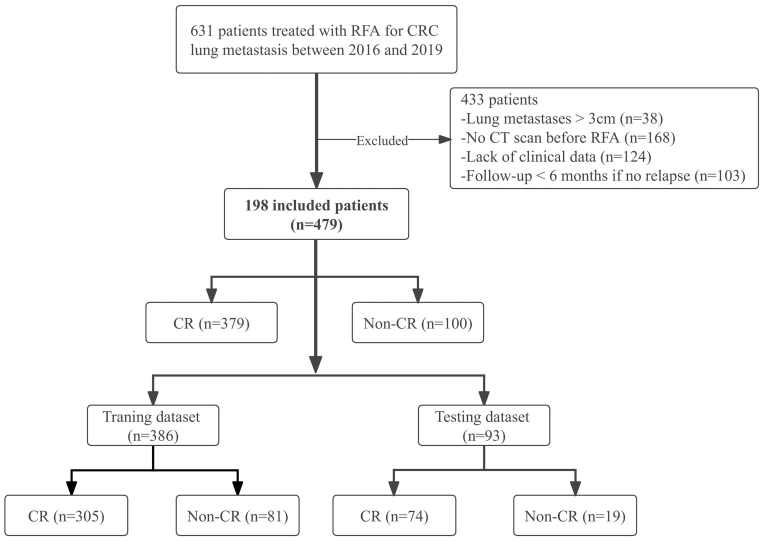


FIGURE 1
Study flow chart. CRC, colorectal cancer; CR, complete response; Non-CR, Non-complete response.

examinations demonstrated the target lesions had irregular enlargement or enhanced solid components, they were classified as a non-complete response (non-CR).

Pre-processing of CT images, radiomics feature extraction, selection and data division

In order to avoid the data bias due to the difference in scanning spacing and slice thickness between preoperative and immediately postoperative CT images, the following preprocessing steps were adopted: the CT images were uniformed to a common resolution of 1 mm × 1 mm × 1 mm by B-spline interpolation algorithm, and then the window width was adjusted within the range of - 1200 Hu to 600 Hu and the intensity was scaled within the range of 0 ~ 255. After normalization of all CT images, the samples containing pulmonary metastases were trimmed to 3D cubes with the size of 40 mm × 40 mm × 40 mm. Finally, the gray values of sample cubes were normalized between 0 and 1 (Figure 2).

To objectively and accurately delineate the target lesions in the preoperative CT images and the boundary of the ablation area immediately after RFA, A 3D U-Net model (43) was used to segment the lesions and ablation region automatically, and two junior radiologists (H.C.and H.H.) verified the segmentations and made the necessary adjustments to guarantee the accurancy and repeatability (Figure 3). If they had disagreements, it would be determined in consultation with the senior expert (W.L.).

For each segmented preoperative lesions and ablation region, 1252 radiomics features were extracted through the open-source feature toolboxes PyRadiomics (44) and PREDICT. The radiomics features were comprised of 13 intensity features, 35 shape features, 9 orientation features and 507 texture features which contained 144 Gray Level Co-occurence Matrix (GLCM) features, 16 Gray Level Size Zone Matrix (GLSZM) features, 16 Gray Level Run Length Matrix (GLRLM) features, 14 Gray Level Dependence Matrix (GLDM) features, 5 Neighborhood Grey Tone Difference Matrix (NGTDM) features, 156 Gabor filters features, 39 Laplacian of Gaussian (LoG) filters features, 39 Local Binary Patterns (LBP) features (32, 44–47), and 688 wavelet features.

In order to reduce unnecessary, redundant information and complexity in the process of calculation and modeling, the maximum correlation and minimum redundancy algorithm (MRMRA) (48) was used for features selection. There are five common variants under the MRMRA framework (49): mutual information difference (MID), mutual information quotient (MIQ), F-test correlation difference (FCD), F-test correlation quotient (FCQ), and random-forest correlation quotient (RFCQ). The formulas were as follows:

Assuming that there were m features in total, for a given feature Xi, i∈ (1, 2,…, m), the importance of the feature could be determined by MRMRA, commonly in the following five forms:

$$f^{MID}(X_i) = I(Y, X_i) - \frac{1}{|S|}\sum X_s \in SI(X_s, X_i)$$
$$f^{MIQ}(X_i) = I(Y, X_i)/[\frac{1}{|S|}\sum X_s \in SI(X_S, X_i)]$$
$$f^{FCD}(X_i) = F(Y, X_i) - \frac{1}{|S|}\sum X_S \in SP(X_S, X_i)$$
$$f^{FCQ}(X_i) = F(Y, X_i)/[\frac{1}{|S|}\sum X_S \in SP(X_S, X_i)]$$
$$f^{RCQ}(X_i) = I_{RF}(Y, X_i)/[\frac{1}{|S|}\sum X_S \in SP(X_S, X_i)]$$

where, Y is the category label corresponding to the variable, S is the selected feature set, $|S|$ is the size of the feature set, $Xs \in S$ is a feature outside the feature set S, and Xi represents a feature that was not currently selected; the function $I(\cdot,\cdot)$ represents mutual information, $p(\cdot,\cdot)$ is the Pearson correlation coefficient, $F(\cdot,\cdot)$ is the Fstatistics, and $I_{RF}(\cdot,\cdot)$ is the random forest feature importance score. Since inconsistent results of various methods under different super parameter conditions, we utilized the above 5 methods to filter features. The frequency of the top 5, top 10, and top 15 features was counted in the importance ranking, and the experiments were conducted from 5 to 15 features with the highest frequency to obtain the best performance, and eventually to confirm the 9 selected features.

As the Gaussian mixture model (GMM) had good performance in the evaluation of sample distribution and similarity in high-

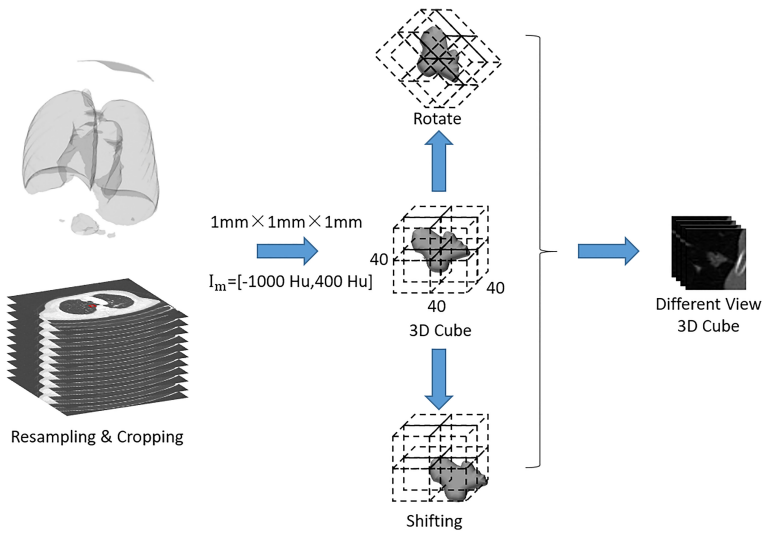


FIGURE 2
Flow chart of image preprocessing.

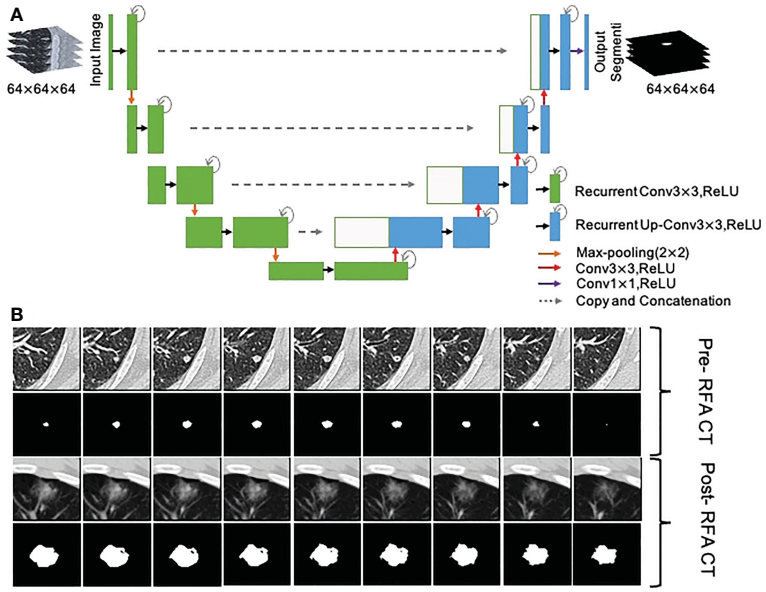


FIGURE 3
Segmentation of metastasis and ablation area. **(A)** Flow chart of the 3D U-Net model; **(B)** CT images and segmentation images of colorectal cancer lung metastasis and ablation area immediate after RFA.

dimensional space (50–52), we used distance metric learning based on the Gaussian mixed model (DML-GMM) rather than random splitting to divide data according to our previous research results (53). We demonstrated that when the sample size was large, there was little difference between random splitting and the DML-GMM model. As for a smaller sample size, however, the DML-GMM model could obtain more stable results than random splitting. Therefore, the log-likelihood of the extracted radiomic features was calculated by DML-GMM model to describe the distribution, the data was split into multiple clusters and then was divided into 5 groups by stratified sampling. One group was selected as the testing set and the remaining 4 groups were used as the training set. Actually, we did five-fold cross-validation and chose a single split data including 386 lesions for the training set and 93 for the testing set.

Model building and performance evaluation

Due to the unbalanced distribution of case counts in CR and non-CR, we adopted the oversampling method (synthetic minority over-sampling technique, SMOTE) (54) to mitigate the biased impact of data imbalance on the models during training.

Clinical model: all the clinical and radiological features were included in the univariate Logistic regression analysis, after which the variables with P< 0.1 were included in the multivariate analysis. Finally, the independent factors with P< 0.05 were selected for modeling. The random forest technique was a regression tree technique which utilized bootstrap aggregation and randomization of predictors to achieve a high degree of predictive accuracy (55). Since the random forest algorithm has been proven to be effective and superior in building clinical models (56, 57), our clinical model was also built on it.

Radiomics model: A random forest, which was the most common classifier used for radiomics features classification, contained multiple decision trees, and the total output result was determined by the subcategories of each decision tree. When processing high-dimensional data, it had a strong ability of anti-interference and anti-overfitting, especially for unbalanced medical data. Several studies have confirmed that the random forest model could be used to predict the survival rate, recurrence risk, and efficacy evaluation of lung cancer patients (58–60).Multimodal data fusion models: the random forest model was integrated based on radiomics and clinical models (Figure 4). The weighted fusion strategy (61) adopted in our study was decision level fusion (late fusion) (62). This level of fusion allowed features from different representations to be combined in the same format of representation, which had more and better scalability and flexibility (63). The exact formula was $confidence = \omega_1 \cdot confidence_{image} + \omega_2 \cdot confidence_{clinical}$. The fusion prediction score was calculated to obtain the final prediction result.

In order to evaluate the performance of various models, we validated them on an independent test dataset, drew receiver operating characteristic curves (ROC) with the obtained prediction scores and calculated the corresponding area under curve (AUC). The difference in the predictive performance of models was compared by the Delong test (64). Meanwhile, the accuracy (ACC = $\frac{TP+TN}{TP+FP+TN+FN}$), sensitivity (Sensitivity = $\frac{TP}{TP+FN}$), and specificity (Specificity = $\frac{TN}{TN+FP}$) were also calculated. In the formula, TP was true positive, FP was false positive, TN was true negative and FN was false negative.

Statistical analysis

Statistical analyses were performed using IBM SPSS (version 26.0, Chicago, USA). Man-Whitney U test was used for continuous variables which were presented as mean ± standard deviation (SD).

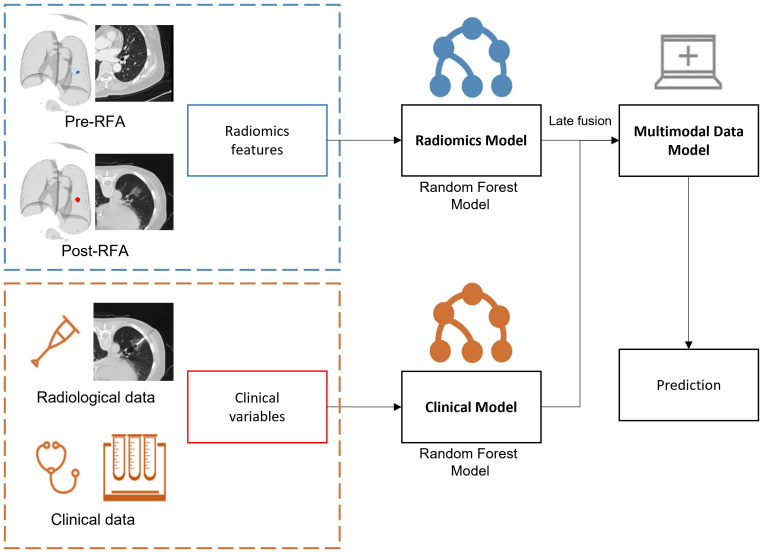


FIGURE 4
Fusion framework of radiomics features and clinical information.

Chi-square or Fisher test was used for categorical variables. All statistical tests were conducted at a two-sided significance level of P<0.05. All the medical image processing procedures and evaluation processes were performed on Python 3.6. In order to build the models and calculate the evaluation scores, we used publicly available packages such as SimpleITK, PyTorch, scikit-learn, numpy, and scipy.

Results

Characteristics of patients and lesions

A total of 198 patients with 479 lung metastases from CRC were enrolled; the detailed demographic characteristics are listed in Table 1. After RFA treatment, there were 379 CR lesions and 100 non-CR lesions. Due to the small sample size, we analyzed each lesion individually in the same patient with multiple metastases as the recent literatures (20, 65–67). Through the GMM method, 386 lesions (305 CR and 81 non-CR) were selected to constitute the training dataset, and 93 lesions (74 CR and 19 non-CR) were chosen to constitute the independent testing dataset (Figure 1). There were 227 lesions (47.4%)< 10 mm, and most lesions (399, 83.3%) were not close to the mediastinum or great vessels (diameter greater than 3 mm), but close to the pleura or diaphragm (287, 59.9%). The incidences of IAH and pneumothorax were 25.9% (124/479) and 24.0% (115/479), respectively.

Clinical and radiomics feature selection

Univariate logistic regression analysis in Table 2 showed that CEA, CA19-9, lesion location (including upper lobe of the right lung, right lower lobe, and left lower lobe), and intra-alveolar hemorrhage (P< 0.1) could completely identify ablated lesions. Furthermore, multivariate regression analysis demonstrated that CA19-9 (odds ratio [OR] = 1.007, P< 0.001) and lesion location (including right upper lobe [OR = 1, P = 0.005], right lower lobe [OR = 2.997, P = 0.003], and left lower lobe [OR = 2.498, P = 0.011]) were independent risk factors for incomplete ablation. These two clinical variables were used to construct a clinical model.

In order to prevent the model from overfitting because of the small sample size, 5 to 15 vital features with the highest scores were selected by MRMRA, and the five forms of MRMRA features importance scores were calculated separately, and compared with the default important feature s of the random forest model as the benchmark. The results demonstrated that the important features selected by MRMRA in the form of MID, MIQ, FCQ, and RFCQ had better performance than the features automatically selected by random forest, and the experimental model with 9 selected features had achieved better stability and smaller deviation. The selected feature results are shown in Table 3.

Prediction performance comparison

The AUC values of each model were calculated in an independent testing dataset, and the DeLong test compared the corresponding P values (Tables 4, 5 and Figure 5). When radiomics features were integrated with clinical variables, and the coefficient of the radiomics model was 0.7 and the coefficient of clinical model was 0.3, the resulting AUC value achieved the highest (0.921) with the statistically significant difference (P values of 0.043) compared with the clinical model alone (0.830). In addition, the accuracy, sensitivity, and specificity of this multimodal data fusion model were also the best (82.6%, 80.3%, and 81.4%, respectively).

Figure 6 presents one example of a patient with post-lung RFA CR, with a nodule in contact with a vessel, complicated by IAH. In contrast, Figure 7 illustrates another example with post-lung RFA non-CR.

10.3389/fonc.2023.1107026

TABLE 1 Characteristics of patients and colorectal cancer lung metastases.

Characteristics		Training dataset(N= 386)	Testing dataset(N=93)	P	Total(N=479)
Pre-RFA clinical features					
Gender	Male	221	48	0.0347	269
	Female	165	45		210
Age (years, mean ± SD)		57.9±10.3	57.3±11.1	0.5860	
Tumor markers	CEA (ng/ml)	4.6±4.1	3.9±3.6	0.4594	
	CA19-9 (U/ml)	10.8±6.2	10.9±6.4	0.8112	
Pre-RFA characteristics of the lung metastasis					
Nodule size (mm)	< 10	186	41	0.0622	227
	10–19	141	37		178
	20 - 30	59	15		74
Location	RUL	91	21	0.0011	112
	RML	41	11		52
	RLL	76	12		88
	LUL	72	36		108
	LLL	106	13		119
Distance 1 (cm)	> 1	329	70	0.3997	399
	< 1	57	23		80
Distance 2 (cm)	> 1	151	41	0.0739	192
	< 1	235	52		287
Immediate post-RFA features					
Pneumothorax	Yes	92	23	0.2932	115
	No	294	70		364
Intra-alveolar	Yes	100	24	0.2654	124
Hemorrhage	No	286	69		355

RUL, right upper lobe; RML, right middle lobe; RLL, right lower lobe; LUL, left upper lobe; LLL, left lower lobe; SD, standard deviation; Distance 1, the distance between the lesion and the large vessels or mediastinum; Distance 2, the distance between the lesion and the pleura or diaphragm.

Discussion

In 2016, the ESMO proposed a toolbox for oligometastases of CRC, which emphasized the clinical value of local therapy (38). In patients who are not eligible for surgery, RFA seems to have more evidence as a locoregional alternative for tumors< 3 cm (6).

After ablation, lung tumors undergo a natural evolution of the outcome process: in the early stage (within 1 week), the lesions are

TABLE 2 Uni- and multi-variate analysis of clinical and radiological characteristics.

Characteristics		Univariate analysis		Multivariate analysis	
		OR (95% CI)	P	OR (95% CI)	P
Clinical features					
Gender	Male	1.259 (0.803-1.974)	0.316		
	Female	1			
Age		0.994 (0.974-1.015)	0.588		
Tumor biomarkers	CEA	1.007 (1.001-1.012)	**0.028**	1.004 (0.999-1.009)	0.166
	CA 19-9	1.006 (1.002-1.009)	**0.001**	1.007 (1.003-1.011)	**<0.001**

(Continued)

TABLE 2 Continued

Characteristics		Univariate analysis		Multivariate analysis	
		OR (95% CI)	P	OR (95% CI)	P
Pre-RFA features of the lung metastases					
Location	RUL	1	**0.013**	1	**0.005**
	RML	1.323 (0.538-3.253)	0.542	0.949 (0.357-2.525)	0.917
	RLL	2.968 (1.468-6.002)	**0.002**	2.997 (1.442-6.23)	**0.003**
	LUL	1.293 (0.615-2.718)	0.497	1.216 (0.557-2.654)	0.624
	LLL	2.23 (1.125-4.419)	**0.022**	2.498 (1.228-5.08)	**0.011**
Distance 1 (cm)	> 1	1			
	< 1	0.693 (0.402-1.197)	0.189		
Distance 2 (cm)	> 1	1			
	< 1	0.957 (0.608-1.506)	0.848		
Immediate post-RFA features					
Pneumothorax	Yes	1.08 (0.652-1.789)	0.764		
	No	1			
IAH	Yes	0.612 (0.354-1.059)	**0.079**	0.644 (0.364-1.138)	0.130
	No	1		1	

The bold p vales in the univariate analysis (in the first column) mean < 0.1, and those in the multivariate analysis (in the second column) mean < 0.05.

covered by GGO, with larger scopes than the lesions, and the interior of the lesions presents a low-density honeycomb appearance. In the middle stage (1 week to 2 months), the ablation area becomes larger, and an enhanced ring appears due to the absorption of inflammation around the lesion. Finally, in the late stage (after 2 months), the ablation area remains relatively stable or slightly larger, gradually shrinking or stabilizing after 6 months (24, 25). Therefore, contrast-enhanced CT of the chest at least 6 months after RFA was used to evaluate the efficacy of RFA in this study, so as to determine whether the lesions were completely ablated.

We found that the level of CA19-9 and location of the metastases were significant correlations with complete ablation. In terms of recurrence and survival prognosis, the combined evaluation of CEA and CA19-9 could obtain more relevant information than the evaluation of CEA alone (68, 69). However, this study found no significant association between CEA levels and complete ablation based on the multivariate logistic regression analysis. On the other hand, the location of nodules, including lung lower lobes, was an independent risk factor with values of OR > 2, possibly due to the influence of patient's respiratory movement on the correct positioning of the probe. Also, IAH was associated with a higher risk of local recurrence, which reached significance in the univariate analysis, likely because of the increasing difficulties in locating the target nodule in the background of the dense and radiopaque zone. In addition, the heat sink effect was associated with a higher risk of incomplete ablation for tumors with blood vessel contact resulting from the blood flow and microscopic extension (12, 70, 71). However, our variables relating to vessels did not reach significance, probably with the influence caused by the enrolled cases close to the mediastinum.

TABLE 3 Radiomics features selected by MRMRA.

pre-RFA radiomics features	post-RFA radiomics features
shape_Elongation	GLCM_Idmn
GLCM_Idmn	GLRLM_RunEntropy
GLCM_Imc1	GLCM_Imc2
GLCM_InverseVariance	
GLCM_ClusterShade	
GLDM_DependenceEntropy	

MRMRA, maximum relevance and minimum redundancy algorithm; RFA, radiofrequency ablation; GLCM, Gray Level Co-occurence Matrix; GLDM, Gray Level Dependence Matrix; GLRLM, Gray Level Run Length Matrix.

In contrast to conventional CT-based imaging features, radiomics analysis enables a greater degree of information reflecting underlying biologic heterogeneity to be derived and qualified at a low cost (27, 46). A radiomics signature, as a panel of multiple features, has been regarded as a more powerful prognostic biomarker, which could provide additional information to clinical data, and has reportedly been a significant predictor for clinically relevant factors (72–74). Previous studies have demonstrated that the size and shape of metastases are the important risk factors for local recurrence (9, 12, 75), as the shape feature selected by MRMRA. In addition, GLRLM features could quantify gray level runs, defined as the length in a

TABLE 4 AUC values of different models in the testing dataset.

Models	AUC
Clinical	0.830
Radiomics	0.887
Radiomics + clinical	0.921
0.1× Radiomics + 0.9 × clinical	0.839
0.2× Radiomics + 0.8 × clinical	0.852
0.3× Radiomics + 0.7 × clinical	0.869
0.4× Radiomics + 0.6 × clinical	0.885
0.5× Radiomics + 0.5 × clinical	0.904
0.6× Radiomics + 0.4 × clinical	0.913
0.7× Radiomics + 0.3 × clinical	**0.921**
0.8× Radiomics + 0.2 × clinical	0.916
0.9× Radiomics + 0.1 × clinical	0.903

The bold value means the highest AUC value of the best model.

TABLE 5 Comparison of prediction performance of different models in the testing dataset.

Models	ACC (%)	AUC	Sensitivity (%)	Specificity (%)
Clinical	71.4	0.830	69.6	75.3
Radiomics	80.8	0.887	79.1	80.6
Radiomics + Clinical	**82.6**	**0.921**	**80.3**	**81.4**

ACC, accuracy; AUC, area under ROC curve.
The bold values mean the best performance of the multimodal data fusion model integrating radiomic features and clinical variables.

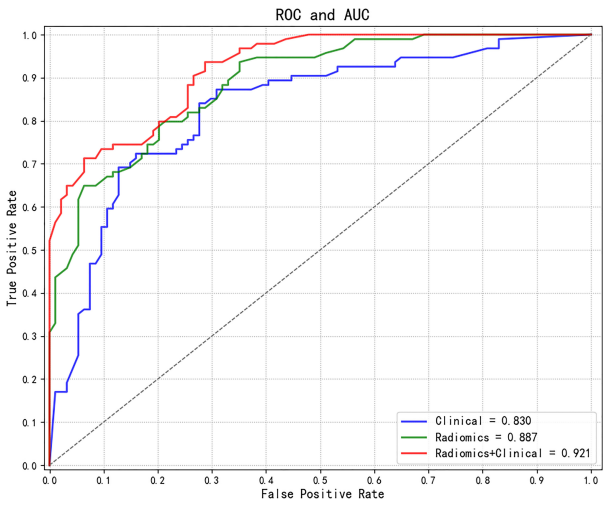


FIGURE 5
Comparisons of ROC curves of different models. ROC, receiver operating characteristic; AUC, area under the curve.

number of consecutive pixels that have the same gray level value (76) and could reflect the volumetric texture of the early ablation zone (77). A GLCM feature, which could reflect and quantify homogeneity to reflect the risk of local recurrence, is a common method of describing texture by studying the spatial correlation characteristics of gray levels (78). A GLDM feature quantifies gray level dependencies, which correspond the number of connected voxels within distance δ dependent on the center voxel (79), likely reflecting the difficulties in identifying the nodule.

Recent work has highlighted important efficacy and prognostic information captured in radiological, clinicogenomic, and histopathological data, which can be exploited through machine learning. However, little is known about the capacity of combining features from these disparate sources to improve the prediction of treatment response. Therefore, we combined radiomics with patients' clinical variables to construct multimodal data fusion models to objectively and accurately evaluate the immediate efficacy of RFA for CRC lung metastases.

An observer study was conducted by testing an independent dataset to validate the performance of models (i.e., results shown in Tables 4, 5 and Figure 5). Compared with the baseline model only based on clinical variables, the radiomics-based models showed further improvement in performance with a significant statistical difference ($P<0.05$). Compared with the model only based on radiomics features, the corresponding performance indicators of the multimodal data fusion model (Radiomics + Clinical) were higher, but the Delong test did not confirm significant difference ($P>0.05$) between the models, indicating that the radiomics features have a dominant role in the models. At the same time, it suggests that the clinical variables could provide supplementary information to improve the predictive performance of the models, although they could not reach significance, possibly because of the limited sample size.

The main advantages of this study are as follows: first of all, different types of data might contain complementary information;

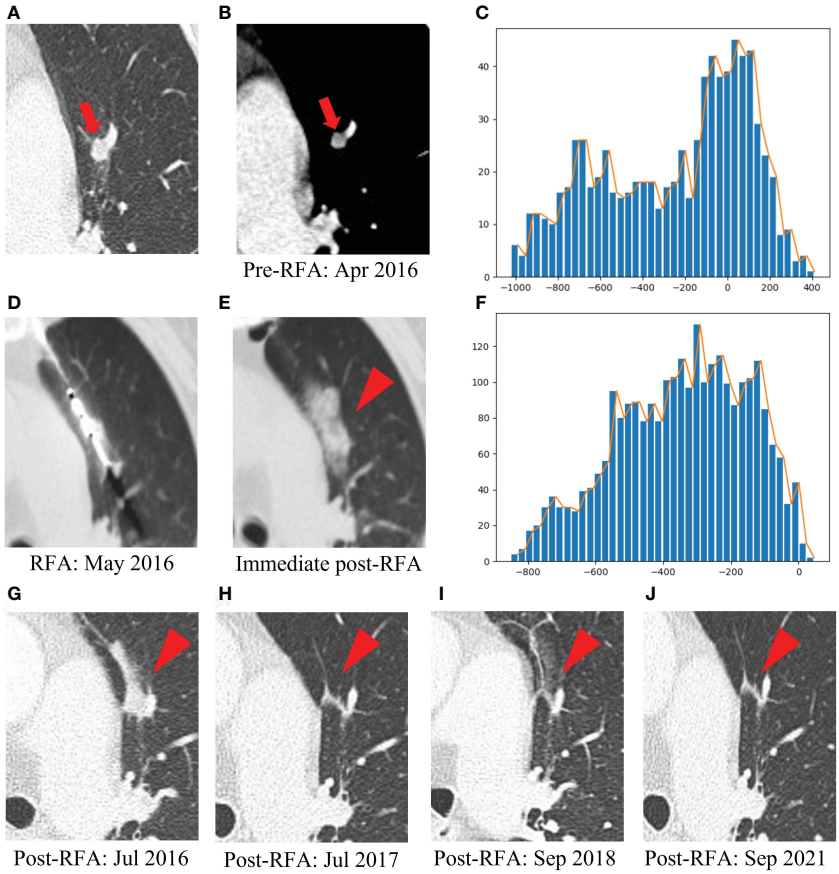


FIGURE 6
Example of a patient with post-lung RFA complete response. The multimodal data fusion model predicted the results: CR: 0.87, Non-CR: 0.13. **(A)** A 77-year-old female pT2N1M1R0 rectal cancer patient with a lung metastasis one year after rection, which was located in the left upper lobe (red arrow). **(B)** The mediastinal window of enhanced CT showed that the lesion was enhanced and adjacent to the blood vessel, with a maximum diameter of 6 mm (red arrow). **(C)** The histogram of the densities within this nodule on the pre-RFA CT scans displayed an asymmetric, skewed distribution corresponding to intra-tumoral enhancement (x-axis: attenuation in Hounsfield units, y-axis: number of voxels). **(D)** The RFA was performed under CT guidance. **(E)** IAH occurred after RFA (red arrowhead). **(F)** The histogram of the densities within the ablation zone on the immediate post-RFA CT scans was rather flat, without peak among the high tissular atenuations (x-axis: attenuation in Hounsfield units, y-axis: number of voxels). **(G)** Chest CT scan showed high density patch shadow in the ablation area one month after RFA (red arrowhead). **(H–J)** One year, two years and five years after RFA, chest CT scans showed that the lesion disappeared (red arrowhead). RFA, radiofrequency ablation; CR, complete response; IAH, intra-alveolar hemorrhage.

therefore, we developed novel multimodal data fusion models integrating radiomics features based on radiological data and clinical variables originating from textual data for evaluating early ablation efficacy. In the second place, we proposed an information fusion scheme based on preoperative and immediately postoperative CT images, which could integrate the characteristics of the same target area in different periods. Finally, we adopted the GMM method (80) proposed in the previous study to conduct more reasonable data division to improve the model's stability, accuracy and generalization, and minimize the deviation problem resulting from limited sample size when training the model.

There are few studies on the application of artificial intelligence methods to evaluate the efficacy of radiofrequency ablation for CRC lung metastases. A recent study (81) has retrospectively observed the instantaneous changes in intratumor density heterogeneity after MWA of pulmonary tumors *via* radiomics-based CT features and determined the prognostic value in predicting treatment response and local tumor progression (LTP). However, only 50 patients with different diseases (39 primary and 11 metastatic) were enrolled, which could not guarantee a sufficient sample size and the homogeneity of disease. In addition, it was not appropriate to evaluate ablation efficacy by chest contrast-enhanced CT afterablation, which was usually used as the baseline for evaluation (82). Another retrospective study (20) utilized radiomics, clinical, radiological, and technical features to access local control of 48 CRC patients with 119 lung metastases treated by RFA. In order to observe the nodule position in the ablation zone (categorized as nodule seen and remote from borders, or not [i.e., hidden or marginal]), patients underwent chest CT 48 hours after RFA. However, the related results might be subjective among doctors because of different experiences, so they could not assist operators in evaluating the ablation efficacy during the operation, thus allowing for more timely interventions, and in turn, reducing tumor load and prolonging overall survival (83).

Despite the promising results, our study has several limitations. Firstly, the sample size was relatively small because of strict exclusion

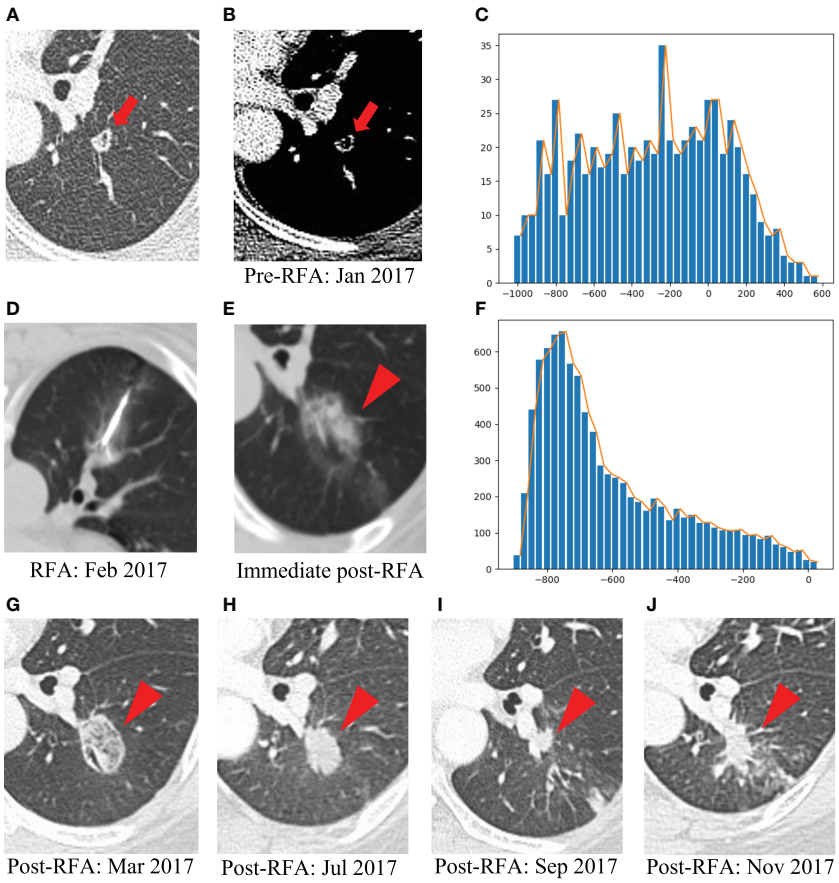


FIGURE 7
Example of a patient with post-lung RFA non-CR. The multimodal data fusion model predicted the results: CR: 0.06, Non-CR: 0.94. **(A)** A 52-year-old female pT4N1M1R0 rectal cancer patient with a lung metastasis eight months after rection, which was located in the left lower lobe (red arrow). **(B)** The mediastinal window of enhanced CT showed that the lesion was accompanied by small cavities with a maximum diameter of 9 mm (red arrow). **(C)** The histogram of the densities within this nodule on the pre-RFA CT scans was flat, without peak among the high tissular attenuations (x-axis: attenuation in Hounsfield units, y-axis: number of voxels). **(D)** The RFA was performed under CT guidance. **(E)** GGO occurred after RFA (red arrowhead). **(F)** The histogram of the densities within the ablation zone on the immediate post-RFA CT scans displayed an asymmetric, skewed distribution (x-axis: attenuation in Hounsfield units, y-axis: number of voxels). **(G)** Chest CT scan showed high density GGO with clear boundary one month after RFA (red arrowhead). **(H)** Five months after RFA, the GGO became a high-density nodule (red arrowhead). **(I)** Seven months after RFA, the high-density nodule shrank, but there was an irregular nodule near the vessel in the ablation area (red arrowhead). **(J)** Nine months after RFA, the irregular nodule was progressively enlarged and the recurrence was considered (red arrowhead). RFA, radiofrequency ablation; CR, complete response; GGO, ground glass opacity.

criteria regarding imaging follow-up. Secondly, the immediate chest CT after RFA was a non-contrast-enhanced CT which might result in the loss of some potentially valuable information related to efficacy. Thirdly, the absent of deep learning algorithm which could identify non-specific features of target lesions and surrounding tissues through automatic learning to achieve information complementation. Thus, a larger patient population from multicenter with deep learning algorithm might further improve the performance in future studies.

In conclusion, the novel multimodal data fusion model (combining radiomics features and clinical variables) was developed to assess the early ablation efficacy. Based on these promising results, our study provides evidence that could assist interventional physicians in objectively and accurately evaluating the immediate efficacy of RFA for CRC lung metastases so as to make necessary supplementary treatment during operation.

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 Fudan University Shanghai Cancer Center. 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

HH, DZ and HC: Methodology, Writing-Reviewing and Editing; CC and YiW: Data Collection and statistical analysis; HH and HC: Evaluation and Validation; DZ: Software and Models Building; LX, YaW and XH: Ablation Operations; WL and YY: Conceptualization and Supervision. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by grants from the National Natural Science Foundation of China (No. 82272095).

Acknowledgments

We thank the Department of Medical Imaging, Mental Health Center, Shanghai Jiao Tong University School of Medicine, and Laboratory for Medical Imaging Informatics, Shanghai Institute of Technical Physics, Chinese Academy of Science for their assistance with this work.

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. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: Cancer J Clin* (2021) 71(3):209–49. doi: 10.3322/caac.21660

2. Nozawa H, Tanaka J, Nishikawa T, Tanaka T, Kiyomatsu T, Kawai K, et al. Predictors and outcome of complete removal of colorectal cancer with synchronous lung metastases. *Mol Clin Oncol* (2015) 3(5):1041–7. doi: 10.3892/mco.2015.599

3. Vatandoust S, Price TJ, Karapetis CS. Colorectal cancer: Metastases to a single organ. *World J Gastroenterol* (2015) 21(41):11767–76. doi: 10.3748/wjg.v21.i41.11767

4. Mitry E, Guiu B, Cosconea S, Jooste V, Faivre J, Bouvier AM. Epidemiology, management and prognosis of colorectal cancer with lung metastases: A 30-year population-based study. *Gut* (2010) 59(10):1383–8. doi: 10.1136/gut.2010.211557

5. Nordholm-Carstensen A, Krarup PM, Jorgensen LN, Wille-Jørgensen PA, Harling H. Occurrence and survival of synchronous pulmonary metastases in colorectal cancer: A nationwide cohort study. *Eur J Cancer* (2014) 50(2):447–56. doi: 10.1016/j.ejca.2013.10.009

6. Ibrahim T, Tselikas L, Yazbeck C, Kattan J. Systemic versus local therapies for colorectal cancer pulmonary metastasis: What to choose and when? *J Gastrointest Cancer* (2016) 47(3):223–31. doi: 10.1007/s12029-016-9818-4

7. McGahan JP, Browning PD, Brock JM, Tesluk H. Hepatic ablation using radiofrequency electrocautery. *Invest Radiol* (1990) 25(3):267–70. doi: 10.1097/00004424-199003000-00011

8. Dupuy DE, Zagoria RJ, Akerley W, Mayo-Smith WW, Kavanagh PV, Safran H. Percutaneous radiofrequency ablation of malignancies in the lung. *AJR Am J roentgenology* (2000) 174(1):57–9. doi: 10.2214/ajr.174.1.1740057

9. de Baère T, Aupérin A, Deschamps F, Chevallier P, Gaubert Y, Boige V, et al. Radiofrequency ablation is a valid treatment option for lung metastases: Experience in 566 patients with 1037 metastases. *Ann oncology: Off J Eur Soc Med Oncology/ESMO* (2015) 26(5):987–91. doi: 10.1093/annonc/mdv037

10. Markezana A, Goldberg SN, Kumar G, Zorde-Khvalevsky E, Gourevtich S, Rozenblum N, et al. Incomplete thermal ablation of tumors promotes increased tumorigenesis. *Int J Hyperthermia* (2021) 38(1):263–72. doi: 10.1080/02656736.2021.1887942

11. Shi L, Wang J, Ding N, Zhang Y, Zhu Y, Dong S, et al. Inflammation induced by incomplete radiofrequency ablation accelerates tumor progression and hinders pd-1 immunotherapy. *Nat Commun* (2019) 10(1):5421. doi: 10.1038/s41467-019-13204-3

12. Yang Q, Qi H, Zhang R, Wan C, Song Z, Zhang L, et al. Risk factors for local progression after percutaneous radiofrequency ablation of lung tumors: Evaluation based on a review of 147 tumors. *J Vasc interventional radiology: JVIR* (2017) 28(4):481–9. doi: 10.1016/j.jvir.2016.11.042

13. Yan P, Tong AN, Nie XL, Ma MG. Assessment of safety margin after microwave ablation of stage I nsclc with three-dimensional reconstruction technique using ct imaging. *BMC Med Imaging* (2021) 21(1):96. doi: 10.1186/s12880-021-00626-z

14. Li C, Wang J, Shao JB, Zhu LM, Sun ZG, Zhang N. Microwave ablation combined with chemotherapy improved progression free survival of iv stage lung adenocarcinoma patients compared with chemotherapy alone. *Thorac Cancer* (2019) 10(7):1628–35. doi: 10.1111/1759-7714.13129

15. Yamamoto A, Nakamura K, Matsuoka T, Toyoshima M, Okuma T, Oyama Y, et al. Radiofrequency ablation in a porcine lung model: Correlation between ct and histopathologic findings. *AJR Am J roentgenology* (2005) 185(5):1299–306. doi: 10.2214/ajr.04.0968

16. Cheng Z, Zhengzhang GU, Shi L, Shan FRadiology DO. The advance of imaging evaluation after ct-guided percutaneous radiofrequency ablation for lung tumors. *International Journal of Medical Radiology* (2016) 39(4):382–5. doi: 10.19300/j.2016.Z4042

17. de Baère T, Palussière J, Aupérin A, Hakime A, Abdel-Rehim M, Kind M, et al. Midterm local efficacy and survival after radiofrequency ablation of lung tumors with minimum follow-up of 1 year: Prospective evaluation. *Radiology* (2006) 240(2):587–96. doi: 10.1148/radiol.2402050807

18. Lee JM, Jin GY, Goldberg SN, Lee YC, Chung GH, Han YM, et al. Percutaneous radiofrequency ablation for inoperable non-small cell lung cancer and metastases: Preliminary report. *Radiology* (2004) 230(1):125–34. doi: 10.1148/radiol.2301020934

19. Matsui Y, Hiraki T, Gobara H, Iguchi T, Fujiwara H, Nagasaka T, et al. Long-term survival following percutaneous radiofrequency ablation of colorectal lung metastases. *J Vasc interventional radiology: JVIR* (2015) 26(3):303–10. doi: 10.1016/j.jvir.2014.11.013

20. Markich R, Palussière J, Catena V, Cazayus M, Fonck M, Bechade D, et al. Radiomics complements clinical, radiological, and technical features to assess local control of colorectal cancer lung metastases treated with radiofrequency ablation. *Eur Radiol* (2021) 31(11):8302–14. doi: 10.1007/s00330-021-07998-4

21. Wei Z, Ye X, Yang X, Zheng A, Huang G, Li W, et al. Microwave ablation in combination with chemotherapy for the treatment of advanced non-small cell lung cancer. *Cardiovasc interventional Radiol* (2015) 38(1):135–42. doi: 10.1007/s00270-014-0895-0

22. Lencioni R, Llovet JM. Modified recist (Mrecist) assessment for hepatocellular carcinoma. *Semin liver Dis* (2010) 30(1):52–60. doi: 10.1055/s-0030-1247132

23. Fernando HC, De Hoyos A, Landreneau RJ, Gilbert S, Gooding WE, Buenaventura PO, et al. Radiofrequency ablation for the treatment of non-small cell lung cancer in marginal surgical candidates. *J Thorac Cardiovasc Surg* (2005) 129(3):639–44. doi: 10.1016/j.jtcvs.2004.10.019

24. Abtin FG, Eradat J, Gutierrez AJ, Lee C, Fishbein MC, Suh RD. Radiofrequency ablation of lung tumors: Imaging features of the postablation zone. *Radiographics: Rev Publ Radiological Soc North America Inc* (2012) 32(4):947–69. doi: 10.1148/rg.324105181

25. Palussière J, Marcet B, Descat E, Deschamps F, Rao P, Ravaud A, et al. Lung tumors treated with percutaneous radiofrequency ablation: Computed tomography imaging follow-up. *Cardiovasc interventional Radiol* (2011) 34(5):989–97. doi: 10.1007/s00270-010-0048-z

26. Suzuki K. Overview of deep learning in medical imaging. *Radiol Phys Technol* (2017) 10(3):257–73. doi: 10.1007/s12194-017-0406-5

27. 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

28. Rios Velazquez E, Parmar C, Liu Y, Coroller TP, Cruz G, Stringfield O, et al. Somatic mutations drive distinct imaging phenotypes in lung cancer. *Cancer Res* (2017) 77(14):3922–30. doi: 10.1158/0008-5472.Can-17-0122

29. Zhang B, He X, Ouyang F, Gu D, Dong Y, Zhang L, et al. Radiomic machine-learning classifiers for prognostic biomarkers of advanced nasopharyngeal carcinoma. *Cancer Lett* (2017) 403:21–7. doi: 10.1016/j.canlet.2017.06.004

30. Sakellaropoulos T, Vougas K, Narang S, Koinis F, Kotsinas A, Polyzos A, et al. A deep learning framework for predicting response to therapy in cancer. *Cell Rep* (2019) 29 (11):3367–73.e4. doi: 10.1016/j.celrep.2019.11.017

31. Xu Y, Hosny A, Zeleznik R, Parmar C, Coroller T, Franco I, et al. Deep learning predicts lung cancer treatment response from serial medical imaging. *Clin Cancer research: an Off J Am Assoc Cancer Res* (2019) 25(11):3266–75. doi: 10.1158/1078-0432.Ccr-18-2495

32. 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

33. Limkin EJ, Sun R, Dercle L, Zacharaki EI, Robert C, Reuzé S, et al. Promises and challenges for the implementation of computational medical imaging (Radiomics) in oncology. *Ann oncology: Off J Eur Soc Med Oncology/ESMO* (2017) 28(6):1191–206. doi: 10.1093/annonc/mdx034

34. Gong J, Liu J, Hao W, Nie S, Wang S, Peng W. Computer-aided diagnosis of ground-glass opacity pulmonary nodules using radiomic features analysis. *Phys Med Biol* (2019) 64(13):135015. doi: 10.1088/1361-6560/ab2757

35. Beig N, Khorrami M, Alilou M, Prasanna P, Braman N, Orooji M, et al. Perinodular and intranodular radiomic features on lung ct images distinguish adenocarcinomas from granulomas. *Radiology* (2019) 290(3):783–92. doi: 10.1148/radiol.2018180910

36. Li Q, Fan L, Cao ET, Li QC, Gu YF, Liu SY. Quantitative ct analysis of pulmonary pure ground-glass nodule predicts histological invasiveness. *Eur J Radiol* (2017) 89:67–71. doi: 10.1016/j.ejrad.2017.01.024

37. Boehm KM, Khosravi P, Vanguri R, Gao J, Shah SP. Harnessing multimodal data integration to advance precision oncology. *Nat Rev Cancer* (2022) 22(2):114–26. doi: 10.1038/s41568-021-00408-3

38. Van Cutsem E, Cervantes A, Adam R, Sobrero A, Van Krieken JH, Aderka D, et al. Esmo consensus guidelines for the management of patients with metastatic colorectal cancer. *Ann oncology: Off J Eur Soc Med Oncology/ESMO* (2016) 27(8):1386–422. doi: 10.1093/annonc/mdw235

39. Goldberg SN, Gazelle GS, Mueller PR. Thermal ablation therapy for focal malignancy: A unified approach to underlying principles, techniques, and diagnostic imaging guidance. *AJR Am J roentgenology* (2000) 174(2):323–31. doi: 10.2214/ajr.174.2.1740323

40. Jaskolka JD, Kachura JR, Hwang DM, Tsao MS, Waddell TK, Asch MR, et al. Pathologic assessment of radiofrequency ablation of pulmonary metastases. *J Vasc interventional radiology: JVIR* (2010) 21(11):1689–96. doi: 10.1016/j.jvir.2010.06.023

41. Yoo DC, Dupuy DE, Hillman SL, Fernando HC, Rilling WS, Shepard JA, et al. Radiofrequency ablation of medically inoperable stage ia non-small cell lung cancer: Are early posttreatment pet findings predictive of treatment outcome? *AJR Am J roentgenology* (2011) 197(2):334–40. doi: 10.2214/ajr.10.6108

42. Lencioni R, Crocetti L, Cioni R, Suh R, Glenn D, Regge D, et al. Response to radiofrequency ablation of pulmonary tumours: A prospective, intention-to-Treat, multicentre clinical trial (the rapture study). *Lancet Oncol* (2008) 9(7):621–8. doi: 10.1016/s1470-2045(08)70155-4

43. Alom MZ, Yakopcic C, Taha TM, Asari VK. Nuclei Segmentation with Recurrent Residual Convolutional Neural Networks Based U-Net (R2u-Net). *NAECON 2018 - IEEE National Aerospace and Electronics Conference* (2018) 228–33. doi: 10.1109/NAECON.2018.8556686

44. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res* (2017) 77(21):e104–e7. doi: 10.1158/0008-5472.Can-17-0339

45. Zwanenburg A, Vallières M, Abdalah MA, Aerts H, Andrearczyk V, Apte A, et al. The image biomarker standardization initiative: Standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology* (2020) 295(2):328–38. doi: 10.1148/radiol.2020191145

46. 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

47. Kumar V, Gu Y, Basu S, Berglund A, Eschrich SA, Schabath MB, et al. Radiomics: The process and the challenges. *Magn Reson Imaging* (2012) 30(9):1234–48. doi: 10.1016/j.mri.2012.06.010

48. Parmar C, Grossmann P, Bussink J, Lambin P, Aerts H. Machine learning methods for quantitative radiomic biomarkers. *Sci Rep* (2015) 5:13087. doi: 10.1038/srep13087

49. Ding C, Peng H. Minimum redundancy feature selection from microarray gene expression data. *J Bioinform Comput Biol* (2005) 3(2):185–205. doi: 10.1142/s0219720005001004

50. Silva D, Deutsch CV. Multivariate data imputation using Gaussian mixture models. *Spatial Stat* (2018) 27:74–90. doi: 10.1016/j.spasta.2016.11.002

51. Lishuai L, Hansman RJ, Palacios R, Welsch R. Anomaly detection *Via* a Gaussian mixture model for flight operation and safety monitoring - sciencedirect. *Transportation Res Part C: Emerging Technol* (2016) 64:45–57. doi: 10.1016/j.trc.2016.01.007

52. Yaxiang F, Wen G, Li D, Qiu S, Levine MD, Xiao F. Video anomaly detection and localization *Via* Gaussian mixture fully convolutional variational autoencoder - sciencedirect. *Comput Vision Image Understanding* (2020) 195(11):102920. doi: 10.1016/j.cviu.2020.102920

53. Zheng DZ YY, Xie Z, Ni YF, Li WT. Data splitting method of distance metric learning based on Gaussian mixed model. *Journal of Shanghai Jiaotong University* (2021) 55(2):131–40. doi: 10.16183/j.cnki.jsjtu.2020.082

54. Khushi M, Shaukat K, Alam TM, Hameed IA, Uddin S, Luo SH, et al. A comparative performance analysis of data resampling methods on imbalance medical data. *IEEE Access* (2021) 9:109960–75. doi: 10.1109/access.2021.3102399

55. Rigatti SJ. Random forest. *J insurance Med* (2017) 47(1):31–9. doi: 10.17849/insm-47-01-31-39.1

56. Yang L, Wu H, Jin X, Zheng P, Hu S, Xu X, et al. Study of cardiovascular disease prediction model based on random forest in Eastern China. *Sci Rep* (2020) 10(1):5245. doi: 10.1038/s41598-020-62133-5

57. Qian L, Zhou Y, Zeng W, Chen X, Ding Z, Shen Y, et al. A random forest algorithm predicting model combining intraoperative frozen section analysis and clinical features guides surgical strategy for peripheral solitary pulmonary nodules. *Trans Lung Cancer Res* (2022) 11(6):1132–44. doi: 10.21037/tlcr-22-395

58. He B, Zhao W, Pi JY, Han D, Jiang YM, Zhang ZG, et al. A biomarker basing on radiomics for the prediction of overall survival in non-small cell lung cancer patients. *Respir Res* (2018) 19(1):199. doi: 10.1186/s12931-018-0887-8

59. Zhang Y, Oikonomou A, Wong A, Haider MA, Khalvati F. Radiomics-based prognosis analysis for non-small cell lung cancer. *Sci Rep* (2017) 7:46349. doi: 10.1038/srep46349

60. Lee SL, Kouzani AZ, Hu EJ. Random forest based lung nodule classification aided by clustering. *Comput Med Imaging Graph* (2010) 34(7):535–42. doi: 10.1016/j.compmedimag.2010.03.006

61. Gong J, Liu JY, Jiang YJ, Sun XW, Zheng B, Nie SD. Fusion of quantitative imaging features and serum biomarkers to improve performance of computer-aided diagnosis scheme for lung cancer: A preliminary study. *Med Phys* (2018) 45(12):5472–81. doi: 10.1002/mp.13237

62. Llinas J, Hall DL. An introduction to multi-sensor data fusion. *Proceedings of the 1998 IEEE International Symposium on Circuits and Systems (Cat. No. 98CH36187)* (1998) 6:537–40. doi: 10.1109/5.554205.

63. Rashinkar P, Krushnasamy VS. An overview of data fusion techniques. International Conference on Innovative Mechanisms for Industry Applications (ICIMIA) (2017) (Bengaluru, India). doi: 10.1109/ICIMIA.2017.7975553.

64. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics* (1988) 44(3):837–45. doi: 10.2307/2531595

65. Staal FCR, Taghavi M, van der Reijd DJ, Gomez FM, Imani F, Klompenhouwer EG, et al. Predicting local tumour progression after ablation for colorectal liver metastases: Ct-based radiomics of the ablation zone. *Eur J Radiol* (2021) 141:109773. doi: 10.1016/j.ejrad.2021.109773

66. Qin S, Hu H, Cui R, Lin J, Liu Y, Wang Y, et al. A prognostic nomogram for intrahepatic progression-free survival in patients with colorectal liver metastases after ultrasound-guided percutaneous microwave ablation. *Int J Hyperthermia* (2022) 39 (1):144–54. doi: 10.1080/02656736.2021.2023226

67. Taghavi M, Staal F, Gomez Munoz F, Imani F, Meek DB, Simões R, et al. Ct-based radiomics analysis before thermal ablation to predict local tumor progression for colorectal liver metastases. *Cardiovasc interventional Radiol* (2021) 44(6):913–20. doi: 10.1007/s00270-020-02735-8

68. Grotowski M. [Antigens (Cea and Ca 19-9) in diagnosis and prognosis colorectal cancer]. *Pol Merkur Lekarski* (2002) 12(67):77–80.

69. Duffy MJ, Lamerz R, Haglund C, Nicolini A, Kalousová M, Holubec L, et al. Tumor markers in colorectal cancer, gastric cancer and gastrointestinal stromal cancers: European group on tumor markers 2014 guidelines update. *Int J Cancer* (2014) 134 (11):2513–22. doi: 10.1002/ijc.28384

70. Steinke K, Haghighi KS, Wulf S, Morris DL. Effect of vessel diameter on the creation of ovine lung radiofrequency lesions in vivo: Preliminary results. *J Surg Res* (2005) 124(1):85–91. doi: 10.1016/j.jss.2004.09.008

71. Najafi A, de Baere T, Purenne E, Bayar A, Al Ahmar M, Delpla A, et al. Risk factors for local tumor progression after rfa of pulmonary metastases: A matched case-control study. *Eur Radiol* (2021) 31(7):5361–9. doi: 10.1007/s00330-020-07675-y

72. Coroller TP, Grossmann P, Hou Y, Rios Velazquez E, Leijenaar RT, Hermann G, et al. Ct-based radiomic signature predicts distant metastasis in lung adenocarcinoma. *Radiotherapy oncology: J Eur Soc Ther Radiol Oncol* (2015) 114(3):345–50. doi: 10.1016/j.radonc.2015.02.015

73. Huang YQ, Liang CH, He L, Tian J, Liang CS, Chen X, et al. Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal cancer. *J Clin oncology: Off J Am Soc Clin Oncol* (2016) 34(18):2157–64. doi: 10.1200/jco.2015.65.9128

74. Liang C, Huang Y, He L, Chen X, Ma Z, Dong D, et al. The development and validation of a ct-based radiomics signature for the preoperative discrimination of stage I-ii and stage iii-iv colorectal cancer. *Oncotarget* (2016) 7(21):31401–12. doi: 10.18632/oncotarget.8919

75. Gillams AR, Lees WR. Radiofrequency ablation of lung metastases: Factors influencing success. *Eur Radiol* (2008) 18(4):672–7. doi: 10.1007/s00330-007-0811-y

76. Galloway MM. Texture analysis using grey level run lengths. *Nasa Sti/recon Technical Report N* (1974) 75:18555.

77. Xu DH, Kurani AS, Furst JD, Raicu DS. Run-length encoding for volumetric texture. *Heart* (2004) 27(25):452–8.

78. Haralick RMJS. Textural features for image classification. *IEEE Transaction Systems Man Cybernetics.* (1973) SMC-3 (6):610–21. doi: 10.1109/TSMC.1973.4309314

79. Zwanenburg A, Leger S, Vallières M, Löck S, Initiative F. Image biomarker standardisation initiative. *Radiotherapy & Oncology* (2016). doi: 10.1016/S0167-8140(18)31291-X

80. Huang H, Zheng D, Chen H, Wang Y, Chen C, Xu L, et al. Fusion of ct images and clinical variables based on deep learning for predicting invasiveness risk of stage I lung adenocarcinoma. *Med Phys* (2022) 49(10):6384–94. doi: 10.1002/mp.15903

81. Liu B, Li C, Sun X, Zhou W, Sun J, Liu H, et al. Assessment and prognostic value of immediate changes in post-ablation intratumor density heterogeneity of pulmonary tumors *Via* radiomics-based computed tomography features. *Front Oncol* (2021) 11:615174. doi: 10.3389/fonc.2021.615174

82. Nian-Long L, Bo Y, Tian-Ming C, Guo-Dong F, Na Y, Yu-Huang W, et al. The application of magnetic resonance imaging-guided microwave ablation for lung cancer. *J Cancer Res Ther* (2020) 16(5):1014–9. doi: 10.4103/jcrt.JCRT_354_20

83. Wang Y, Li G, Li W, He X, Xu L. Radiofrequency ablation of advanced lung tumors: Imaging features, local control, and follow-up protocol. *Int J Clin Exp Med* (2015) 8(10):18137–43.

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 13 February 2023
DOI 10.3389/fonc.2023.1085874

OPEN ACCESS

EDITED BY
Jasper Nijkamp,
Aarhus University, Denmark

REVIEWED BY
Anando Sen,
Newcastle University, United Kingdom
Dario Baldi,
IRCCS SYNLAB SDN, Italy

*CORRESPONDENCE
Evis Sala
es220@medschl.cam.ac.uk

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

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

RECEIVED 31 October 2022
ACCEPTED 24 January 2023
PUBLISHED 13 February 2023

CITATION
Delgado-Ortet M, Reinius MAV, McCague C, Bura V, Woitek R, Rundo L, Gill AB, Gehrung M, Ursprung S, Bolton H, Haldar K, Pathiraja P, Brenton JD, Crispin-Ortuzar M, Jimenez-Linan M, Escudero Sanchez L and Sala E (2023) Lesion-specific 3D-printed moulds for image-guided tissue multi-sampling of ovarian tumours: A prospective pilot study. *Front. Oncol.* 13:1085874. doi: 10.3389/fonc.2023.1085874

Lesion-specific 3D-printed moulds for image-guided tissue multi-sampling of ovarian tumours: A prospective pilot study

Maria Delgado-Ortet[1,2†], Marika A. V. Reinius[2,3,4,5†], Cathal McCague[1,2,4], Vlad Bura[1,2,4,6], Ramona Woitek[1,2,4,7], Leonardo Rundo[1,2,8], Andrew B. Gill[1,4], Marcel Gehrung[2,3], Stephan Ursprung[1,2,4], Helen Bolton[4], Krishnayan Haldar[4], Pubudu Pathiraja[4], James D. Brenton[2,3,4,5], Mireia Crispin-Ortuzar[2,5], Mercedes Jimenez-Linan[2,4], Lorena Escudero Sanchez[1,2] and Evis Sala[1,2,4,9,10*]

[1]Department of Radiology, University of Cambridge, Cambridge, United Kingdom, [2]Cancer Research UK Cambridge Centre, Cambridge, United Kingdom, [3]Cancer Research UK Cambridge Institute, University of Cambridge, Cambridge, United Kingdom, [4]Cambridge University Hospitals NHS Foundation Trust, Cambridge, United Kingdom, [5]Department of Oncology, University of Cambridge, Cambridge, United Kingdom, [6]Department of Radiology, Clinical Emergency Children's Hospital, Cluj-Napoca, Romania, [7]Research Center for Medical Image Analysis & Artificial Intelligence (MIAAI), Danube Private University, Krems, Austria, [8]Department of Information and Electrical Engineering and Applied Mathematics, University of Salerno, Fisciano, SA, Italy, [9]Dipartimento Diagnostica per Immagini, Radioterapia Oncologica ed Ematologia, Policlinico Universitario A. Gemelli IRCCS, Rome, Italy, [10]Dipartimento di Scienze Radiologiche ed Ematologiche, Università Cattolica del Sacro Cuore, Rome, Italy

Background: High-Grade Serous Ovarian Carcinoma (HGSOC) is the most prevalent and lethal subtype of ovarian cancer, but has a paucity of clinically-actionable biomarkers due to high degrees of multi-level heterogeneity. Radiogenomics markers have the potential to improve prediction of patient outcome and treatment response, but require accurate multimodal spatial registration between radiological imaging and histopathological tissue samples. Previously published co-registration work has not taken into account the anatomical, biological and clinical diversity of ovarian tumours.

Methods: In this work, we developed a research pathway and an automated computational pipeline to produce lesion-specific three-dimensional (3D) printed moulds based on preoperative cross-sectional CT or MRI of pelvic lesions. Moulds were designed to allow tumour slicing in the anatomical axial plane to facilitate detailed spatial correlation of imaging and tissue-derived data. Code and design adaptations were made following each pilot case through an iterative refinement process.

Results: Five patients with confirmed or suspected HGSOC who underwent debulking surgery between April and December 2021 were included in this prospective study. Tumour moulds were designed and 3D-printed for seven pelvic lesions, covering a range of tumour volumes (7 to 133 cm^3) and compositions (cystic and solid proportions). The pilot cases informed

innovations to improve specimen and subsequent slice orientation, through the use of 3D-printed tumour replicas and incorporation of a slice orientation slit in the mould design, respectively. The overall research pathway was compatible with implementation within the clinically determined timeframe and treatment pathway for each case, involving multidisciplinary clinical professionals from Radiology, Surgery, Oncology and Histopathology Departments.

Conclusions: We developed and refined a computational pipeline that can model lesion-specific 3D-printed moulds from preoperative imaging for a variety of pelvic tumours. This framework can be used to guide comprehensive multi-sampling of tumour resection specimens.

KEYWORDS

precision oncology, ovarian cancer, cancer imaging, radiogenomics, co-registration, 3D-printing, custom tumour moulds, tumour sampling

1 Introduction

High-Grade Serous Ovarian Carcinoma (HGSOC) is the most prevalent and lethal subtype of ovarian cancer (1). While high levels of genomic complexity and clonal expansion are associated with poor outcome (2), comprehensive tissue multisampling to quantify cellular and molecular tumour heterogeneity is beyond the scope of current clinical diagnostic workflows, thereby limiting our understanding of the landscape of drug resistance mechanisms and potentially actionable targets in HGSOC.

Biomarkers that integrate routinely collected radiological data with molecular features may improve prediction of patient outcome and treatment response (3, 4). However, radiogenomic studies to date predominantly rely on retrospective cohorts and a single tumour sample from a single site per case – thus introducing an unquantified risk of sampling bias, and offering limited insight into the spatial relationship between radiomic features (5) and molecular heterogeneity at the whole-tumour level. This represents one major barrier to clinical implementation of radiogenomic biomarkers.

Biological validation of radiomic habitats –regions that share quantitative imaging characteristics (6, 7) – requires fine co-registration between tissue and imaging coordinates to allow systematic multi-sampling between as well as *within* radiomic habitats (8) emphasising the need for dedicated and improved co-registration methods. Patient-specific three-dimensional (3D) moulds have been proposed to allow co-registration between tumour tissue biopsies and preoperative imaging for subsequent multimodal data correlation and radiogenomic (6, 7) studies in several tumour-types, counting prostate (9–16), hepatic (17, 18) and renal (7, 19) cancers as well as a case report in ovarian cancer (6). These approaches represent important steps toward a detailed 3D spatial understanding of the wide relationship between molecular and radiomic heterogeneity.

In practice, however, two major factors limit wider implementation of 3D mould-based methods across institutions. First, most published works discuss only the end product and not the technical development process. Second, key anatomical, biological and clinical pathway-related aspects of ovarian tumour diagnosis and treatment present unique challenges to the implementation of previously proposed methods. For example, once adnexal tumours are resected, the orientation of those specimens for tissue sampling is often challenging due to the lack of anatomical landmarks. The purpose of this paper is therefore to provide a detailed, comprehensive illustration of the stages of development performed for our use case, to benefit future 3D mould-based work in the gynaecological cancer research community.

To generate a framework for 3D mould-based tumour sampling that can cater for the diversity of ovarian tumours encountered in clinical practice, we conducted a pilot study of five illustrative cases in their primary diagnostic phase. Through an adaptive process described in this work, we have developed and refined an automated computational pipeline that computes the shape and size of tumour delineated on routine computed tomography (CT) or magnetic resonance imaging (MRI) to inform lesion-specific mould printing prior to planned surgical resection. Our method is specifically aimed to allow for detailed spatial correspondence between imaging and tissue-derived data. Multidisciplinary work of this kind is inherently operationally complex – to maximise the reproducibility of our work, we provide a detailed overview of case selection and evaluation, computational modelling, tissue processing and critical considerations around the research pathway development.

2 Methods

2.1 Study design and patient cohort

Five patients with confirmed or suspected HGSOC undergoing debulking surgery between April and December 2021 at Addenbrooke's Hospital (Cambridge University Hospital NHS Foundation Trust, Cambridge, UK) were enrolled in this prospective pilot study. Inclusion required informed written consent to the CTCR-OV04 observational study (Cambridgeshire and Hertfordshire Research Ethics Committee approval reference 08/H0306/61).

Representative cases of the two major treatment pathways in HGSOC (3), i.e. immediate (IPS) and delayed primary surgery (DPS), were included to explore unique challenges associated with each. Inclusion of DPS cases required prior histopathological confirmation of HGSOC. IPS procedures are often performed both as a diagnostic and therapeutic intervention if prior laparoscopic biopsy had not been indicated or possible. The IPS cases without prior histological diagnosis were selected on the basis of clinical suspicion for HGSOC due to significantly elevated serum CA125 (> 4 times upper limit of normal at 35 units per millilitre) and CT imaging features highly suspicious for HGSOC assessed by a Consultant Radiologist with special expertise in gynaecologic oncological imaging (RW/ES).

2.2 Imaging review and segmentation

Preoperative CT and MRI scans were anonymised prior to downloading from the hospital PACS system. This included removal of all directly identifiable information from the images themselves, and deletion of all directly identifiable information from the DICOM headers in the accompanying image metadata. De-identified images were then downloaded in DICOM format from the hospital PACS system and separated into different series using OsiriX DICOM Viewer (version 12.0, Pixmeo SARL, Geneva, CH).

Manual segmentations were performed by a Radiologist in training (CM: 3 years' experience; VB: 6 years' experience) using Microsoft Advanced Medical Image Labeler (version 1.0.0.0, project InnerEye, Microsoft, Redmond, WA, USA) or the Open Health Imaging Foundation viewer (Open Health Imaging Foundation, Massachusetts Institute of Technology, Cambridge, MA, USA, version 3.2.0) *via* its plugin to XNAT, hosted at the local node of the repository established by the CRUK National Cancer Imaging Translational Accelerator (NCITA, https://ncita.org.uk) (20). Challenging segmentations were verified by a board certified Gynaecological Radiologist (RW: 8 years' experience as Consultant Radiologist). Pelvic lesions representing confirmed or suspected tumour were segmented as the region(s) of interest (ROIs) in each case. To automate the tumour rotation steps of the computational pipeline (see Section 2.3), the optimal location of the base of the mould (the surface where the tumour will sit on) for the final two cases was added to the manual segmentations by the radiologist as an extra ROI (Figure 1). Optimal base positioning was selected upon tumour shape and composition on imaging to maximise specimen stability within the mould for increased slicing accuracy, prioritising smooth and solid tumoural surfaces to be on the lowest portion.

2.3 Computational pipeline

A computational pipeline was implemented to generate a mould to specifically hold and slice the segmented lesion in the axial plane, using the input DICOM images and the DICOM-RT file containing the segmented ROIs. A series of tunable values (e.g. slice thickness for tumour dissection, mould height) allows the user to customise the final design to case-specific requirements. The pipeline runs on Python with an interface to OpenSCAD for building the mould structures. All code is available on https://github.com/mariadeor/3d-moulds-for-ovarian-cancer.

2.3.1 Tumour rotation

The first step of the pipeline re-slices and interpolates the input images to a standardised isotropic voxel size of 1×1×1 mm using zero-order spline interpolation to homogenise the process independently of the reconstructed slice thickness of the scans. From there, the tumour segmentation is orientated such that it complies with (a) the imposed base of the mould location and (b) the slicing direction, in

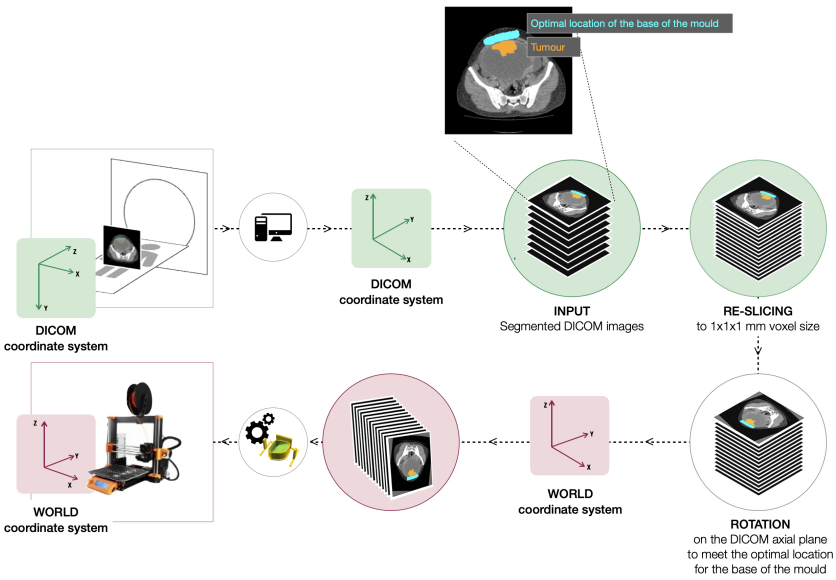


FIGURE 1
Initial steps of the pipeline to re-slice and re-orientate the segmentation to build the mould complying with the imposed base of the mould location (manually segmented, in cyan) and the slicing direction (in the anatomical axial plane). When transforming the image to the world coordinate system (WCS), the slicing direction is perpendicular to the x axis.

the anatomical axial plane (Figure 1). During step (a), the segmentation mask is rotated on the DICOM axial plane to ensure the tumour region adjacent to the segmented optimal location for the base of the mould sits at the bottom. This step could only be automated after the additional ROI outlining the positioning of the base of the mould was added to the input manual segmentation. Step (b) involves the transformation from the DICOM coordinate system to the world coordinate system (WCS), fixed in relation to the print bed.

2.3.2 Tumour modelling

Once the tumour segmentation is appropriately oriented in the WCS, the 3D tumour volume is reconstructed from the 2D segmentations. To minimise the layered appearance of the stack of 2D segmentations and increase the resemblance to the actual specimen, two algorithms are sequentially applied. First, the 3D surface mesh of the tumour volume is extracted using the Lorensen and Cline marching cubes algorithm (21). Next, the mesh is smoothed using Laplacian smoothing ($\lambda = 1$, defined empirically to ensure the smoothing of the tumoural volume while avoiding excessive shrinkage) (22, 23) (Figure 2). The resulting volume (tumour replica) is 3D-printed for improved orientation and visual assessment purposes.

2.3.3 Mould modelling

The mould is designed to aid tumour slicing in the axial plane, which is the standard view for radiological assessment of the abdomen and pelvis. The final mould is an ensemble of three structures: (i) the mould cavity, (ii) the slicing guide and (iii) the orientation guide (Figure 3). Baseline structural design choices (open mould cavity and slicing guide on a single side) were based upon previous optimisation for renal cancers (7) and adapted to the challenges uniquely posed by pelvic lesions throughout this work. A single-sided mould (open cavity) was preferred as changes in the upper side (e.g. cyst rupture, specimen deformation due to the gravity effect on change of orientation) do not affect the accuracy and fitting of the specimen to the mould. Additionally, it requires shorter printing times, especially relevant for the integration of the research pathway in the clinical setting.

First, the cavity of the mould is carved around the convex hull of the tumour replica (Figure 2) and its height is defined through a tunable parameter specifying the percentage of tumour height to cover (Figure 3).

Next, the slicing guide is modelled: it covers the whole length of the tumour on the x axis (the equivalent on the WCS to the patient's craniocaudal axis) and it has the number of slicing slits resultant from dividing the tumour length by the slice width (tunable). The central slicing slit is aligned with the centre of the modelled tumour, and further slicing slits are placed at set intervals on either side of this. The slicing slits have tunable width and are projected to the mould cavity in order to guide the tissue knife all the way through the tumour upon slicing. In parallel to this process, the slicing slits positions are used to extract the 2D outlines of each tissue slice based on the segmentation. These can be printed in real size, have overlaid the base of the mould and the position of the "orientation incision" –explained below– and are fundamental for allowing the co-registration between the tissue and the imaging slices (See Results, Case 5).

The last structures to be built are the two orientation guides, which are single-slit guide walls located perpendicularly to the x axis at each end of the tumour. They have the purpose of determining the course of the "orientation incision" across the top part of the tumour, to ensure the correct orientation of the slices when removed from the mould. This is key for the registration of histopathological samples with imaging (see Results, Cases 3-4).

Finally, the structures are assembled onto a common baseplate. In order not to increase the printing time and minimise the waste of printing material, the baseplate is not a solid block but a tunable offset to the mould cavity and the walls attached.

2.4 3D-printing

All moulds and tumour replicas were 3D-printed using the Prusa i3 MMU2S MK3S (Prusa Research, Prague, Czech Republic) printer loaded with PLA filament. Preceding 3D model slicing and print preparation was done with PrusaSlicer software (PrusaSlicer version 2.3.1, Prusa Research, Prague, Czech Republic), setting the infill value to 20% and the layer height to 0.3 mm.

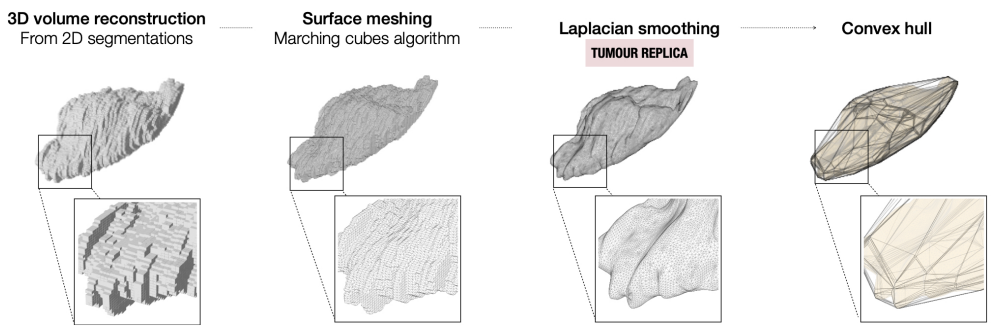


FIGURE 2
Steps for the modelling of the tumour from the 2D segmentations on the preoperative imaging.

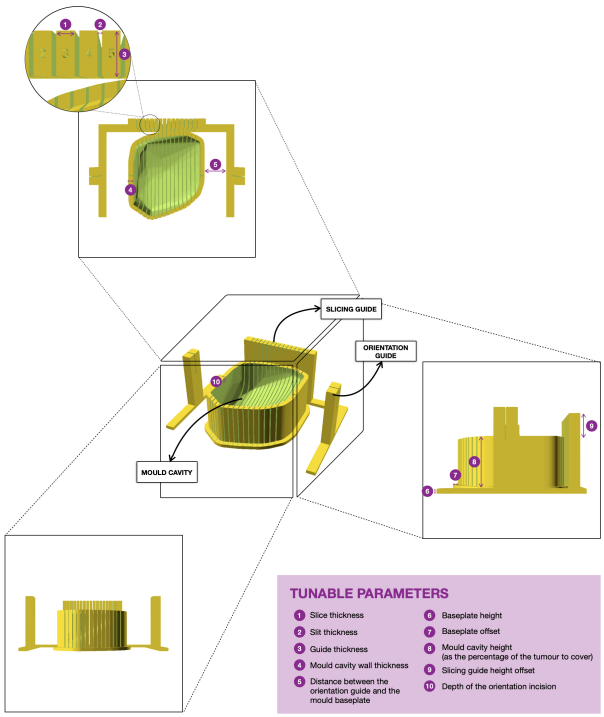


FIGURE 3
3D rendering of the final version of a mould. The tumour sits on the mould cavity and it is sliced by passing a knife through each of the slits of the slicing guide. Two single-slit guide walls are built parallel to the slicing direction for the slitting of the orientation line at the top of the specimen. The orientation guide determines the position of the orientation incision, which facilitates the correct orientation of the slices when removed from the mould. The mould is customisable and the tunable parameters for mould design are in purple.

2.5 Intra- and post-operative workflow

Upon resection by a Specialist Gynae-Oncology Surgeon (HB, KH, PB), tissue specimens were immediately checked visually against their respective tumour replicas and placed within their respective moulds in the operating theatre. As an additional layer of confirmation of correct orientation, orientation sutures were placed by the surgeon prior to resection for the third and subsequent cases. Following confirmation of stability, specimen-bearing moulds were placed in designated clinical specimen containers and transferred to the Cambridge Human Research Tissue Bank (HRTB) following routine tissue transfer procedures.

At the HRTB, all specimens were reviewed and processed by a Consultant Histopathologist specialising in gynaecological cancers (MJ-L) according to standard procedures. Tumours were sliced within their corresponding moulds. Frozen sections were taken for cellularity assessment or to evaluate or confirm likely diagnosis in suspected or confirmed HGSOC, respectively. Benign pathology precluded further tissue processing for research purposes. Malignant tissue sampling for research use was restricted to tissue regions not required for diagnostic purposes in the patient's clinical care.

3 Results

3.1 Clinical characteristics

Five patients were included in this pilot study (Table 1). Mean and median age were 53.8 and 54, respectively, with a range of 47 years between the eldest and youngest patient. Two IPS cases were found to have benign tumours following histopathological review of the IPS resection specimens, while the third IPS case was confirmed to have non-serous high grade pathology. Both DPS cases were treated with four cycles of neoadjuvant carboplatin paclitaxel chemotherapy prior to surgery for HGSOC confirmed on initial diagnostic biopsy. The time interval between imaging and surgery ranged from nine to 53 days (mean: 32.2, median: 40) reflecting individual clinical factors and health system pressures during the COVID-19 pandemic.

3.2 Image segmentation

For two cases undergoing neoadjuvant chemotherapy (NACT) following histopathological confirmation of HGSOC (Cases 1 and 5), routine NACT response assessment images were used for segmenting tumour regions corresponding to expected specimens from DPS. For the remaining three cases undergoing IPS, baseline diagnostic imaging was used.

Tumours were delineated using CT for the majority of cases (Cases 1-4, six moulds). For one patient (Case 5, two moulds) who underwent CT for NACT response assessment, as well as additional MRI prior to surgery, the MRI was used as it was acquired closer to the time of surgery (nine days for MRI, 30 days for CT) (Figure 4). Relevant imaging acquisition parameters are summarised in Supplementary Tables 1, 2.

Segmentations were performed between two and six days before surgery (median five days prior). Segmentation times varied between

TABLE 1 Clinical characteristics of the pilot cohort.

Case	1	2	3	4	5
Prior histology	HGSOC	NA	NA	Insufficient sample but invasive component noted	HGSOC
FIGO stage	IIIC	NA	NA	IIB	IIIB
Surgery type	DPS	IPS	IPS	IPS	DPS
Disease laterality/distribution	Unilateral (left ovary) + peritoneal	Unilateral (right ovary)	Bilateral	Multifocal	Bilateral
Tumour characteristics	Cystic with limited solid component	Solid	Solid	Large cyst with large solid component	Solid
Final histology	HGSOC	Benign	Benign	High grade endometrioid	HGSOC

10.3389/fonc.2023.1085874

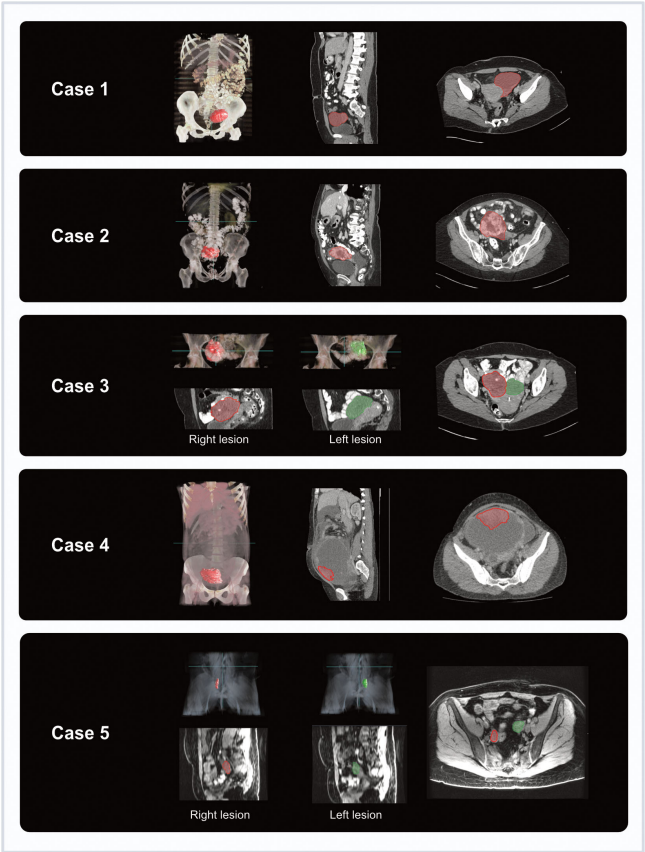


FIGURE 4
Coronal and axial imaging with overlaid tumour segmentations for each case (3D volume rendering from Microsoft Advanced Medical Image Labeler).

cases depending on complexity, taking up to one hour for particularly complex cases.

3.3 Mould generation and refinement

Eight lesion-specific moulds were designed and 3D-printed for a total of seven lesions, covering a range of tumour volumes (7 to 133 cm^3) and compositions (cystic and solid proportions) (Table 2). Tumour replicas were also 3D-printed to visually confirm accurate representation of the shapes and sizes of the resected specimens (accurate in six out of seven lesions).

The computational pipeline for mould design took under two minutes to run in each case. 3D-printing took a median of 4 hours and 41 minutes (range 2-10 h) for the lesion-specific moulds and 2 hours and 32 minutes for the tumour replicas (32 min-7 h).

TABLE 2 Mould and pathway characteristics.

Case	1	2	3	4	5
Mould site(s)	Left ovary	Left ovary	Left ovary Right ovary	Anterior inferior solid mass within cystic lesion (two moulds covering the interior and exterior surface)	Left ovary Right ovary
Tumour volume (cm^3)	125.6	132.9	Left: 69.1 Right: 104.8	Solid mass: 129.8	Left: 10.2 Right: 7.0
Imaging modality	CT	CT	CT	CT	MRI
Time between imaging and surgery	40 days	43 days	15 days	53 days	9 days
Time from study inclusion to surgery	6 days	3 days	3 days	8 days	16 days
Time from image segmentation to surgery	5 days	3 days	3 days	8 days	16 days
Mould 3D-printing time [hours:minutes]	04:30	04:56	Left: 04:13 Right: 04:52	Internal: 10:02 External: 10:30	Left: 02:36 Right: 02:16
Tumour replica 3D-printing time [hours: minutes]	02:36	02:54	Left: 02:12 Right: 02:32	Mass: 06:49	Left: 00:35 Right: 00:32

3.3.1 Case 1

The main aims for the first case were: (a) to build a local research pathway that can be implemented without altering existing standard of care pathways (see Section 3.4); (b) to determine which anatomical structures to include in the mould design; and (c) to test if the meshing and smoothing of the image segmentations produced moulds and tumour replicas that matched the actual size and shape of the resected specimen.

A post-NACT HGSOC case with a sufficiently large lesion to span at least three specimen slicing positions (with a slice width of 10 mm) was selected. Specimen collapse following tissue slicing was expected due to the cystic nature of the lesion of interest, and sufficient tissue for sampling was not anticipated based on imaging features. Instead, overall mould design alone (rather than sampling) was the focus of this first pilot case.

A lesion-specific (left ovary only) mould was 3D-printed and used (an en bloc mould design which was trialled and deemed unsuitable as detailed in Supplementary Material). Given the lack of anatomical landmarks for orientation, the 3D-printed tumour replica was key in facilitating rapid specimen orientation once the tumour was detached from surrounding structures (Figure 5A). Confirming tumour orientation within the mould using the tumour replica also minimised subsequent manipulation of the specimen and the associated risk of cyst rupture. The replica allowed visual confirmation that the modelled volume accurately resembled the actual shape and size of the resected specimen (Figure 5B).

Although the specimen was stable in the mould, slicing resulted in the release of cystic fluid, causing the mass to collapse as expected (Figure 5C). Histopathological review confirmed the lack of sufficient tumour tissue for sampling, as anticipated based on pre-operative CT imaging. However, the test slicing proved useful in highlighting the need for the 0.5 mm knife slots to be widened to accommodate the full depth of the tissue knife.

The smooth symmetric tumour contour, in addition to the lack of anatomical landmarks, made the orientation of the meshed and smoothed tumour replica particularly challenging. While the position of the optimal base of the mould was agreed upon multidisciplinary discussion, it was not included in the segmentation which would have allowed automated processing. Consequently, the computational pipeline required manual input for tumour rotation, a step that could potentially be automated by adding such annotation as shown in later iterations (Cases 4 and 5). Despite the fact that manual rotation on the axial plane was successful, the subsequent transformation to the WCS was unsuccessful, implying that the slicing axis was not aligned to the axial plane. This highlighted the importance of ensuring accurate transformation to the WCS as per enabling specimen slicing along the desired anatomical plane.

3.3.2 Case 2

To test specimen slicing and slice orientation, a suspected malignant right ovarian tumour for IPS resection was selected on the basis of predominantly solid tumour composition. While frozen section assessment and subsequent review of diagnostic formalin-fixed paraffin-embedded (FFPE) blocks confirmed benign tumour pathology and thereby precluded tissue analysis for research purposes, the case provided key learning to improve the mould creation process.

We successfully incorporated our learning from Case 1, including: specimen orientation using the tumour replica, which closely resembled the resection specimen (Figures 6A, B), and tumour slicing through the entire lesion diameter in every slit following increased knife slit width to 1 mm (Figure 6C).

Three key mould design challenges were highlighted by this case. Firstly, insufficient depth and thereby lateral support produced slices that spilled over the mould edge. This led to decreasingly accurate slicing with each cut due to lateral slice movement (Figure 6C). This resulted from setting the maximum mould cavity height to the widest xy dimension of the tumour in the original version of the pipeline –the rationale being that increasing the mould wall height upwards to the level of a narrower point would constrict the mould brim diameter and

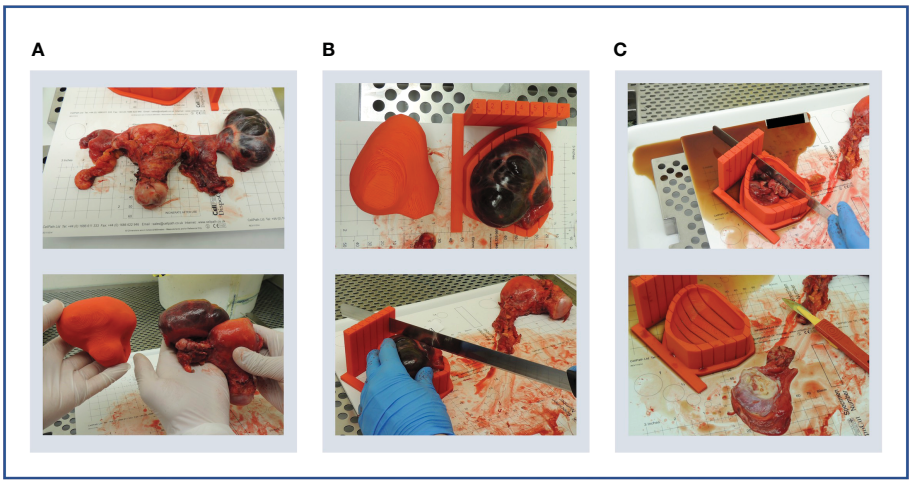


FIGURE 5
Case 1. **(A)** Upper panel: en bloc resection specimen. Lower panel: en bloc resection specimen with left ovarian tumour next to its 3D-printed tumour replica. **(B)** Upper panel: left ovarian tumour in its 3D mould, next to its tumour replica. Lower panel: specimen knife placed in slit prior to slicing. The 5 mm slits were too narrow for the knife. **(C)** Upper panel: the cystic tumour collapsed immediately on capsule rupture as expected, releasing cystic fluid.Lower panel: Tumour tissue was insufficient for research sampling as expected based on the pre-operative CT imaging.

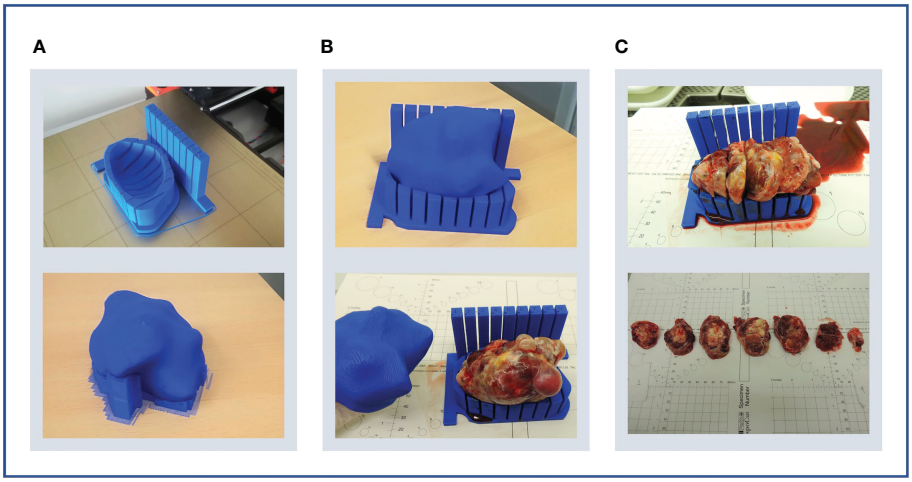


FIGURE 6
Case 2. **(A)** Upper panel: lesion-specific mould on the print bed. Determining tumour specimen placement based on the shape of the mould cavity alone is challenging, hence a tumour replica is also printed (lower panel). A temporary scaffolding to support the emerging replica during printing is added and easily removed when printing is complete. **(B)** Upper panel: the tumour replica fits the mould closely as the mould and replica are both printed based on the same segmentation. Lower panel: the resection specimen is placed directly into its correct position based on the placement of the replica. **(C)** Upper panel: the tumour was successfully sliced, but the shallow tumour cavity led to tissue slices overspilling. Lower panel: even with a tumour with macroscopically visible substructures, slice orientation in relation to the mould and to each other is lost on retrieval from the mould.

impede specimen placement. Secondly, tissue retrieval from the mould after slicing highlighted that slice orientation is easily lost without a method to mark slice orientation in relation to each other and to the mould whilst inside the mould (Figure 6C). Finally, this case also highlighted the need for accurate orientation to WCS as in Case 1.

3.3.3 Case 3

A further case with solid ovarian disease was chosen to address mould design improvements indicated by Case 2.

Firstly, we tested whether the problem with insufficient mould height was an isolated or recurrent problem. When mould cavity height was set to the maximum allowed by the original pipeline algorithm (i.e. the widest xy slice), similarly to the case above, the slice overspill issue recurred. This demonstrated that the mould height would need to be increased perpendicularly upwards (rather than following the narrowing tumour contour) from the widest xy dimension of the tumour (Figure 7) to avoid overspilling (Figure 8B).

Secondly, the loss of slice orientation in relation to each other on removal from the mould was addressed by placing a partial-depth perpendicular incision across the tissue slices while remaining in the mould – we refer to this as the "orientation incision" (Figure 8C). This simultaneously placed a notch in each slice, serving as a physical mark that prevents inadvertent tissue slice rotation on removal from the mould. Additionally, if the position and depth of the orientation incision is known and dictated by the mould, it can also be used as a physical landmark for precise image-tissue co-registration. To achieve this, the orientation guide (Figure 3) was added to the mould structure to aid the slitting of the orientation incision, ensuring that it is both centred to the specimen and at a suitable depth to place a partial notch in the tissue slices without bisecting them (Figure 9).

Thirdly, accurate rotation and transformation to the WCS to orient the mould in relation to the intended slice alignment (axial plane) was once again hampered by the lack of anatomical landmarks or distinctive tumour shape features. An extra ROI drawn during segmentation, marking the optimal base location, was planned for the following case to automate accurate tumour rotation and transformation to WCS.

Additionally, this case presented an opportunity to test pathway compatibility with bilateral disease, with doubled modelling and 3D-printing time. The two moulds and replicas were completed in time for the planned surgery, and tumour replicas closely resembled the corresponding resection specimens as with previous cases (Figure 8A).

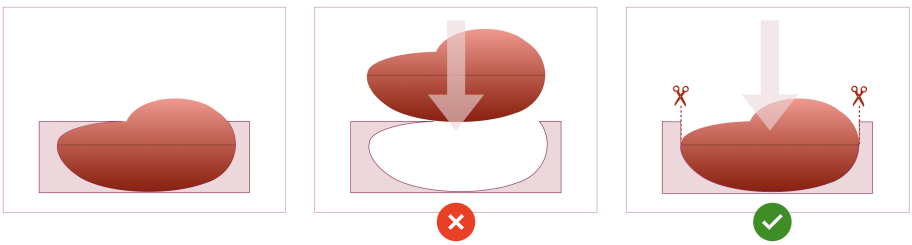


FIGURE 7
Schematic of the design of the mould cavity for cases in which the height of the cavity is greater than the height where the widest surface of the tumour occurs.

10.3389/fonc.2023.1085874

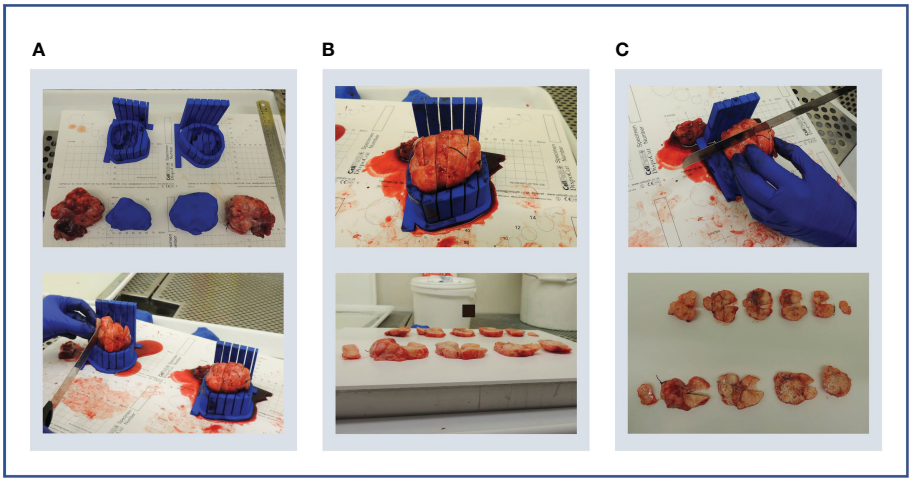


FIGURE 8
Case 3. **(A)** Upper panel: bilateral specimens with respective moulds and replicas. Lower panel: both specimens were sliced within their respective moulds. **(B)** Upper panel: due to shallow mould cavities and slice overspill, accurate cutting into 10 mm slices was challenging, and resulted in unequal slice thickness (lower panel). **(C)** Upper panel: an orientation cut was trialled to mark slice orientation. Lower panel: the orientation cut keeps track of slice orientation to each other.

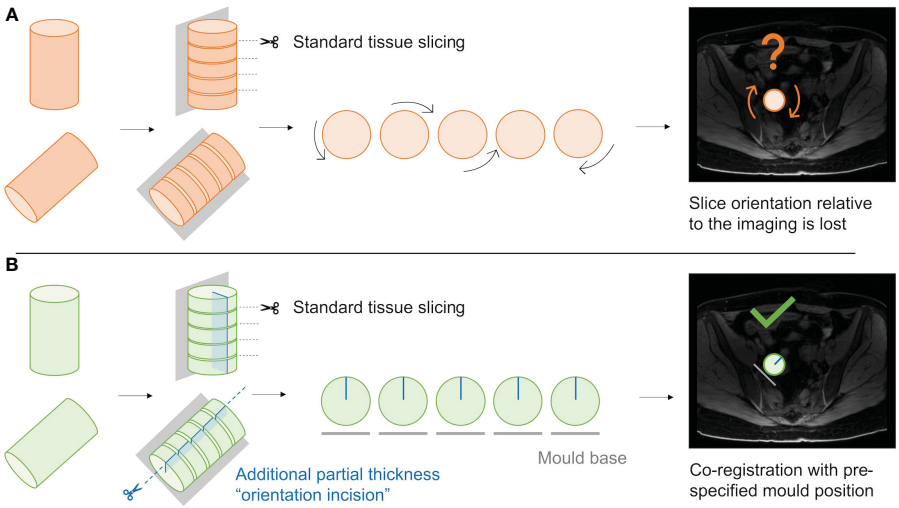


FIGURE 9
Schematic demonstrating the utility of the "orientation incision". **(A)**; Without an orientation incision, any rotation of the tissue slices when removed from the mould cannot be quantified, and co-registration if radiological imaging is lost. **(B)**; a partial-thickness orientation incision across tissue slices allows orientation of slices to each other, and to their corresponding coordinates on the radiological imaging.

3.3.4 Case 4

Cases 2 and 3 highlighted a challenge in terms of being unable to perform tissue sampling when applying this pathway to IPS cases which are only found to be benign at the time of frozen section post-operatively. Given this, a case with confirmed high grade malignant features on biopsy was selected next.

This was an illustrative case of mixed cystic and solid composition often encountered with ovarian tumours (Figure 10A). As the main solid lesion of interest was located adjoining the wall of a very large (233×181×136 mm) cyst which would collapse on slicing, the solid part alone was segmented. Two alternative locations for the mould base were also marked during the segmentation step to enable creation of two alternative moulds to hold this solid component – one each for the external and internal projections of the solid mass in relation to the cyst (Figure 10A). This accounted for preoperative unpredictability regarding the optimal specimen orientation to preserve tissue integrity and stability for this case.

The mould cavity algorithm was modified to allow for the extension of the mould cavity height beyond the point of maximum tumour width to reduce tissue slice overspill as discussed for Cases 2 and 3. Nonetheless, while the resection specimen resembled the tumour replica in shape and contour, the size exceeded the mould dimensions (Figures 10B, C). This reflects the longer time elapsed between the CT imaging used for segmentation and IPS due to individual clinical circumstances (Table 2). Meanwhile, the transformation to WCS was successful for

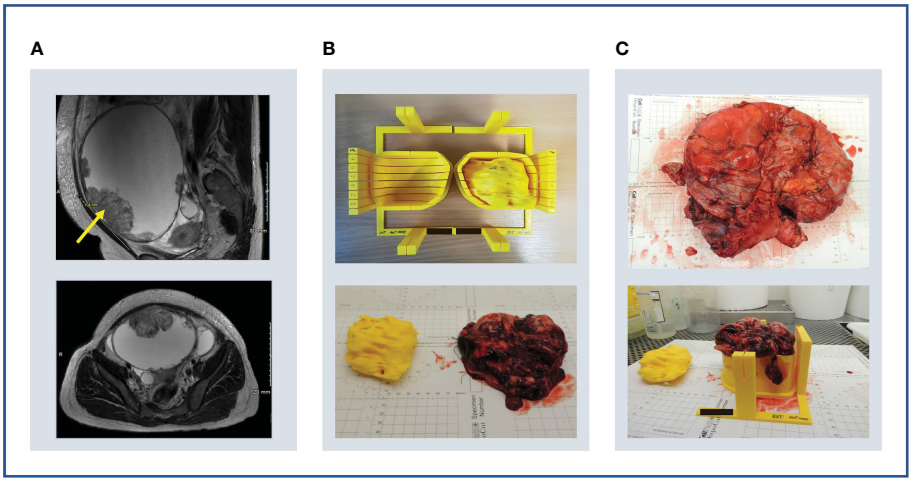


FIGURE 10
Case 4. **(A)** Upper panel: sagittal MRI image showing large cystic lesion with solid component in anterior wall (arrow). Lower panel: axial MRI image showing the same lesion (arrow) and positions of external (blue) and internal (green) mould bases. Note CT imaging was used for segmentation (MRI images used here for illustrative purposes only). **(B)** Upper panel: internal (left) and external (right) moulds with guides (circled) for orientation cut (dotted line). Lower panel: tumour replica and specimen with size discrepancy. **(C)** Upper panel: the ruptured cyst was removed en bloc –two orientation sutures (orange arrow, other suture on underside of specimen) and the resected cervix (blue arrow) were used for specimen orientation. Lower panel: ultimately, specimen size exceeded predicted mould dimensions due to an unusual interval of 53 days between imaging and surgery during which the tumour had increased in size.

this case, demonstrating the utility of adding the ROI of the optimal base of the mould location during the segmentation step.

3.3.5 Case 5

As a final proof of concept of the technical pipeline, we selected a typical HGSOC case with low-volume disease following response to NACT.

The bilateral lesions were segmented on MRI which was performed closer to the time of surgery than CT by 21 days (see Section 3.2). From a technical perspective, the transition from CT to MRI was seamless and did not require any alterations to the pipeline, as the inputs from both modalities are a DICOM set of images and a DICOM-RT file containing the segmented ROIs.

Both lesions were of small size and approximately ovoidal in shape, with their longest axis along the craniocaudal axis (both below 4 cm). This meant that only a limited number of slices with a high thickness-to-diameter ratio would result from using the original slice thickness of 10mm, which would be unideal for multisampling within tissue slices. To enable the option of increasing tissue sampling resolution, the impact of reducing the intervals between slits was explored in preparation for Case 5. A test mould of the previous case was re-modelled and 3D-printed with reduced slice thickness from 10 mm to 5 mm, which confirmed that mould sturdiness was retained.

The two moulds and corresponding tumour replicas for this case were successfully printed in time for the resection procedure (Figure 11A). Structural robustness of the final moulds was maintained with the updated slice thickness of 5 mm, and these slits were successfully used for specimen slicing for both lesions (Figures 11B–E). Whilst tumour cellularity was below the fresh tissue sampling threshold, the reduced slice thickness allowed all slices from each ovary to be placed in tissue cassettes for processing into FFPE blocks. The orientation cut was essential in enabling orientation of tissue sections resulting from these FFPE blocks Figures 11F, L.

3.4 Prospective research pathway development

In addition to providing proof of concept of 3D mould-based multi-sampling in a variety of ovarian lesions, this pilot study sought to test the feasibility of prospective implementation of the method in a busy clinical setting at a tertiary hospital. With patient care remaining the central priority throughout, clinical workflow constraints dictated the following absolute requirements for an ovarian 3D mould research pathway (1): to be compatible with standard clinical procedures without any interference or delay to patients' standard of care treatment pathways (2); to be able to adapt to a variety of real-world timelines, including short treatment planning intervals for cases scheduled for elective surgery on an urgent basis (3); to have no or negligible impact on clinical workload (4); to maximise efficiency and effectiveness of communication between clinical and research team members from diverse disciplines.

To achieve this, we developed a research pathway that fits entirely *around* the standard clinical care pathway for patients with suspected or confirmed ovarian cancer (Figure 12). Following case identification and research consent by a member of our research team, potential cases were selected *via* rapid case review between the Research Radiology team and key members of each research discipline. A designated clinical research fellow (MR: Medical Oncologist in training) coordinated case selection and all communication with clinical teams as a single point of contact, disseminating required case information and confirming availability for participation by each team. Case confirmation triggered the research pathway which interfaces with the clinical care pathway across two phases: from the point of specimen retrieval intraoperatively through to tissue sampling in Histopathology, both coordinated by the same clinical fellow for consistency. The surgical and histopathology teams were briefed at their convenience on the day of surgery with no additional workload or delay.

10.3389/fonc.2023.1085874

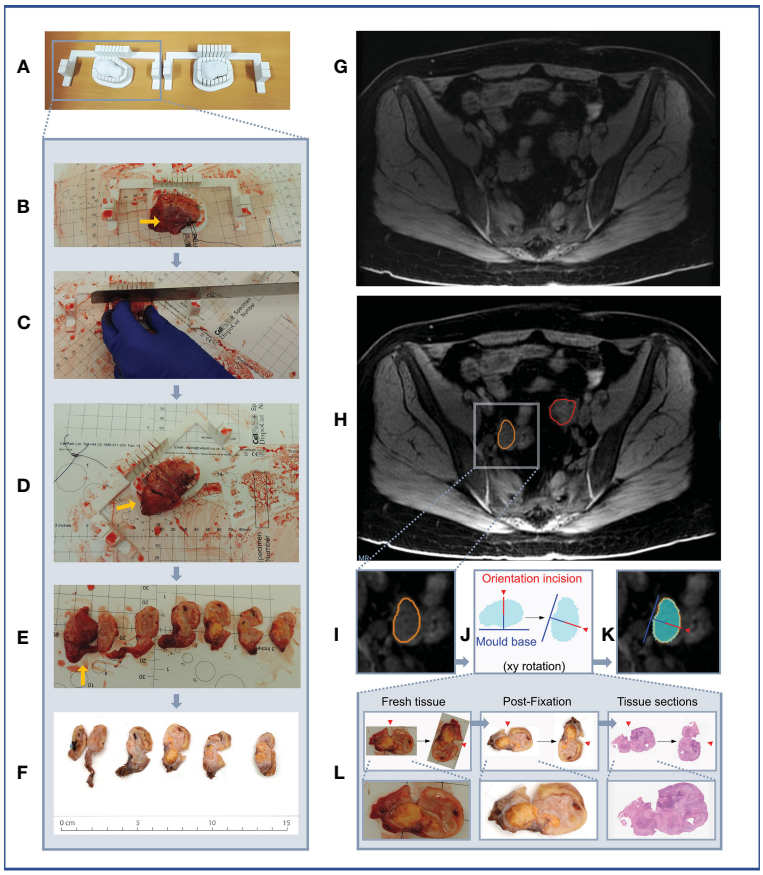


FIGURE 11
Case 5. **(A–F)** Demonstration of 3D mould-based histological processing workflow, with the right ovary as a representative example. Yellow arrow: Fallopian Tube. **(G–L)** Overview of co-registration between radiology and tissues. **(A)** Right and left moulds and corresponding tumour replicas. **(B)** Right ovary placed within corresponding mould. **(C)** Partial orientation incision placed across specimen using in-built orientation guide. **(D)** Specimen sliced within mould. **(E)** Fresh 5mm thickness tissue slices resulting from slicing within mould. Notches from orientation incision allow slices to be orientated in relation to each other. **(F)** Tissue slices following formalin fixation. **(G)** MRI image. **(H)** MRI image with bilateral ovarian segmentations overlayed (orange: right ovary, red: left ovary). **(I)** Magnified image of segmented right ovary. **(J)** Computationally derived expected outline of the corresponding tissue slice and mould base position, rotated on the xy plane to correspond to the base position specified at the segmentation step. The red arrow and segment represent the angle and depth of the in-built orientation incision. **(K)** Right ovarian segmentation with tissue outline and mould base overlayed. **(L)** The orientation incision allowed tissue lice orientation to be preserved across processing steps, from the fresh, fixed to sectioned tissue stages (haematoxylin and eosin stained section shown).

The interval between case confirmation and surgery ranged between 3 and 16 days, and the shortest window for segmentation, modelling and 3D-printing was two and a half days. We demonstrated that all elements of the pathway required for producing moulds and tumour replicas in time for surgery could be completed for every clinical timeline encountered in this pilot study (Figure 13).

4 Discussion

The high levels of heterogeneity at the radiomic, cellular and genomic levels in HGSOC are individually known to carry prognostic significance (24–26), however the spatial relationship between these multiscale features has not been defined. In this work, we developed a computational pipeline for generating lesion-specific 3D-printed moulds to allow for the co-registration of imaging and tissue regions, based directly on insights drawn from a prospective pilot study.

3D moulds have been used in preclinical and translational studies for correlating imaging and tissue-derived data in a number of tumour-types, including ovarian (6), prostate (9–16), hepatic (17, 18) and renal cancers (7, 19). However, no previous study has investigated a variety of ovarian lesion characteristics seen in practice to generate a pipeline that caters to this diversity. Furthermore, most published works to date have presented a final method, without detailing intermediate technical insights relating to both successful and unsuccessful aspects of the mould development process which could be key to reproducing or building upon these existing pipelines. Given this, we provided a detailed account of our entire iterative process of pipeline development, and highlighted critical disease-specific challenges that should be considered in any future 3D mould-based studies of ovarian tumours and HGSOC in particular.

A key anatomical challenge in achieving accurate image-tissue co-registration in ovarian tumours is the lack of universal anatomical structures that can reliably be utilised for computational tumour rotation and specimen orientation. This contrasts with other tumour types for which anatomical contexts are inherently preserved post-resection, as exemplified by the use of the tumour-hilum contact point of radical nephrectomy specimens in determining mould base position and confirming correct specimen positioning (7). We

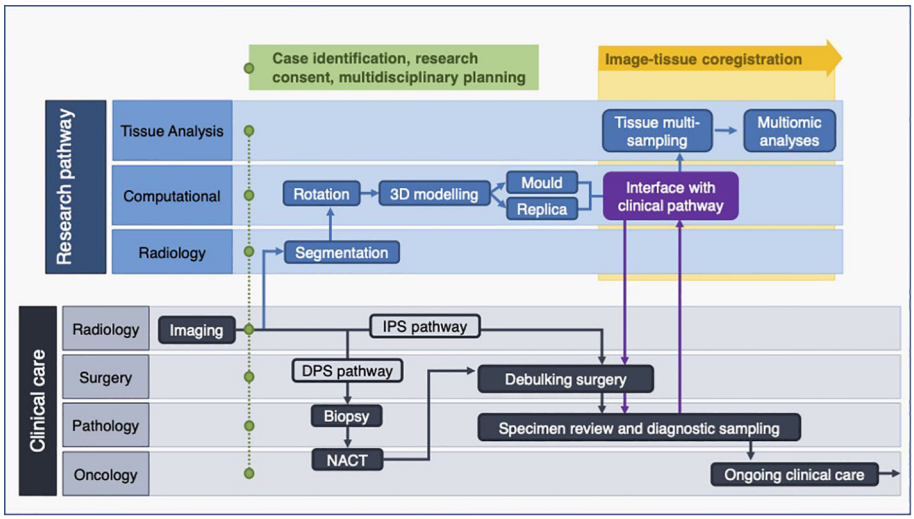


FIGURE 12
Multidisciplinary research pathway for malignant pelvic ovarian tumours. Interface with clinical pathway shown in purple. DPS, (delayed primary surgery); IPS, (immediate primary surgery); NACT, (neoadjuvant chemotherapy).

demonstrated that the utility of en bloc resection for orientation is limited in ovarian tumours (Case 1, Supplementary Material), particularly given the highly tortuous and mobile nature of the Fallopian tube, as well as the separation of ovarian tumours from adjoining structures prior to slicing during standard histopathological tissue processing. Instead, we addressed this through two innovations. Firstly, we achieved automated tumour rotation by introducing an additional annotation to mark optimal base of the mould location at the segmentation stage. This had the advantage of allowing the slicing to be constrained to an anatomical plane of choice. Secondly, by creating 3D-printed tumour replicas, we were able to conduct visual comparison of predicted and actual resection specimen characteristics and guide correct specimen placement in the mould.

From a clinical perspective, the main challenge in implementing 3D mould-based sampling of resection specimens in HGSOC is that an increasing proportion of patients are treated with NACT followed by DPS (27). While the majority of HGSOC cases are diagnosed at an advanced stage and often with large-volume disease (28), tumours resected at DPS in the context of NACT response can be of markedly small volume. The volume of the lesions was of particular relevance during computational tumour modelling, as meshing and smoothing processes have greater effects on small volumes. The use of tumour replicas was thereby key in confirming that the modelled tumour volumes accurately resembled the shape and size of the resected specimen for six out of the seven lesions, including those below 15 cm^3. Unfortunately, given the scope and size (five cases, seven lesions) of the pilot study, we could not establish a meaningful volumetric tolerance to quantitatively study the suitability of a lesion for our mould design. However, other lesion characteristics beyond volume are crucial for the mould design. For instance, the length of the tumour in the axial plane (slicing direction) highlights the importance of the presence of tunable parameters including slice thickness, in order to allow the granularity of tissue sampling to be adjusted on a case-by-case basis and ensure that a sufficient number of slices can be acquired for meaningful tissue multisampling. We have shown that the structural integrity of moulds was retained when halving slice thickness from 10 mm to 5 mm while increasing knife slits from 0.5 mm to 1 mm. This was of particular importance in Case 5, in which the post-NACT tumour was found to have negligible tumour cellularity below the threshold required for fresh tissue-based

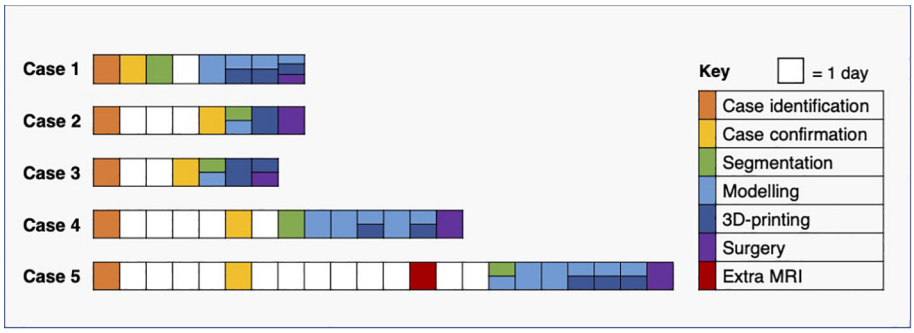


FIGURE 13
Pathway timelines for each case. All baseline CT imaging took place prior to case identification (see Figure 12).

genomic analysis. By reducing the slice thickness, all slices could be placed in standard tissue cassettes to be made into FFPE blocks, which can instead be sectioned and subjected to other analyses such as computational pathology approaches.

Accurate image-tissue co-registration relies on two further technical properties of mould design, namely stability and orientation of tissue slices while within the mould. Firstly, as shown by Cases 2 and 3, the accuracy of specimen slicing is dependent on the stability and support afforded by its mould, which requires both the robustness of the mould itself, as well as a sufficient mould height to provide lateral support. To avoid specimen overspill, we therefore do not constrain mould height to the point at which the widest xy dimension of the tumour occurs (7), but rather allow the percentage height covered by the mould cavity to be tunable according to case-specific requirements. Specimen stability is also dependent on the structural composition across the specimen itself, which may not be possible to determine preoperatively, particularly based on CT imaging alone. The ability to determine mould base location during segmentation, and to specify multiple options in order to 3D-print alternative moulds in advance of surgery, is therefore a key strength that contributes to the flexibility of our pipeline, as confirmed by Case 4.

Secondly, without a method to mark tumour slices within the mould, their orientation to each other and to the radiological imaging is immediately lost upon removal from the mould, particularly for ovoid tissues lacking obvious shape or substructure characteristics (Figure 9). Our in-built partial-thickness perpendicular orientation slit is therefore a powerful improvement on existing co-registration pipelines which allows slice orientation to be maintained both before and after fixation and embedding, but crucially is simple and rapid enough to be implemented in a clinical setting (Figure 11). To perform co-registration, we generated slice-by-slice outlines of tissue slices expected from each slicing position in the mould (Figure 11J). These outlines include the positions of the mould base and orientation incision (the depth of this partial incision is marked in red in Figure 11J), such that the physical orientation incisions of tissue slices can be used to orient them against the slice-by-slice outlines. A critical mould design feature here is that the tissue slicing corresponds to axial slices in the segmented radiological imaging, meaning that each slice outline and corresponding tissue slice simply requires a single rotation by a fixed angle along the xy plane to be aligned with the mould base position as specified at the image segmentation step (Figures 11G–K). Given the persisting physical presence of the orientation incision, this orientation process can be undertaken at the fresh or post-fixation stages of tissue processing. In our final case, we were able to extend this co-registration approach to the subsequent whole-slide histological sections generated from the FFPE blocks, given the context of low-volume residual disease post-NACT where the 5mm-thickness tissue sections could be embedded in full within standard tissue cassettes (Figure 11L). The ability to perform coregistered sampling before or after sample fixation allows an important level of flexibility in being able to adapt tissue sampling to post-operative findings – for instance, when frozen section review of the fresh specimen in Case 5 demonstrated insufficient histopathological tumour cellularity for genomic analysis by taking fresh tissue biopsies, it was possible to make a rapid real-time decision to prioritise whole-slice analyses of FFPE blocks at the next stage of routine clinical tissue processing.

In practice, a research pathway of this kind must be highly adaptable to the requirements of a patient's individual treatment pathway and tumour characteristics, and be implemented efficiently with minimal to no impact on resources and professional workloads within the wider healthcare setting. While previous work has used multiparametric MRI and PET/CT (6, 29), we were able to successfully implement our framework based on more routinely available CT (N = 4), as well as MRI (N = 1). In addition to addressing the aforementioned technical and anatomical challenges by implementing a range of tunable parameters, our research pipeline was able to accommodate further clinical scenarios including mould generation for bilateral disease (Cases 3 and 5) and for the solid component of a large mixed solid-cystic mass (Case 4). All research components were completed in time for each patient's surgical procedure without interference with individual treatment timelines, with a minimum interval of three days between case selection and debulking surgery. Focused and coordinated communication between multidisciplinary clinical professionals from the Radiology, Surgery, Oncology and Histopathology departments was key in enabling timely research pathway implementation.

Our approach is based on clinical standard-of-care imaging and did not include any dedicated imaging sessions for optimising mould design. This was the likely reason for the mismatch between the segmented tumour and actual specimen volume in Case 4, who had the longest elapsed time between imaging and surgery. Of note, in most routine clinical settings, cross-sectional imaging is undertaken in prior preparation for but not on the day of surgery, and the cost of further imaging or risk of radiation exposure is not justified without unique clinical indications. An additional scan, ideally an MRI including T2-weighted images, close to the day of surgery could easily solve this problem in cases with a long delay between the last preoperative standard-of-care scan and surgery. However, we found that all other segmentations generated tumour replicas which closely resembled the resected specimen, including that of a post-NACT case whose imaging was performed 40 days prior to surgery.

A second limitation is that these pilot cases did not yield HGSOC tissue for multi-site genomic profiling. This highlights the inherent challenges of case selection in the context of standard treatment pathways for advanced pelvic ovarian tumours. For cases undergoing IPS, tumours are likely to be of greater volume and cellularity for sampling, but histological diagnosis is often lacking preoperatively. An IPS case may be benign despite significantly elevated CA-125 and suspicious qualitative radiological features (as we found with Cases 2 and 3), however a significant proportion of tumour will be required for diagnostic purposes and may leave a small proportion for research purposes even if malignancy is confirmed. DPS cases are more likely to have limited remaining tumour tissue for sampling (Case 5), or have greater cystic components with or without sufficient tissue for sampling as a feature of NACT response (Case 1). With regard to tumours containing cystic components, we have shown that moulds can be modelled for a discrete solid component, and this can be extended to multiple distinct areas of solid disease. The majority of cases for which tumours are predominantly cystic with minimal or no residual tumour tissue after NACT will be identifiable at the preoperative imaging stage. The manual base placement step allows the base to be placed by the surface with the largest proportion of solid components to maximise stability – however specimen stability would remain challenging in very complex masses containing a large number of discrete cystic and solid areas. Future case selection will be informed by these clinical factors, and may benefit from focusing on IPS cases with prior tissue diagnosis from

laparoscopic biopsy or DPS cases with predominantly solid residual disease post-NACT.

This work was motivated by the need to guide comprehensive multi-sampling of pelvic tumour resected specimens for correlation of imaging and genomic features. A key strength of our work is that the pipeline facilitates direct co-registration of tissue coordinates and imaging slices on the same transverse axis by constraining tissue slicing to the axial plane. Our pipeline is thereby highly suited to implementing systematic grid-based sampling, by aligning a grid-based system with the orientation cut of each tissue slice in order to directly map sample coordinates onto spatial coordinates of the corresponding image slice. Such a coordinate-based system will be well-suited to larger-scale prospective biological validation of previously described radiomic 'habitats' (6).

5 Conclusion

We developed a computational pipeline for modelling lesion-specific 3D-printed moulds to guide slicing and multi-sampling of solid pelvic tumour resection specimens. This work provides a framework for obtaining spatially co-registered imaging and multi-sampled tissue data, thus aiming to perform detailed multi-level characterisation of intratumoural heterogeneity of pelvic gynaecological tumours. We highlight specific challenges pertaining to ovarian tumours, propose a highly tunable design that is adaptable to the specific requirements of a given case, and provide recommendations for pathway implementation. This pipeline can be implemented alongside clinical treatment pathways for patients with newly diagnosed HGSOC and other ovarian tumours.

Data availability statement

The datasets presented in this article are not readily available because of ethical restrictions due to the presence of patient data. Requests to access the datasets should be directed to Prof Evis Sala, es220@medschl.cam.ac.uk.

Ethics statement

The studies involving human participants were reviewed and approved by Cambridgeshire and Hertfordshire Research Ethics Committee approval reference 08/H0306/61. The patients/participants provided their written informed consent to participate in this study.

Author contributions

Conceptualisation: MR, MD-O, CM, VB, RW, LR, MJ-L, MC-O, JB, LES and ES. Methodology: MR, MD-O, CM, VB, RW, LR, MJ-L, MC-O, JB, LES and ES. Software: MD-O (original for this work), AG, MG and MC-O (previous developments). Validation: MD-O, MR, CM, VB, RW, LR, MJ-L, MC-O, LES and ES. Investigation: MR, MD-O, CM, VB, LES, RW, MJ-L, JB and ES. Resources: HB, KH, PP, MJ-L, JB and ES. Data curation: MD-O, MR and LES. Visualisation: MR, MD-O and CM. Writing—original draft: MD-O and MR. Writing—review and editing: MR, MD-O, LES, CM, VB, RW, LR, AG, MJ-L, MG, SU, HB, KH, PP, MC-O, JDB and ES. Supervision: ES, LES, MC-O, JB and RW. Project administration: MR, JB and ES. Funding acquisition: JB and ES. All authors contributed to the article and approved the submitted version.

Funding

MD-O was supported by the W.D. Armstrong Trust Fund. LS and ES were supported by the CRUK National Cancer Imaging Translational Accelerator (NCITA) [C42780/A27066]. MC-O, RW, LR, JB and ES were supported by The Mark Foundation for Cancer Research [C9685]. MC-O, MR, RW, LR, JB and ES were supported by Cancer Research UK and the Cancer Research UK Cambridge Centre [A22905, A25177]. RW was supported by the Austrian Science Fund [J-4025]. Additional support has been provided by the Wellcome Trust Innovator Award, UK [215733/Z/19/Z] and the National Institute of Health Research (NIHR) Cambridge Biomedical Research Centre [BRC-1215-20014]. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

Acknowledgments

We would like to thank the patients who consented to the CTCR-OV04 study. The authors would also like to thank the Cambridge Human Research Tissue Bank for their support. Microsoft Advanced Medical Image Labeler was provided to Addenbrooke's Hospital (Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK) by the Microsoft InnerEye project.

Conflict of interest

MC-O: 52 North Health Ltd. employee and shareholder, GE HealthCare research support, GSK speakers' bureau. JB: Tailor Bio co-founder, stock options, AstraZeneca consulting and advisory roles, honoraria, Clovis Oncology consulting and advisory roles, GSK honoraria, holder of patents TAm-Seq v2 method for ctDNA estimation; enhanced detection of target DNA by fragment size analysis; methods for predicting treatment response in cancers. ES: Lucida Medical co-founder and shareholder, GE HealthCare research support, speakers' bureau, Canon research support, speakers' bureau.

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2023.1085874/full#supplementary-material

References

1. Jayson GC, Kohn EC, Kitchener HC, Ledermann JA. Ovarian cancer. *Lancet Lond Engl* (2014) 384(9951):1376–88. doi: 10.1016/S0140-6736(13)62146-7

2. Schwarz RF, Ng CKY, Cooke SL, Newman S, Temple J, Piskorz AM, et al. Spatial and temporal heterogeneity in high-grade serous ovarian cancer: a phylogenetic analysis. *PloS Med* (2015) 12(2):e1001789. doi: 10.1371/journal.pmed.1001789

3. Funingana IG, Reinius MAV, Petrillo A, Ang JE, Brenton JD. Can integrative biomarker approaches improve prediction of platinum and PARP inhibitor response in ovarian cancer? *Semin Cancer Biol* (2021) 77:67–82. doi: 10.1016/j.semcancer.2021.02.008

4. McCague C, Ramlee S, Reinius M, Selby I, Hulse D, Piyatissa P, et al. Introduction to radiomics for a clinical audience. *Clin Radiol* (2023) 78(2):83–98. doi: 10.1016/j.crad.2022.08.149

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. Weigelt B, Vargas HA, Selenica P, Geyer FC, Mazaheri Y, Blecua P, et al. Radiogenomics analysis of intratumor heterogeneity in a patient with high-grade serous ovarian cancer. *JCO Precis Oncol* (2019) 3:PO.18.00410. doi: 10.1200/PO.18.00410

7. Crispin-Ortuzar M, Gehrung M, Ursprung S, Gill AB, Warren AY, Beer L, et al. Three-dimensional printed molds for image-guided surgical biopsies: An open source computational platform. *JCO Clin Cancer Inform* (2020) 4:736–48. doi: 10.1200/CCI.20.00026

8. Tomaszewski MR, Gillies RJ. The biological meaning of radiomic features. *Radiology* (2021) 298(3):505–16. doi: 10.1148/radiol.2021202553

9. Shah V, Pohida T, Turkbey B, Mani H, Merino M, Pinto PA, et al. A method for correlating *in vivo* prostate magnetic resonance imaging and histopathology using individualized magnetic resonance-based molds. *Rev Sci Instrum* (2009) 80(10):104301. doi: 10.1063/1.3242697

10. Turkbey B, Mani H, Shah V, Rastinehad AR, Bernardo M, Pohida T, et al. Multiparametric 3T prostate magnetic resonance imaging to detect cancer: histopathological correlation using prostatectomy specimens processed in customized magnetic resonance imaging based molds. *J Urol* (2011) 186(5):1818–24. doi: 10.1016/j.juro.2011.07.013

11. Trivedi H, Turkbey B, Rastinehad AR, Benjamin CJ, Bernardo M, Pohida T, et al. Use of patient-specific MRI-based prostate mold for validation of multiparametric MRI in localization of prostate cancer. *Urology* (2012) 79(1):233–9. doi: 10.1016/j.urology.2011.10.002

12. Costa DN, Chatzinoff Y, Passoni NM, Kapur P, Roehrborn CG, Xi Y, et al. Improved magnetic resonance imaging-pathology correlation with imaging-derived, 3D-printed, patient-specific whole-mount molds of the prostate. *Invest Radiol* (2017) 52(9):507–13. doi: 10.1097/RLI.0000000000000372

13. Hurrell SL, McGarry SD, Kaczmarowski A, Iczkowski KA, Jacobsohn K, Hohenwalter MD, et al. Optimized b-value selection for the discrimination of prostate cancer grades, including the cribriform pattern, using diffusion weighted imaging. *J Med Imaging Bellingham Wash* (2018) 5(1):011004. doi: 10.1117/1.JMI.5.1.011004

14. McGarry SD, Hurrell SL, Iczkowski KA, Hall W, Kaczmarowski AL, Banerjee A, et al. Radio-pathomic maps of epithelium and lumen density predict the location of high-grade prostate cancer. *Int J Radiat Oncol Biol Phys* (2018) 101(5):1179–87. doi: 10.1016/j.ijrobp.2018.04.044

15. Baldi D, Aiello M, Duggento A, Salvatore M, Cavaliere C. MR imaging-histology correlation by tailored 3D-printed slicer in oncological assessment. *Contrast Media Mol Imaging* (2019) 2019:1071453. doi: 10.1155/2019/1071453

16. Priester A, Wu H, Khoshnoodi P, Schneider D, Zhang Z, Asvadi NH, et al. Registration accuracy of patient-specific, three-Dimensional-Printed prostate molds for correlating pathology with magnetic resonance imaging. *IEEE Trans BioMed Eng* (2019) 66(1):14–22. doi: 10.1109/TBME.2018.2828304

17. Trout AT, Batie MR, Gupta A, Sheridan RM, Tiao GM, Towbin AJ. 3D printed pathological sectioning boxes to facilitate radiological-pathological correlation in hepatectomy cases. *J Clin Pathol* (2017) 70(11):984–7. doi: 10.1136/jclinpath-2016-204293

18. Mikhail AS, Mauda-Havakuk M, Partanen A, Karanian JW, Pritchard WF, Wood BJ. Liver-specific 3D sectioning molds for correlating *in vivo* CT and MRI with tumor histopathology in woodchucks (Marmota monax). *PloS One* (2020) 15(3):e0230794. doi: 10.1371/journal.pone.0230794

19. Dwivedi DK, Chatzinoff Y, Zhang Y, Yuan Q, Fulkerson M, Chopra R, et al. Development of a patient-specific tumor mold using magnetic resonance imaging and 3-dimensional printing technology for targeted tissue procurement and radiomics analysis of renal masses. *Urology* (2018) 112:209–14. doi: 10.1016/j.urology.2017.08.056

20. Doran S, Al Sa'd M, Petts J, Darcy J, Alpert K, Cho W, et al. Integrating the OHIF viewer into XNAT: Achievements, challenges and prospects for quantitative imaging studies. *Tomography* (2022) 8(1):497–512. doi: 10.3390/tomography8010040

21. Lorensen WE, Cline HE. Marching cubes: A high resolution 3D surface construction algorithm. *ACM SIGGRAPH Comput Graph* (1987) 21(4):163–9. doi: 10.1145/37402.37422

22. Desbrun M, Meyer M, Schröder P, Barr AH. Implicit fairing of irregular meshes using diffusion and curvature flow. In: *Proceedings of the 26th annual conference on computer graphics and interactive techniques - SIGGRAPH '99*. ACM Press (1999). p. 317–24. Available at: http://portal.acm.org/citation.cfm?doid=311535.311576.

23. Vollmer J, Mencl R, Muller H. Improved laplacian smoothing of noisy surface meshes. *Comput Graph Forum* (1999) 18(3):131–8. doi: 10.1111/1467-8659.00334

24. Schwarz RF, Trinh A, Sipos B, Brenton JD, Goldman N, Markowetz F. Phylogenetic quantification of intra-tumour heterogeneity. *PloS Comput Biol* (2014) 10(4):e1003535. doi: 10.1371/journal.pcbi.1003535

25. Vargas HA, Veeraraghavan H, Micco M, Nougaret S, Lakhman Y, Meier AA, et al. A novel representation of inter-site tumour heterogeneity from pre-treatment computed tomography textures classifies ovarian cancers by clinical outcome. *Eur Radiol* (2017) 27(9):3991–4001. doi: 10.1007/s00330-017-4779-y

26. Meier A, Veeraraghavan H, Nougaret S, Lakhman Y, Sosa R, Soslow RA, et al. Association between CT-texture-derived tumor heterogeneity, outcomes, and BRCA mutation status in patients with high-grade serous ovarian cancer. *Abdom Radiol N Y* (2019) 44(6):2040–7. doi: 10.1007/s00261-018-1840-5

27. Knisely AT, St. Clair CM, Hou JY, Collado FK, Hershman DL, Wright JD, et al. Trends in primary treatment and median survival among women with advanced-stage epithelial ovarian cancer in the US from 2004 to 2016. *JAMA Netw Open* (2020) 3(9):e2017517. doi: 10.1001/jamanetworkopen.2020.17517

28. Prat J. FIGO committee on gynecologic oncology. staging classification for cancer of the ovary, fallopian tube, and peritoneum. *Int J Gynecol Obstet* (2014) 124(1):1–5. doi: 10.1016/j.ijgo.2013.10.001

29. Jiménez-Sánchez A, Cybulska P, Mager KL, Koplev S, Cast O, Couturier DL, et al. Unraveling tumor–immune heterogeneity in advanced ovarian cancer uncovers immunogenic effect of chemotherapy. *Nat Genet* (2020) 52(6):582–93. doi: 10.1038/s41588-020-0630-5

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 20 February 2023
DOI 10.3389/fonc.2023.1079040

OPEN ACCESS

EDITED BY
Jasper Nijkamp,
Aarhus University, Denmark

REVIEWED BY
Petra J. Van Houdt,
The Netherlands Cancer Institute (NKI), Netherlands
Catarina Fernandes,
Eindhoven University of Technology, Netherlands

*CORRESPONDENCE
Johan Bengtsson
johan.bengtsson@med.lu.se

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

RECEIVED 24 October 2022
ACCEPTED 03 February 2023
PUBLISHED 20 February 2023

CITATION
Bengtsson J, Thimansson E, Baubeta E, Zackrisson S, Sundgren PC, Bjartell A and Flondell-Sité D (2023) Correlation between ADC, ADC ratio, and Gleason Grade group in prostate cancer patients undergoing radical prostatectomy: Retrospective multicenter study with different MRI scanners.
Front. Oncol. 13:1079040.
doi: 10.3389/fonc.2023.1079040

Correlation between ADC, ADC ratio, and Gleason Grade group in prostate cancer patients undergoing radical prostatectomy: Retrospective multicenter study with different MRI scanners

Johan Bengtsson[1,2]*, Erik Thimansson[3,4], Erik Baubeta[2,3], Sophia Zackrisson[3,5], Pia Charlotte Sundgren[1,2,6], Anders Bjartell[3,7] and Despina Flondell-Sité[3,7]

[1]Department of Clinical Sciences, Radiology, Lund, Lund University, Lund, Sweden, [2]Department of Medical Imaging and Physiology, Skåne University Hospital, Lund, Sweden, [3]Department of Translational Medicine, Lund University, Malmö, Sweden, [4]Department of Radiology, Helsingborg Hospital, Helsingborg, Sweden, [5]Department of Medical Imaging and Physiology, Skåne University Hospital, Malmö, Sweden, [6]Lund Bioimaging Center (LBIC), Lund University, Lund, Sweden, [7]Department of Urology, Skåne University Hospital, Malmö, Sweden

Background: MRI is an important tool in the prostate cancer work-up, with special emphasis on the ADC sequence. This study aimed to investigate the correlation between ADC and ADC ratio compared to tumor aggressiveness determined by a histopathological examination after radical prostatectomy.

Methods: Ninety-eight patients with prostate cancer underwent MRI at five different hospitals prior to radical prostatectomy. Images were retrospectively analyzed individually by two radiologists. The ADC of the index lesion and reference tissues (contralateral normal prostatic, normal peripheral zone, and urine) was recorded. Absolute ADC and different ADC ratios were compared to tumor aggressivity according to the ISUP Gleason Grade Groups extracted from the pathology report using Spearman's rank correlation coefficient (ρ). ROC curves were used to evaluate the ability to discriminate between ISUP 1-2 and ISUP 3-5 and intra class correlation and Bland-Altman plots for interrater reliability.

Results: All patients had prostate cancer classified as ISUP grade ≥ 2. No correlation was found between ADC and ISUP grade. We found no benefit of using the ADC ratio over absolute ADC. The AUC for all metrics was close to 0.5, and no threshold could be extracted for prediction of tumor aggressivity. The interrater reliability was substantial to almost perfect for all variables analyzed.

Conclusions: ADC and ADC ratio did not correlate with tumor aggressiveness defined by ISUP grade in this multicenter MRI study. The result of this study is opposite to previous research in the field.

KEYWORDS

MRI, MR-diffusion, ADC, neoplasms, prostate

Introduction

Prostate cancer (PCa) is the most common cancer in men worldwide (GLOBOCAN 2020) (1). However, most men with PCa have low-grade, indolent tumors. Therefore, discriminating between indolent and aggressive tumors is a diagnostic issue. With the traditional diagnostic approach, which includes a blood test of prostate specific antigen (PSA), digital rectal examination, and systematic transrectal ultrasound-guided biopsies, only a small and randomly distributed fraction of the gland is examined, resulting in a substantial risk of both over- and under-sampling. A more modern pathway involves magnetic resonance imaging (MRI) to detect clinically significant prostate cancer (csPCa) and rule out other causes of elevated PSA levels. On pathology, csPCa is defined as a Gleason score ≥7 (including 3 + 4 with a prominent but not predominant Gleason 4 component), volume ≥0.5 mL, and/or extra prostatic extension (2). Today, the International Society of Urological Pathology (ISUP) grade is often used to categorize different Gleason score patterns (3). When using MRI as a triage tool, unnecessary biopsies can be avoided, and targeted when required. This approach was investigated in the PRECISION study, which showed that MRI followed by targeted biopsies detected more significant tumors (38% versus 26%, p=0.005) and fewer insignificant tumors (9% versus 22%, p<0.001) compared to systematic biopsies (4). In the group that had an MRI in the work up, 28% had a negative MRI and, thus, did not have to undergo biopsy. These results changed the work-up routine, and MRI is now a cornerstone of PCa diagnosis. Therefore, the demands on MRI are high in terms of technical quality and radiological interpretation for correctly detecting or excluding csPCa.

Prostate Imaging – Reporting and Data System (PI-RADS, version 2.1) is a system that describes how to perform, interpret, and report MRI of the prostate (2). The most important MRI sequence is diffusion-weighted imaging (DWI), which is the deciding sequence in the peripheral zone (PZ) and the secondary sequence in the transition zone (TZ). DWI provides information on tissue composition and tumor cellularity (5). The signal intensity on DWI reflects the motion of water molecules in the tissue. The concept is based on the theory that a tumor consists of more dense tissue than normal prostatic tissue.

Several studies have shown that the ADC value inversely correlates with ISUP grade and is often used as a marker of aggressiveness (5–9). Several cut-off values have been proposed; in PIRADS 2.0, a threshold of 750-900 $\mu m^2/s$ was suggested as a pathological ADC value, but no consensus has been reached (8, 10). The concept is associated with several difficulties. First, the ADC varies substantially depending on several factors, including the b-values used, scanner field strength, patient and coil geometry, temporal fluctuations in the magnet, and variations in measurements between readers. Furthermore, non-cancerous lesions, such as benign prostate hypertrophy, may also exhibit decreased ADC values, and there is a substantial overlap in ADC values and PCa (11). ADC is sometimes used as a marker of aggressiveness in other organs and diseases. For example, in rectal adenocarcinoma, a lower ADC value is associated with a more aggressive tumor and poorer survival rate. Similar correlations have been found in certain types of breast cancer, ovarian cancer, lung cancer, and gliomas (12–15).

A common way to overcome the differences in absolute ADC values is to normalize the ADC by using different ADC ratios (10, 16). The ADC ratio is expressed as the ratio between the ADC value of the tumor and the ADC value of another location, such as non-cancerous tissue in the same organ or other organs in the same patient (5, 17).

In recent years, several studies have investigated the potential benefit of using the ADC ratio over absolute ADC values. Some authors have affirmed that the ADC ratio is the preferred method and demonstrated significant capability in discriminating Gleason 3 + 4 from 4 + 3 PCa (5, 8, 9, 16, 18). Other authors have been more doubtful (19).

The aim of the present study was to investigate, in a consecutive patient cohort imaged using different MRI scanners, how absolute ADC value and ADC ratios correlate with ISUP grade following robot-assisted laparoscopic prostatectomy (RALP). A secondary aim was to assess the potential inter-observer variability.

Material and methods

The study was a retrospective cohort study approved by the local ethics review committee at Lund University (Dnr 2014-886) and the Swedish ethical review authority (entry no. 2019-03674).

Study population

All consecutive patients who underwent RALP for biopsy proven PCa at Skåne University Hospital in Malmö, Sweden, during 2018 were identified and assessed for eligibility. Patients

were included if they had undergone MRI less than 1 year before surgery at five different hospitals. Patients were excluded if the index lesion described in the pathology report was not identified on MRI, severe artifacts were present on MRI, the MRI was performed outside of Region Skåne, or the patient opted out. Lesions were excluded based on consensus between two readers (JB and ET). The data collection algorithm is presented in Figure 1. Patient data were obtained from medical records.

Pathological examination

The surgical specimens were handled according to clinical routines and fixed in formalin. Lesions were examined by experienced pathologists using hematoxylin and eosin staining. Pathological data and whole mount (WM) tumor maps were obtained from the pathology report. The location and Gleason score of the index lesion were recorded using the ISUP category classification (Table 1).

MRI acquisition and image analysis

Preoperative MRI of the prostate was performed within Region Skåne using one of eight MRI scanners at five sites. Both 3T and 1.5T scanners were used. According to local routines, different imaging acquisition parameters were used at different sites. All protocols included transverse, coronal, and sagittal T2-weighted turbo spin-echo images, transverse T1-weighted images, diffusion-weighted images with a high b-value of 1500 s/mm^2, and a calculated ADC map. A list of MRI scanners and imaging acquisition parameters for the DWI are presented in Table 2.

Two readers, both specialists in radiology with 4 and 5 years of experience in reading prostate MRI, performed all imaging analyses as described below. The examinations were reviewed using the clinical Picture Archiving and Communication System, Sectra IDS7.

First, and in consensus, the two readers matched the index lesion in the surgical specimen with the corresponding lesion on MRI using the

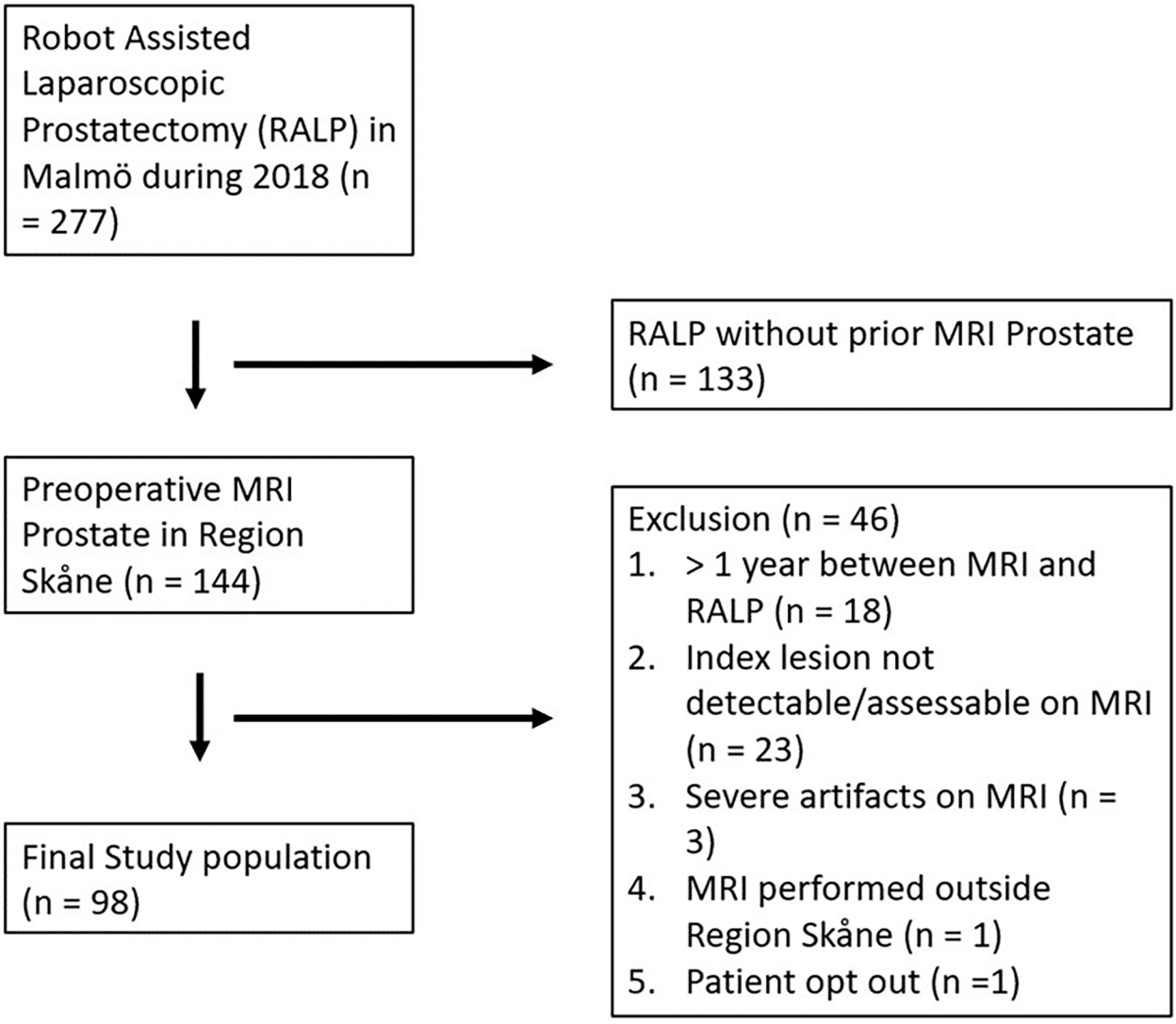


FIGURE 1
Flow diagram of patient selection.

TABLE 1 ISUP grade groups and the corresponding Gleason scores and patterns.

ISUP grade group	Gleason score	Gleason pattern
1	≤6	≤3+3
2	7	3+4
3	7	4+3
4	8	4+4, 3 + 5, 5 + 3
5	9 or 10	4+5, 5 + 4, or 5 + 5

pathological report and the whole-mount tumor map. In a second step, the remaining interpretation and image analyses were performed individually. For the index lesion, each reader recorded the maximum diameter in millimeters, zone location (PZ or TZ), and PI-RADS score (version 2.1). A circular region of interest (ROI) was placed in the index lesion in the ADC map on the slice with the largest cross-sectional area of tumor (ADC_{lesion}). The ROI was drawn to include only the lesion without any surrounding parenchyma. The size of the ROI was not fixed, it was drawn as big as possible within the defined lesion. A second ROI ($ADC_{contralat\ ref}$) of the same size was placed at the contralateral position on the same slice, that is in the same zone as the index lesion. A third and fourth ROI was placed in the most homogenous area in the PZ ($ADC_{PZ\ ref}$) and in the urinary bladder ($ADC_{urine\ ref}$), respectively. For each ROI, the mean ADC value was recorded (Figure 2).

Statistical analysis

Descriptive statistics were used to present the study population. Box plots and Spearman's rank correlation coefficient (ρ) were used to evaluate the association between ISUP grade and ADC variables. Measurements from reader 1 were used for the analyses of ADC metrics. These analyses were repeated and stratified by scanner field strength (1.5 vs. 3T) and tumor location (PZ vs. TZ). Receiver operating characteristic (ROC) curves were used to evaluate the ability to discriminate between ISUP 1-2 and ISUP 3-5 based on ADC variables. Interrater reliability was evaluated using Bland-Altman plots and intraclass correlation (ICC) based on the formula for random effects, absolute agreement, and single rater measurements. The ICC values were rated as follows: slight agreement, 0 – 0.20; fair agreement, 0.21 – 0.40; moderate agreement, 0.41 – 0.60; substantial agreement, 0.61 – 0.80; almost perfect agreement, 0.81 – 1.

All statistical analyses were performed in R version 4.0.2. The pROC package was used for ROC curves and the irr package to calculate ICC.

Results

A total of 144 men underwent RALP due to biopsy proven PCa and had an MRI prior to the procedure. After exclusion for different reasons (Figure 1), 98 patients were included in the final study analysis. The patient and tumor characteristics are presented in Table 3. No specimen was classified as ISUP 1. Most index lesions were located in the PZ of the prostate. Patients with different ISUP grades were relatively evenly distributed over the eight scanners, details are available in Supplementary Table 1.

ADC measurements vs. ISUP grade

The average ADC_{lesion} was 652×10^{-6} mm^2/s (range 396×10^{-6} mm^2/s to 1271×10^{-6} mm^2/s), whereas the average $ADC_{contralat\ ref}$ tissue was 1275×10^{-6} mm^2/s (range 779×10^{-6} mm^2/s to 1794×10^{-6} mm^2/s). The average $ADC_{PZ\ ref}$ was 1478×10^{-6} mm^2/s (range 779×10^{-6} mm^2/s to 2155×10^{-6} mm^2/s) and of $ADC_{urine\ ref}$ was 2021×10^{-6} mm^2/s (range 861×10^{-6} mm^2/s to 3368×10^{-6} mm^2/s; Figure 3). We found no significant negative correlation, between absolute the ADC value of the index lesion and the ISUP grade. The observed spearman correlation between the ADC of the index lesion and ISUP grade was low ($\rho = -0.18$) and not significant. Furthermore, the ADC of the index lesion did not perform well in discriminating between ISUP 1-2 and ISUP 3-5 (AUC= 0.62 [95% CI 0.51-0.74]). A tendency for a negative correlation was observed when the results from the 3T scanners were analyzed separately ($\rho = -0.27$; $p<0.05$), but not for the 1.5 T scanners ($\rho = -0.01$). Tables reporting the correlation values stratified by field strength are available in Supplementary Table 2. We found no correlation in separate analyses of the PZ and TZ.

The three different ADC ratios were calculated for each lesion ($ADC_{lesion}/ADC_{contralat\ ref}$, $ADC_{lesion}/ADC_{urine\ ref}$, and $ADC_{lesion}/ADC_{PZ\ ref}$ in relation to tumor aggressiveness. None of them showed any discriminatory effect (Figures 3, 4).

The agreement between the two readers in the ADC measurements was almost perfect for ADC_{lesion} (ICC of 0.80 [95% CI 0.72 – 0.86]), $ADC_{contralat\ ref}$ (ICC of 0.82 [95% CI 0.75 – 0.88]), and $ADC_{urine\ ref}$ (ICC of 0.96 [95% CI 0.94 – 0.97]). For $ADC_{PZ\ ref}$, the agreement was substantial (ICC of 0.75 [95% CI 0.65 – 0.86], Figure 5).

Discussion

This multi-scanner cohort study of 98 consecutive patients with MRI of the prostate before RALP showed no correlation between the absolute ADC value of the tumor and tumor aggressivity determined by pathology. No improvement was noted when the ADC value was normalized by applying different ADC ratios. Thus, no threshold values for ADC or ADC ratio were determined to discriminate significant from non-significant PCa. The inter-reader agreement between the two observers was substantial to almost perfect.

Different methods of interpretation have been applied to predict whether a lesion found on MRI represents benign tissue, non-significant cancer, or significant cancer. When comparing the results from the different studies, the definition of csPCa is crucial, as most authors try to define a threshold value for different ADC metrics in relation to tumor aggressivity. Some

TABLE 2 Overview of scanners and diffusion-weighted imaging (DWI) acquisition parameters.

Scanner	Patients, n (%)	Vendor	Model	Field strength, T	Sequence	B-values*, s/mm^2	TR, ms	TE, ms	Acquisition matrix	FOV, mm^2	Slice thickness, mm	Interslice gap, mm	DWI acquisition time, min
1	36 (36.7)	Siemens	TrioTim	3	2D EPI	50, 400, 1500	5200	87	128	240 x 240	3	3.6	Not applicable
2	12 (12.2)	Siemens	Skyra	3	2D EPI	0, 800, 1500	6700	75	130	240 x 240	3	4	5:50
3	12 (12.2)	Siemens	Prisma	3	2D EPI	0, 800, 1500	4500	66	130	240 x 240	3	3	3:56
4	24 (24.5)	Siemens	Avanto fit	1.5	2D EPI	50, 300, 1000, 1500	3300	67	128	240 x 240	3	4	8:15 + 5:33**
5	7 (7.1)	Siemens	Avanto fit	1.5	2D EPI	50, 300, 1000	2900	62	128	240 x 240	3	3	5:33 + 2:37**
6	3 (3.1)	GE	Optima4550r	1.5	2D EPI	50, 400, 800	4746	72	256	280 x 280	4	4.5	3:52
7	2 (2.0)	Siemens	Aera	1.5	Resolve	50, 400, 800	4620	57	116	200 x 200	4	4	6:25 + 6:19**
8	2 (2.0)	Siemens	Avanto fit	1.5	Resolve	50, 400, 800	4620	58	116	200 x 200	4	4	6:41 + 4:41**

EPI, echo-planar imaging; TR, repetition time; TE, echo time; FOV, field of view. *B-values included in the ADC calculation, **separate b1500.

10.3389/fonc.2023.1079040

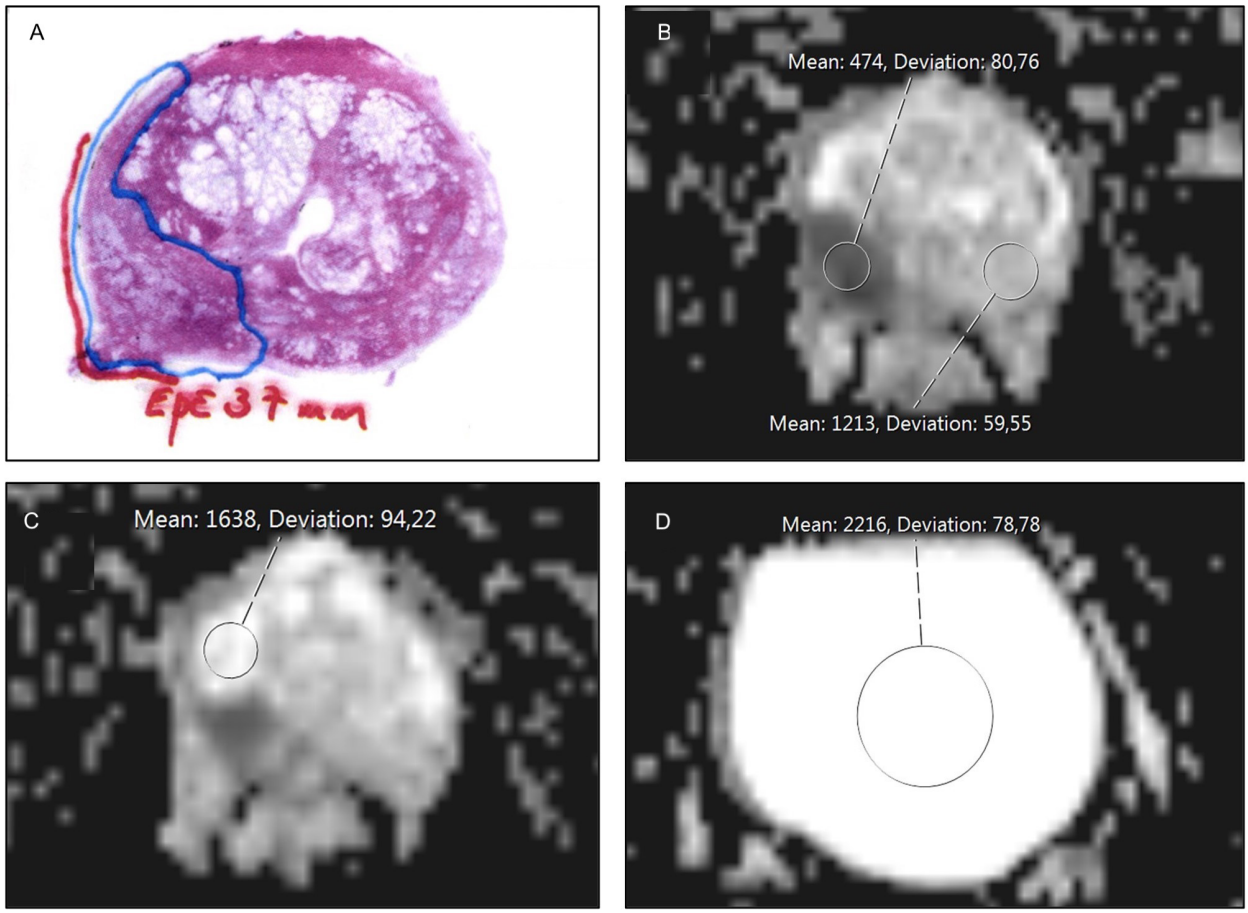


FIGURE 2
Example of a whole mount pathology specimen and placement of a region of interest (ROI) in the ADC map. The specimen was from a 70-year-old man with prostate cancer, PSA level 6.2 ng/mL, and clinical stage T3b. Systematic biopsies showed Gleason 4 + 5 (ISUP grade 5) in 7 of 12 cores. MRI was performed for staging and revealed a 2 x 3 cm PI-RADS 5 lesion in the right peripheral zone (PZ) with findings in line with extraprostatic extension (EPE) and seminal vesicle invasion (SVI). The final staging was pT3a. **(A)** Midgland whole mount specimen with a large tumor in the right PZ (blue border) with 37 mm EPE (red line). **(B)** Circular ROI in tumor (ADC_{lesion} = 474 x 10^{-6} mm^2/s) and in contralateral non-tumorous tissue ($ADC_{contralat\ ref}$ = 1213 x 10^{-6} mm^2/s). **(C)** Circular ROI drawn in non-tumorous PZ ($ADC_{PZ\ ref}$ = 1638 x 10^{-6} mm^2/s). **(D)** Circular ROI in urinary bladder ($ADC_{urine\ ref}$ = 2216 x 10^{-6} mm^2/s). Tumor to non-tumor ratio = 0.36, tumor to PZ ratio = 0.29, and tumor to urinary bladder ratio = 0.21.

papers have used ISUP grade 1 as non-significant and ISUP 2 and higher as significant (20–24), whereas others have included ISUP 2 in the non-significant group. One study even included all ISUP 2 and 3 in the more harmless group and used the terms intermediate and high-risk cancer as the border between the two groups (25). Boesen et al. performed their analyses on two different cut-offs with ISUP 2 in both the significant and non-significant groups (8). In our study, all resected prostates were ISUP 2 or higher, which gave us no choice to use only ISUP 1 in the non-significant group. This was also true for the 23 patients in whom the index lesion could not be identified on MRI.

Regardless of which definition of csPCa is used, several authors have reported a strong inverse correlation between ADC metrics and tumor aggressivity, with a reported AUC of up to 0.94 (26) or 0.96 (17). This contrasts with the results of our study, as we found an AUC of 0.62, which would suggest that the absolute ADC value is not useful for predicting the presence of csPCa. The reasons for these results can be debated. We used eight different MRI scanners with different acquisition parameters. Disparate absolute ADC values are not unexpected with these settings. Barret et al. calculated different ADC values from the same scans by combining four b-values in different ways, thereby simulating different parameters (5). Most combinations showed a relatively good inverse correlation with tumor aggressivity. When they used the ratio between tumorous and non-tumorous ADC values, the differences in acquisition parameters were less obvious. Thus, they stated that the ADC ratio may be considered a more robust tool for assessing restricted diffusion in the prostate (5). With the same intention, we evaluated whether the disparate ADC values between our scanners could be more useful when different ratios were applied. However, despite using three different tissues as denominators in the creation of the ratios, no added value or better performance were found for the metrics. In fact, the AUC was even smaller, close to 0.5 for all three ratios, which is slightly smaller than for the absolute ADC. For the 1.5T scanners there was a tendency of positive correlation, instead of the expected negative correlation, between ADC ratio and ISUP grade.

TABLE 3 Patient characteristics (n=98).

Characteristic	*Mean ± SD (min – max)*
Age, years	66.3 ± 6.4 (45 – 76)
Time between MRI and RALP, months	4.08 ± 2.6 (1 – 11)
Preoperative PSA, ng/mL	9.26 ± 6.8 (1.8 – 39.0)
	n (%)
Clinical T-stage	
T0	4 (4.1)
T1	13 (13.3)
T1c	29 (29.6)
T2	36 (36.7)
T2b	4 (4.1)
T2c	2 (2.0)
T3	9 (9.2)
T3a	1 (1.0)
Pathological T-stage	
T1	0 (0)
T2	51 (52.0)
T3a	34 (34.7)
T3b	12 (12.2)
T3	0 (0)
Missing	1 (1.0)
Biopsy ISUP grade	
1	7 (7.1)
2	41 (41.8)
3	24 (24.5)
4	9 (9.2)
5	17 (17.3)
Pathological ISUP grade	
1	0 (0)
2	39 (39.8)
3	41 (41.8)
4	3 (3.1)
5	15 (15.3)
MRI field strength	
1.5 Tesla	38 (38.8)
3 Tesla	60 (60.1)
Zone location	
Peripheral zone	68 (69.4)
Transitional zone	30 (30.6)

Several other authors have claimed that the ratio, often tumor versus the contralateral normal appearing tissue, is better than the absolute ADC value. Lebovici et al. showed the usefulness of an ADC ratio in differentiating low-grade and high-grade disease (25). Similar results were reported by Boesen et al. and Litjens et al. (8, 27). Interestingly, both absolute ADC values and the ADC ratios differed considerably between these studies. Itatani et al. assessed 58 men who underwent RALP after MRI and used the internal obturator muscle as the ADC reference, finding superior use of the ratio (AUC 0.85 vs. 0.71) (28). Bajgiran et al. concluded that the ADC ratio is a more robust biomarker of PCa aggressiveness (21). Conversely, Rosencrantz et al. found no benefit of using the ADC ratios with urine ADC as the denominator for differentiating benign and malignant tissue in the PZ (17). Woo et al. (20) included 165 men, and DeCobelli 72 men (26), with contralateral prostatic tissue as the reference and found no benefit of the ADC ratio compared to standalone ADC.

Woo et al. pointed out several reasons why the use of the ADC value for internal reference organs may not yield helpful ADC ratios and thereby add, rather than reduce, sources of error in the interpretation (20). For example, they emphasize that the ADC value of the non-tumor PZ can vary according to age, and that the intrinsically organized chaos of the TZ results in a wide range of normal ADC values (29). Moreover, post-biopsy changes can alter the signal intensity of DWI in the prostatic tissue for several weeks. Finally, as hypothesized by DeCobelli, non-tumorous tissue can be affected by nearby non-visible tumor infiltration or by peritumoral fibrosis and inflammation, which all affect the ADC (26). The b-values that were used to estimate the ADC (Table 2) varied across MRI systems and sites, and several were inconsistent with PI-RADS recommendations (2). For example, the estimation of ADC based on data acquired at low b-values (<100 s/mm^2) may introduce a positive bias due to incoherent blood perfusion (30). Furthermore, when the ADC is based on high b-values (>1000 s/mm^2), the estimation in normal tissue may be negatively biased due to the rectified noise floor (31). These factors may explain why the ratio did not show a better inverse correlation with cancer aggressiveness than standalone ADC. Moreover, in a systematic review of 39 papers with 2457 patients, Surov et al. identified only a moderate correlation between ADC and Gleason score in PCa located in the PZ, and an even worse correlation in the TZ (32).

Harmonizing MRI parameters between centers is important, especially since the ADC values are used for deciding PI-RADS category and hence, affects the clinical decision. In 2007, the Radiological Society of North America organized The Quantitative Imaging Biomarkers Alliance ® (QIBA). QIBA strives for standardization of image acquisition and assesses whether imaging metrics have clinical value (33). Their ongoing work includes evaluation and standardization of DWI in for example MRI Prostate.

In our study, the interrater agreements for different ADC metrics were strong, suggesting that factors other than differences in radiologists' measurements are the reason for the lack of correlation with pathology. Our results are in line with similar previous studies (19, 23, 34).

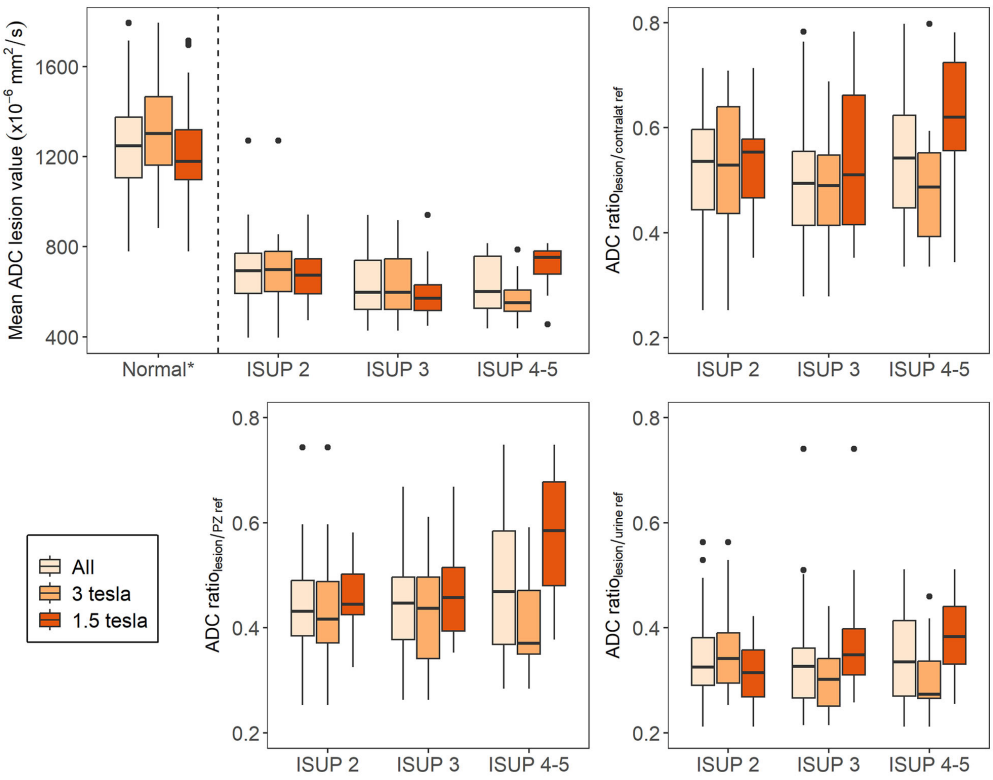


FIGURE 3
Box-and-whisker plots of apparent diffusion coefficient (ADC) metrics for tumors stratified by ISUP grade. (*) *Normal* represents the absolute ADC value of the normal appearing tissue in the contralateral position of the index lesion.

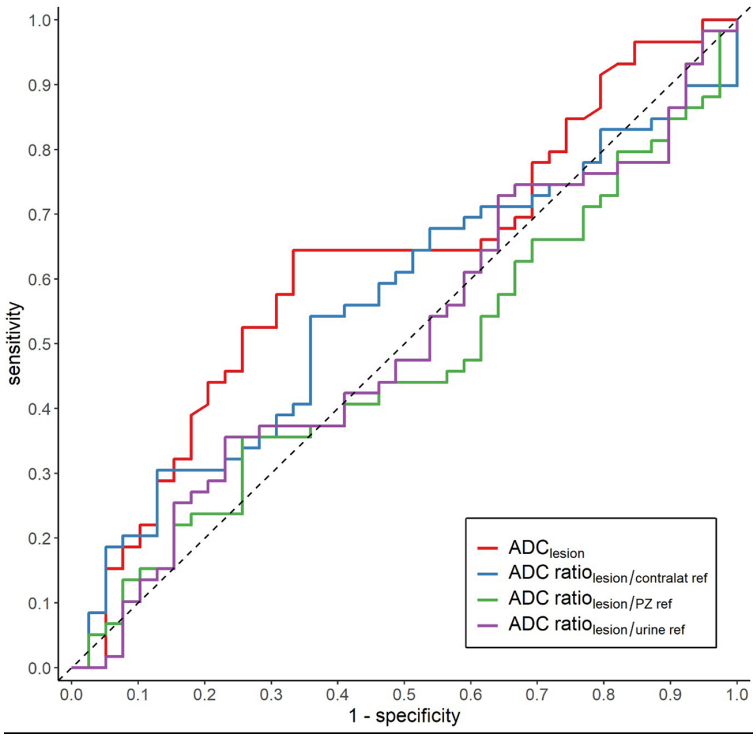


FIGURE 4
ROC curves comparing absolute ADC and three different ADC ratios in discriminating ISUP 1-2 from ISUP 3-5.

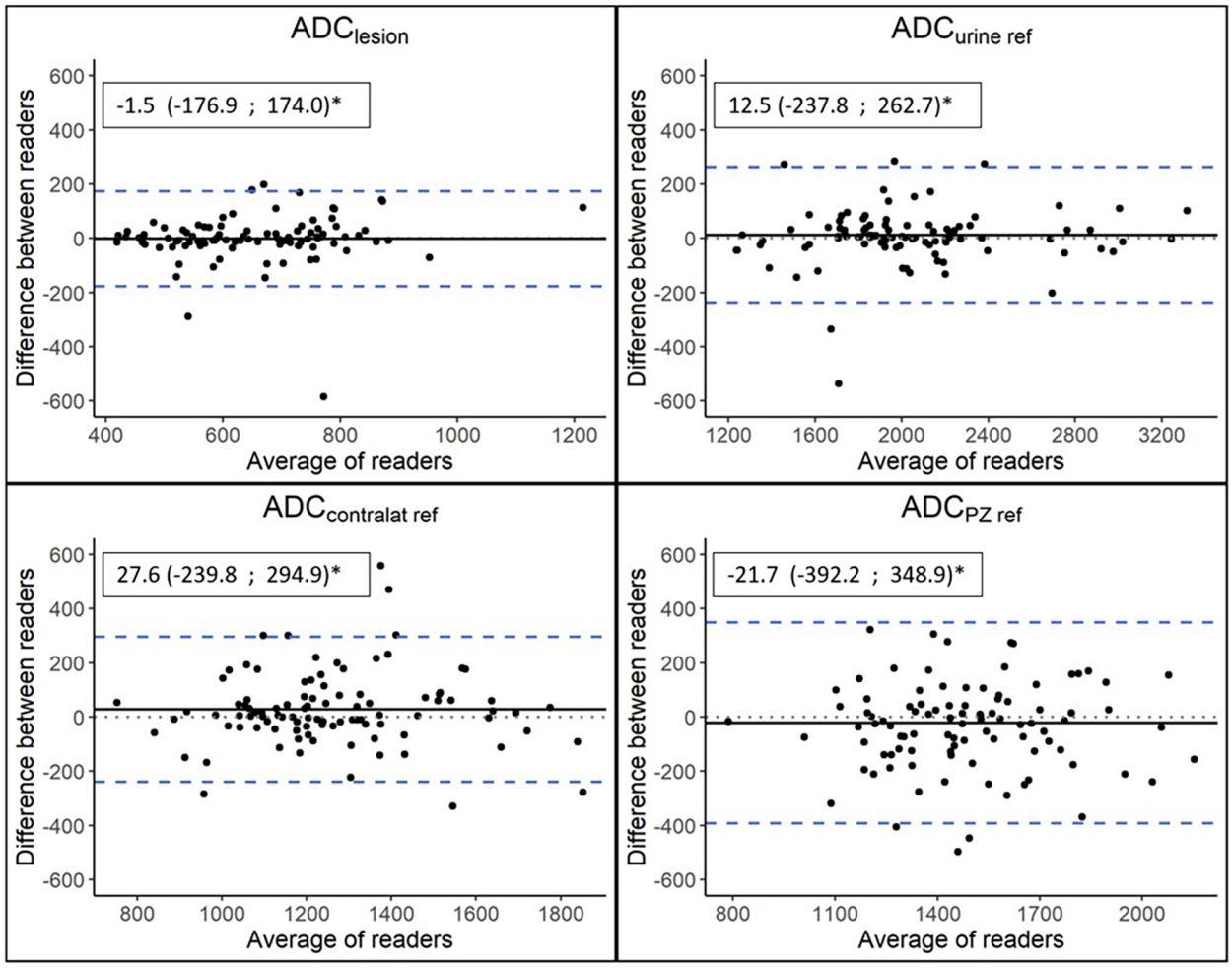


FIGURE 5
Bland-Altman plots. The dotted lines represent no difference between readers, the solid lines represent mean differences between readers, and the dashed blue lines represent limits of agreement, calculated as the mean difference ± 1.96SD of the mean difference. *Mean differences between readers (95% limits of agreement).

Our study has several limitations. First, the study group was small. In addition, the quality of the MRI scans was generally lower than would have been acceptable today. Another limitation is that all included patients had csPCa; therefore, we only obtained data from the more advanced and aggressive tumors. In contrast to previous articles on this topic, no patients with ISUP 1 tumors were subject to prostate resection. This is in line with current clinical treatment guidelines (35). Furthermore, we did not have information on the fraction of Gleason 4 in the ISUP 2 group (Gleason 3 + 4). A lower percentage of Gleason 4 could have put these patients in the group with non-significant cancers. Moreover, the results from pathology were extracted from the original pathology reports, which were produced in a clinical setting by different pathologists with different levels of experience. That is, no study-dedicated pathology examination was performed.

There is potential for improvement, which we will implement in a forthcoming study. Most important is to include the whole range of benign to the most aggressive tumors. This can be achieved by including core biopsies performed using the MR – ultrasound fusion technique. Furthermore, with new digital pathology archives, high precision correlations can be made between the WM RALP specimen and corresponding MR slice. A dedicated revaluation of a specific location in the WM specimen, including tumor subtype and tumor cell growth pattern, can be made.

Conclusions

In conclusion, our study did not find any correlation between the ADC value and ISUP grade in a multi-scanner setting. We found no benefit of using ADC ratios, so-called normalized ADC values, even with good agreement between the two experienced readers. This contradicts previous single-center studies published research in the field. Therefore, in a clinical situation with different MRI scanner types, measurements of ADC must be used with caution. It also highlights the importance of harmonizing the parameters of the MRI sequences across centers.

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 Lund University (Dnr 2014-886) Swedish ethical review authority (entry no. 2019-03674). Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

Author contributions

JB, ET, DF-S, AB and SZ conceived of the presented idea. JB and ET performed the measurements. DFS retrieved clinical data from patient records. JB, ET, SZ processed the experimental data and performed the analyses. SZ and EB supervised the work. All authors contributed to the writing of the manuscript and approved the submitted version.

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2023.1079040/full#supplementary-material

References

1. Cancer IAfRo. *Global cancer observatory*. Lyon, France: International Agency for Research on Cancer (2020).

2. American College of Radiology Committee on PI-RADS® (Prostate). (2019). Available at: https://www.acr.org/-/media/ACR/Files/RADS/PI-RADS/PIRADS-V2-1.pdf Accessed on October 1, 2022.

3. Epstein JI, Egevad L, Amin MB, Delahunt B, Srigley JR, Humphrey PA. The 2014 international society of urological pathology (ISUP) consensus conference on Gleason grading of prostatic carcinoma. *Am J Surg pathol* (2016) 40(2):244–52. doi: 10.1097/PAS.0000000000000530

4. Kasivisvanathan V, Rannikko AS, Borghi M, Panebianco V, Mynderse LA, Vaarala MH, et al. MRI-Targeted or standard biopsy for prostate-cancer diagnosis. *N Engl J Med* (2018) 378(19):1767–77. doi: 10.1056/NEJMoa1801993

5. Barrett T, Priest AN, Lawrence EM, Goldman DA, Warren AY, Gnanapragasam VJ, et al. Ratio of tumor to normal prostate tissue apparent diffusion coefficient as a method for quantifying DWI of the prostate. *AJR Am J Roentgenol* (2015) 205(6):W585–93. doi: 10.2214/AJR.15.14338

6. Hambrock T, Somford DM, Huisman HJ, van Oort IM, Witjes JA, Hulsbergen-van de Kaa CA, et al. Relationship between apparent diffusion coefficients at 3.0-T MR imaging and Gleason grade in peripheral zone prostate cancer. *Radiology* (2011) 259(2):453–61. doi: 10.1148/radiol.11091409

7. Verma S, Rajesh A, Morales H, Lemen L, Bills G, Delworth M, et al. Assessment of aggressiveness of prostate cancer: Correlation of apparent diffusion coefficient with histologic grade after radical prostatectomy. *AJR Am J Roentgenol* (2011) 196(2):374–81. doi: 10.2214/AJR.10.4441

8. Boesen L, Chabanova E, Logager V, Balslev I, Thomsen HS. Apparent diffusion coefficient ratio correlates significantly with prostate cancer gleason score at final pathology. *J Magn Reson Imaging* (2015) 42(2):446–53. doi: 10.1002/jmri.24801

9. Hötker AM, Mazaheri Y, Aras Ö, Zheng J, Moskowitz CS, Gondo T, et al. Assessment of prostate cancer aggressiveness by use of the combination of quantitative DWI and dynamic contrast-enhanced MRI. *Am J Roentgenol* (2016) 206(4):756–63. doi: 10.2214/AJR.15.14912

10. Manetta R, Palumbo P, Gianneramo C, Bruno F, Arrigoni F, Natella R, et al. Correlation between ADC values and Gleason score in evaluation of prostate cancer: Multicentre experience and review of the literature. *Gland Surg* (2019) 8(Suppl 3):S216–S22. doi: 10.21037/gs.2019.05.02

11. Hoeks CM, Hambrock T, Yakar D, Hulsbergen–van de Kaa CA, Feuth T, Witjes JA, et al. Transition zone prostate cancer: Detection and localization with 3-T multiparametric MR imaging. *Radiology* (2013) 266(1):207–17. doi: 10.1148/radiol.12120281

12. Costantini M, Belli P, Rinaldi P, Bufi E, Giardina G, Franceschini G, et al. Diffusion-weighted imaging in breast cancer: Relationship between apparent diffusion coefficient and tumour aggressiveness. *Clin Radiol* (2010) 65(12):1005–12. doi: 10.1016/j.crad.2010.07.008

13. Wu C-C, Jain R, Radmanesh A, Poisson LM, Guo W-Y, Zagzag D, et al. Predicting genotype and survival in glioma using standard clinical MR imaging apparent diffusion coefficient images: A pilot study from the cancer genome atlas. *Am J Neuroradiol* (2018) 39(10):1814–20. doi: 10.3174/ajnr.A5794

14. Surov A, Meyer HJ, Wienke A. Correlation between apparent diffusion coefficient (ADC) and cellularity is different in several tumors: A meta-analysis. *Oncotarget* (2017) 8(35):59492. doi: 10.18632/oncotarget.17752

15. Sun Y, Tong T, Cai S, Bi R, Xin C, Gu Y. Apparent diffusion coefficient (ADC) value: a potential imaging biomarker that reflects the biological features of rectal cancer. *PloS One* (2014) 9(10):e109371. doi: 10.1371/journal.pone.0109371

16. Alessandrino F, Taghipour M, Hassanzadeh E, Ziaei A, Vangel M, Fedorov A, et al. Predictive role of PI-RADSv2 and ADC parameters in differentiating Gleason pattern 3 + 4 and 4 + 3 prostate cancer. *Abdom Radiol (NY)* (2019) 44(1):279–85. doi: 10.1007/s00261-018-1718-6

17. Rosenkrantz AB, Kopec M, Kong X, Melamed J, Dakwar G, Babb JS, et al. Prostate cancer vs. post-biopsy hemorrhage: Diagnosis with T2- and diffusion-weighted imaging. *J Magn Reson Imaging* (2010) 31(6):1387–94. doi: 10.1002/jmri.22172

18. Tamada T, Prabhu V, Li J, Babb JS, Taneja SS, Rosenkrantz AB. Assessment of prostate cancer aggressiveness using apparent diffusion coefficient values: Impact of patient race and age. *Abdom Radiol (NY)* (2017) 42(6):1744–51. doi: 10.1007/s00261-017-1058-y

19. Abreu-Gomez J, Walker D, Alotaibi T, McInnes MDF, Flood TA, Schieda N. Effect of observation size and apparent diffusion coefficient (ADC) value in PI-RADS v2.1 assessment category 4 and 5 observations compared to adverse pathological outcomes. *Eur Radiol* (2020) 30(8):4251–61. doi: 10.1007/s00330-020-06725-9

20. Woo S, Kim SY, Cho JY, Kim SH. Preoperative evaluation of prostate cancer aggressiveness: Using ADC and ADC ratio in determining Gleason score. *AJR Am J Roentgenol* (2016) 207(1):114–20. doi: 10.2214/AJR.15.15894

21. Bajgiran AM, Mirak SA, Sung K, Sisk AE, Reiter RE, Raman SS. Apparent diffusion coefficient (ADC) ratio versus conventional ADC for detecting clinically significant prostate cancer with 3-T MRI. *AJR Am J Roentgenol* (2019) 213(3):W134–W42. doi: 10.2214/AJR.19.21365

22. Ragheb SR, Bassiouny RH. Can mean ADC value and ADC ratio of benign prostate tissue to prostate cancer assist in the prediction of clinically significant prostate

cancer within the PI-RADSv2 scoring system? *Egypt J Radiol Nucl Med* (2020) 51(1). doi: 10.1186/s43055-020-00347-3

23. Falaschi Z, Valenti M, Lanzo G, Attanasio S, Valentini E, Garcia Navarro LI, et al. Accuracy of ADC ratio in discriminating true and false positives in multiparametric prostatic MRI. *Eur J Radiol* (2020) 128:109024. doi: 10.1016/j.ejrad.2020.109024

24. Pepe P, D'Urso D, Garufi A, Priolo G, Pennisi M, Russo G, et al. Multiparametric MRI apparent diffusion coefficient (ADC) accuracy in diagnosing clinically significant prostate cancer. *vivo* (2017) 31(3):415–8. doi: 10.21873/invivo.11075

25. Lebovici A, Sfrangeu SA, Feier D, Caraiani C, Lucan C, Suciu M, et al. Evaluation of the normal-to-diseased apparent diffusion coefficient ratio as an indicator of prostate cancer aggressiveness. *BMC Med imaging* (2014) 14(1):1–7. doi: 10.1186/1471-2342-14-15

26. De Cobelli F, Ravelli S, Esposito A, Giganti F, Gallina A, Montorsi F, et al. Apparent diffusion coefficient value and ratio as noninvasive potential biomarkers to predict prostate cancer grading: comparison with prostate biopsy and radical prostatectomy specimen. *AJR Am J Roentgenol* (2015) 204(3):550–7. doi: 10.2214/AJR.14.13146

27. Litjens GJ, Hambrock T, Hulsbergen–van de Kaa C, Barentsz JO, Huisman HJ. Interpatient variation in normal peripheral zone apparent diffusion coefficient: Effect on the prediction of prostate cancer aggressiveness. *Radiology* (2012) 265(1):260–6. doi: 10.1148/radiol.12112374

28. Itatani R, Namimoto T, Yoshimura A, Katahira K, Noda S, Toyonari N, et al. Clinical utility of the normalized apparent diffusion coefficient for preoperative evaluation of the aggressiveness of prostate cancer. *Japan J Radiol* (2014) 32 (12):685–91. doi: 10.1007/s11604-014-0367-0

29. Zhang J, Tian W-Z, Hu C-H, Niu T-L, Wang X-L, Chen X-Y. Age-related changes of normal prostate: Evaluation by MR diffusion tensor imaging. *Int J Clin Exp Med* (2015) 8(7):11220.

30. Le Bihan D. What can we see with IVIM MRI? *Neuroimage* (2019) 187:56–67. doi: 10.1016/j.neuroimage.2017.12.062

31. Gudbjartsson H, Patz S. The rician distribution of noisy MRI data. *Magn Reson Med* (1995) 34(6):910–4. doi: 10.1002/mrm.1910340618

32. Surov A, Meyer HJ, Wienke A. Correlations between apparent diffusion coefficient and Gleason score in prostate cancer: A systematic review. *Eur Urol Oncol* (2020) 3(4):489–97. doi: 10.1016/j.euo.2018.12.006

33. Shukla-Dave A, Obuchowski NA, Chenevert TL, Jambawalikar S, Schwartz LH, Malyarenko D, et al. Quantitative imaging biomarkers alliance (QIBA) recommendations for improved precision of DWI and DCE-MRI derived biomarkers in multicenter oncology trials. *J Magn Reson Imaging* (2019) 49(7):e101–e21. doi: 10.1002/jmri.26518

34. Park KJ, Kim MH, Kim JK. Extraprostatic tumor extension: Comparison of preoperative multiparametric MRI criteria and histopathologic correlation after radical prostatectomy. *Radiology* (2020) 296(1):87–95. doi: 10.1148/radiol.2020192133

35. Mottet N, Van den Bergh R, Briers E, Cornford P, De Santis M, Fanti S, et al. Eau-Eanm-Estro-Esur-Siog guidelines on prostate cancer. *Eur Assoc Urol* (2020) 1–182.

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 28 February 2023
DOI 10.3389/fonc.2023.1120392

OPEN ACCESS

EDITED BY
Jasper Nijkamp,
Aarhus University, Denmark

REVIEWED BY
Mark J. Gooding,
Inpictura Ltd, United Kingdom
Elisa D'Angelo,
University Hospital of Modena, Italy

*CORRESPONDENCE
Clifton D. Fuller
cdfuller@mdanderson.org
Kimmo Kaski
kimmo.kaski@aalto.fi

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

RECEIVED 09 December 2022
ACCEPTED 13 February 2023
PUBLISHED 28 February 2023

CITATION
Sahlsten J, Wahid KA, Glerean E, Jaskari J, Naser MA, He R, Kann BH, Mäkitie A, Fuller CD and Kaski K (2023) Segmentation stability of human head and neck cancer medical images for radiotherapy applications under de-identification conditions: Benchmarking data sharing and artificial intelligence use-cases. *Front. Oncol.* 13:1120392. doi: 10.3389/fonc.2023.1120392

Segmentation stability of human head and neck cancer medical images for radiotherapy applications under de-identification conditions: Benchmarking data sharing and artificial intelligence use-cases

Jaakko Sahlsten[1], Kareem A. Wahid[2], Enrico Glerean[3], Joel Jaskari[1], Mohamed A. Naser[2], Renjie He[2], Benjamin H. Kann[4], Antti Mäkitie[5], Clifton D. Fuller[2*] and Kimmo Kaski[1*]

[1]Department of Computer Science, Aalto University School of Science, Espoo, Finland, [2]Department of Radiation Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, United States, [3]Department of Neuroscience and Biomedical Engineering, Aalto University, Espoo, Finland, [4]Artificial Intelligence in Medicine Program, Brigham and Women's Hospital, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA, United States, [5]Department of Otorhinolaryngology, Head and Neck Surgery, University of Helsinki and Helsinki University Hospital, Helsinki, Finland

Background: Demand for head and neck cancer (HNC) radiotherapy data in algorithmic development has prompted increased image dataset sharing. Medical images must comply with data protection requirements so that re-use is enabled without disclosing patient identifiers. Defacing, i.e., the removal of facial features from images, is often considered a reasonable compromise between data protection and re-usability for neuroimaging data. While defacing tools have been developed by the neuroimaging community, their acceptability for radiotherapy applications have not been explored. Therefore, this study systematically investigated the impact of available defacing algorithms on HNC organs at risk (OARs).

Methods: A publicly available dataset of magnetic resonance imaging scans for 55 HNC patients with eight segmented OARs (bilateral submandibular glands, parotid glands, level II neck lymph nodes, level III neck lymph nodes) was utilized. Eight publicly available defacing algorithms were investigated: afni_refacer, DeepDefacer, defacer, fsl_deface, mask_face, mri_deface, pydeface, and quickshear. Using a subset of scans where defacing succeeded (N=29), a 5-fold cross-validation 3D U-net based OAR auto-segmentation model was utilized to perform two main experiments: 1.) comparing original and defaced data for training when evaluated on original data; 2.) using original data for training and comparing the model evaluation on original and defaced data. Models were primarily assessed using the Dice similarity coefficient (DSC).

Results: Most defacing methods were unable to produce any usable images for evaluation, while mask_face, fsl_deface, and pydeface were unable to remove the face for 29%, 18%, and 24% of subjects, respectively. When using the original data for evaluation, the composite OAR DSC was statistically higher ($p \leq 0.05$) for the model trained with the original data with a DSC of 0.760 compared to the mask_face, fsl_deface, and pydeface models with DSCs of 0.742, 0.736, and 0.449, respectively. Moreover, the model trained with original data had decreased performance ($p \leq 0.05$) when evaluated on the defaced data with DSCs of 0.673, 0.693, and 0.406 for mask_face, fsl_deface, and pydeface, respectively.

Conclusion: Defacing algorithms may have a significant impact on HNC OAR auto-segmentation model training and testing. This work highlights the need for further development of HNC-specific image anonymization methods.

KEYWORDS

anonymization, radiotherapy, head and neck cancer, MRI, medical imaging, artificial intelligence (AI), autosegmentation, defacing

Introduction

The landscape of data democratization is rapidly changing. The rise of open science practices, inspired by coalitions such as the Center for Open Science (1), and the FAIR (Findable, Accessible, Interoperable, and Reusable) guiding principles (2), has spurred interest in public data sharing. Subsequently, the medical imaging community has increasingly adopted these practices through initiatives such as The Cancer Imaging Archive (3). Given the appropriate removal of protected health information through anonymization techniques, public repositories have democratized the access to medical imaging data such that the world at large can now help develop algorithmic approaches to improve clinical decision-making. Among the medical professions seeking to leverage these large datasets, radiation oncology has the potential to vastly benefit from these open science practices (4). Imaging is crucial to radiotherapy workflows, particularly for organ at risk (OAR) and tumor segmentation (5, 6). Moreover, in recent years public data competitions, such as the Head and Neck Tumor Segmentation and Outcome Prediction in positron emission tomography/computed tomography (PET/CT) Images (HECKTOR) challenge (7–9), have been targeted to improve the radiotherapy workflow. However, there is a particular facet of medical image dissemination for radiotherapy applications that has spurred controversy, namely the anonymization of head and neck cancer (HNC) related images.

While the public dissemination of HNC image data is invaluable to improve the radiotherapy workflow, concerns have been raised regarding readily identifiable facial features on medical imaging. Importantly, the U.S. Health Insurance Portability and Accountability Act references "full-face photographs and any comparable images" as a part of protected health information (10). This policy introduces some uncertainty in the dissemination of high-resolution images, where the intricacies of facial features can be reconstructed to generate similar or "comparable" visualizations with relative ease. Several studies have shown the potential danger in releasing unaltered medical images containing facial features, as they can often be easily recognized by humans and/or machines (11–15). For example, using facial recognition software paired with image-derived facial reconstructions, one study found up to 83% of research participants could be identified from their magnetic resonance imaging (MRI) scans (13). Similar alarming results have been demonstrated for CT images (14). While brain images are often processed such that obvious facial features are removed (i.e., skull stripping), these crude techniques remove large anatomic regions necessary for building predictive models with HNC imaging data. "Defacing" tools, where voxels that correspond to the areas of the patient's facial features are either removed or altered, offer one solution. However, they may still engender the potential loss of voxel-level information needed for predictive modeling or treatment planning, thereby prohibiting their use in data resharing strategies for radiotherapy applications. While several studies have investigated the effects of defacing for neuroimaging (16–21), there have not yet been any systematic studies on the effects of defacing tools for radiotherapy applications.

Inspired by the increasing demand for public HNC imaging datasets and the importance of protecting the privacy of patients, a systematic analysis of a number of existing methods for facial anonymization on HNC MRI images was performed. Through qualitative and quantitative analysis using open-source datasets and tools, the efficacies of defacing approaches on whole images and structures relevant to radiation treatment planning were determined. Moreover, the effects of these approaches on auto-segmentation, a specific domain application that is increasingly relevant for HNC public datasets, were also examined. This study is

an important first step towards the development of robust approaches for the safe and trusted democratization of HNC imaging data.

Methods

Dataset

For this analysis, a publicly available dataset hosted on the TCIA, the American Association of Physicists in Medicine RT-MAC Grand Challenge 2019 (AAPM) dataset (22), was utilized. The AAPM dataset consists of T2-weighted MRI scans of 55 HNC patients that are labeled for OAR segmentations of bilateral: i) submandibular glands, ii) level II neck lymph nodes, iii) level III neck lymph nodes, and iv) parotid glands. Structures were annotated as being on the right or left side of the patient anatomy. The spatial resolution of the scans is 0.5 mm × 0.5 mm with 2.0 mm spacing. Additional technical details on the AAPM images and segmentations can be found in the corresponding data descriptor (22). Defacing experiments were also attempted using the HECKTOR 2021 training dataset (8) containing 224 HNC patients with CT scans. Additional technical details on the HECKTOR dataset can be found in the corresponding overview papers (8, 9).

Defacing methods

For defacing the images, the same methods as taken into consideration by Schwartz et al. (16), as well as novel tools that benefit from recent advances in deep learning were used. The most popular tools use a co-registration to a template in order to identify face and ears and then identify those structures in the original image, which should be removed or blurred. The following 6 co-registration based methods: *afni_refacer*, *fsl_deface* (23), *mask_face* (24), *mri_deface* (18), *pydeface* (25), and *quickshear* were implemented. Two more recent methods using deep learning technology were also included: *defacer* (26) and *DeepDefacer* (27). These methods utilize pre-trained deep learning models using data from public neuroimaging datasets to identify facial features to be removed. An automated pipeline for applying all these defacing methods is available at https://github.com/eglerean/faceai_testingdefacing. Each defacing method was tested with all subjects such that, for each subject, a defaced volume was produced as well as a volumetric mask of which voxels were affected by defacing. All methods were run with the default parameters and standard reference images.

Defacing performance

After applying the defacing methods, the success or failure of a defacing method was determined by visually inspecting all the defaced volumes (i.e., performing scanwise quality control). Specifically, a binary categorization of each scan was implemented: "1" if the eyes, nose, and mouth were removed (i.e., defacing succeeded), "0" if the eyes, nose, or mouth were not removed (i.e., defacing failed). Subsequently, the amount of voxels present in the structures after application of the defacing algorithm were quantitatively measured.

Deep learning model for OAR segmentation reliability

To evaluate the OAR segmentation performance under different defacing schemes from volumetric MRI data, a convolutional neural network architecture, 3D U-net, which has found wide success in HNC-related segmentation tasks (28–33), was utilized. Both contractive and expansive pathways include four blocks, where each block consists of two convolutional layers with a kernel size of 3, and each convolution is followed by an instance normalization layer and a LeakyReLU activation with 0.1 negative slope. The max-pooling and transpose convolutional layers have a kernel size and stride of 2. The last convolutional layer has a kernel size and stride of 1 with 9 output channels and a softmax activation. The model architecture is shown in Figure 1. Experiments were developed in Python v. 3.6.10 (34) using Pytorch 1.8.1 (35) with a U-net model from Project MONAI 0.7.0 (36) and data preprocessing and augmentation with TorchIO 0.18.61 (37).

A subset of patients for which defacing was deemed successful were used for building the segmentation models. The subset was randomly split with 5-fold cross validation: for each cross-validation iteration one fold was used for model testing, one fold was used for model validation, and the remaining three folds were used for model training. The reported segmentation performance

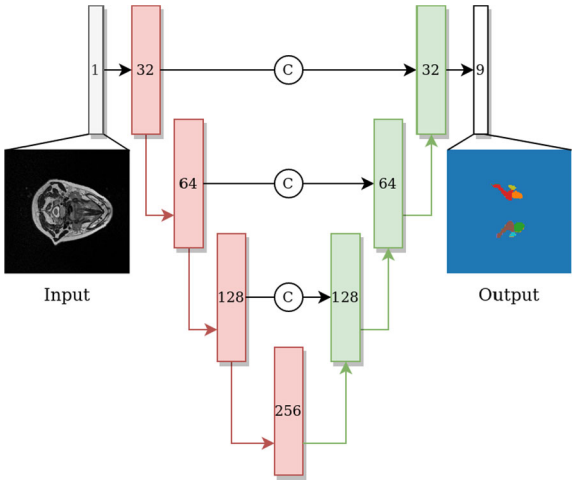


FIGURE 1
U-net network architecture with blocks on the contractive path colored in red and blocks on the expanding path colored in green. Each block includes two convolutions, each followed by instance normalization and Leaky ReLU activation, subsequently followed by a max-pool layer (red arrow) or transpose convolution layer (green arrow) on contractive and expanding paths, respectively. The number shown in each block indicates the number of channels of the feature map. Arrows with the letter C indicate concatenation.

was based on the test fold that was not used for model development. The same random splits were used for training and evaluating the models trained on original or defaced data.

Data preprocessing after the defacing included linear resampling to 2 mm isotropic resolution with the intensity scaled into a range of [-1,1]. The training data was augmented with random transforms that were applied with a probability (p), independently of each other. The used transforms were random elastic deformations (p=10%) for all axes, random flips for inferior-superior and anterior-posterior axes (p=50%), random rotation (-10° to 10°) of all axes (p=50%), random bias field (p=50%), and random gamma (p=50%). The model was trained using the cross-entropy loss for the 8 OAR classes and background with parameter updates computed using the Adam optimizer with (0.001 learning rate, 0.9 β_1, 0.999 β_2, and AMSGrad). The model training was stopped early after 60 epochs for non-improvement of the validation loss.

Segmentation evaluation

Two experiments to evaluate the impact of defacing on the resulting segmentations were performed. In order to determine the impact of defacing on algorithmic development, models were trained on original or defaced data using the original target data for evaluation. Subsequently, in order to determine the impact of defacing on algorithms not originally developed for defaced data, a model was trained using the original data and its performance was evaluated by using the original data or the defaced data.

For both experiments, the performance of the models were quantified primarily with the Dice similarity coefficient (DSC) and the mean surface distance (MSD), defined as follows:

$$DSC = \frac{2\ \ TP}{2\ \ TP + FP + FN} \quad ,$$

$$MSD = \frac{1}{2}\left(\sum_{t \in T} \frac{d(t,\ P)}{|T|} + \sum_{p \in P} \frac{d(p,\ T)}{|P|}\right) \quad ,$$

where TP denotes true positives, FP false positives, FN false negatives, P the set of segmentation surface voxels of the model output, and T the set of segmentation surface voxels of the annotation. The distance from the surface metric is defined as:

$d(a, B)=min_{b \in B}\{\|a-b\|_2\}$. These metrics were selected because of their ubiquity in literature and ability to capture both volumetric overlap and boundary distances (38, 39). The model output was resampled into the original resolution with the nearest-neighbor sampling and evaluated against the original resolution segmentations. MSD was measured in millimeters. When comparing the performance measures between the segmentation models, Wilcoxon signed rank tests (40) were implemented, with p-values less than or equal to 0.05 considered as significant. To correct for multiple hypotheses, a Benjamini-Hochberg false discovery rate procedure (41) was implemented by taking into account all the OARs and models compared. Statistical comparisons were performed using the statannotations 0.4.4 Python package (https://github.com/trevismd/statannotations). Notably, any ROI metrics that yielded empty outputs were omitted from the comparisons. Additional surface metric values (mean Hausdorff distance at 95% and Hausdorff distance at 95%) were also calculated as part of the supplementary analysis (details in Appendix A).

Results

Defacing performance

Five of the methods tested (afni_refacer, quickshear, mri_deface, DeepDefacer, and defacer) failed for all subjects in the AAPM dataset. Therefore, for all subsequent analyses only the mask_face, fsl_deface, and pydeface methods were considered. There was scanwise quality control to remove the defaced scans with poor quality from the analyses, which resulted in 16 (29%), 10 (18%), and 13 (24%) scans removed from mask_face, fsl_deface, and pydeface, respectively, with all these methods working on 29 patient scans. A barplot comparison of the ratio of remaining OAR voxels after defacing and quality control is depicted in Figure 2. In addition, the defacing methods removed some OARs completely, which were also omitted from the segmentation evaluation. After filtering unusable data, the total number of OARs available for use in segmentation experiments was 232 for the original data and mask_face, 231 for fsl_deface, and 169 for pydeface. A full comparison of omitted OARs is shown in Table 1.

All of the tested defacing methods were unable to provide sufficient data for segmentation analysis in the HECKTOR CT dataset. Specifically, fsl_deface and pydeface methods successfully defaced 18 (8%) and 102 (46%) scans, respectively. All other methods (afni_refacer, quickshear, mri_deface, DeepDefacer, defacer, and mask_face) failed to correctly deface any of the scans. Although pydeface had the highest success rate on defacing, it only preserved the brain. Thus, no further analysis was performed for this dataset.

Segmentation performance

The 29 patient scans for which the defacing was deemed successful were used to construct and evaluate segmentation models for the mask_face, fsl_deface, and pydeface methods. The model DSC performances pooled across all structures based on training input and valid evaluation target combinations are shown in Table 2. The models trained using the original, mask_face, and fsl_deface input data had the highest composite mean DSC when evaluated on the original target data with values of 0.760, 0.742, and 0.736, respectively, while the model trained on pydeface input data had the highest composite mean DSC of 0.653 when evaluated on pydeface target data. In contrast, the models trained using original mask_face, and fsl_deface input data had the lowest composite mean DSC when evaluated on pydeface target data with values of 0.406, 0.413, 0.465, respectively, while the model trained using pydeface input data had the lowest composite mean DSC of 0.395

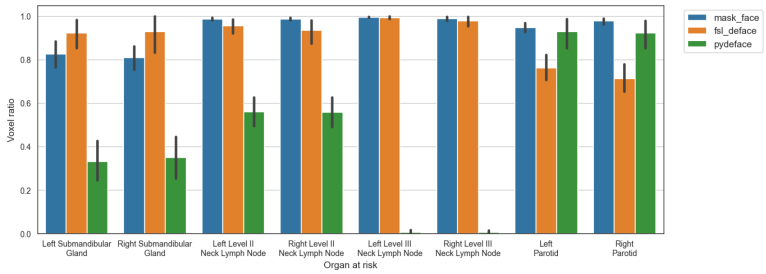


FIGURE 2
Ratio of preserved voxels in comparison to the original segmentation mask after defacing (mask_face, fsl_deface, and pydeface) for each of the organs at risk, where defacing was successful for N=39, N=42, and N=45, respectively. The mean and standard deviation are represented as the center and extremes of the error bars, respectively.

when evaluated on fsl_deface target data. All comparisons within the same evaluation data are statistically different from each other ($p \leq 0.05$) with the exception of mask_face and fsl_deface trained models evaluated on original data, and original as well as mask_face trained models evaluated on pydeface data.

Defacing impact on model training

The analysis was based on eight OAR structure segmentations from 29 patients totaling 232 evaluations. The MSD of left and right level III neck lymph nodes for pydeface trained models were omitted from the analysis as all the model outputs were empty. Full comparisons of the model performance for each OAR are depicted in Figure 3. Additional surface distance metrics are shown in Appendix A (Figure A1). Overall, the model trained with the original data performed better than the models trained with the defaced data for the majority of structures and evaluation metrics. Both metrics were significantly better for the model trained with the original data compared to the model trained with mask_face data for the left submandibular gland and right level II neck lymph node, while only the DSC was significantly better for the right submandibular gland and right level III neck lymph node. Similarly, both metrics were significantly better for the model trained with the original data compared to the model trained with fsl_deface data for the right level II neck lymph node, left parotid, and right parotid, while only the DSC was significantly better for the right level III neck lymph node. Moreover, both metrics were significantly better for the model trained with the original data compared to the model trained with pydeface data for all the structures.

Defacing impact on model testing

In these results, only valid target data with successful defacing on all three methods using non-empty segmentation structures were included. This was obtained using results from 26 left submandibular glands, 27 right submandibular glands, 1 left neck level III lymph nodes, 2 right neck level III lymph nodes, and 28 of each of the remaining structures. Due to the low number of cases for the right and left level III lymph nodes, they were omitted from the comparison. In addition, for the MSD metric, empty model output segmentations were discarded resulting in evaluation of 1 left submandibular gland for fsl_deface and mask_face and 14 for pydeface, 1 and 6 right submandibular glands on fsl_deface and

TABLE 1 Quantitative details on the number of organs at risk available after the defacing was applied for all 55 patient scans.

	Completely removed after successful defacing			Unavailable for segmentation analysis*		
Organ at risk/Defacing method	mask_face	fsl_deface	pydeface	mask_face	fsl_deface	pydeface
Left Submandibular Gland	0 (0%)	2 (4%)	6 (11%)	16 (29%)	14 (25%)	13 (24%)
Right Submandibular Gland	0 (0%)	1 (2%)	7 (13%)	16 (29%)	14 (25%)	14 (25%)
Left Neck Lymph Node Level II	0 (0%)	0 (0%)	4 (7%)	16 (29%)	13 (24%)	11 (20%)
Right Neck Lymph Node Level II	0 (0%)	0 (0%)	4 (7%)	16 (29%)	13 (24%)	11 (20%)
Left Neck Lymph Node Level III	0 (0%)	0 (0%)	45 (82%)	16 (29%)	13 (24%)	51 (93%)
Right Neck Lymph Node Level III	0 (0%)	1 (2%)	44 (80%)	16 (29%)	13 (24%)	50 (91%)
Left Parotid	0 (0%)	0 (0%)	6 (11%)	16 (29%)	13 (24%)	11 (20%)
Right Parotid	0 (0%)	0 (0%)	6 (11%)	16 (29%)	13 (24%)	11 (20%)
Total omitted	0 (0%)	4 (1%)	122 (28%)	128 (29%)	106 (24%)	172 (39%)

Only the mask_face, fsl_deface, and pydeface methods yielded usable data. The first group of columns correspond to the organs at risk that were completely removed from the cases with successful defacing. The second group of columns correspond to all items in the first group of columns plus incorporating any of the cases where defacing failed. Defacing success or failure was counted from scanwise quality control. *Organs at risk in these columns were omitted for all the subsequent segmentation-related experiments.

TABLE 2 Composite DSC performance - mean (standard deviation) - of all structures for all combinations of training data (rows) and evaluation data (columns).

	Evaluated on original (N =2 32)	Evaluated on mask_face (N = 232)	Evaluated on fsl_deface (N = 231)	Evaluated on pydeface (N = 169)
Trained on original	0.760 (0.112)	0.673 (0.181)	0.693 (0.140)	0.406 (0.304)
Trained on mask_face	0.742 (0.115)	0.733 (0.120)	0.668 (0.143)	0.413 (0.312)
Trained on fsl_deface	0.736 (0.108)	0.643 (0.185)	0.733 (0.122)	0.465 (0.293)
Trained on pydeface	0.449 (0.333)	0.417 (0.325)	0.395 (0.301)	0.653 (0.258)

The number of total segmentation maps evaluated is shown in brackets on the header. All comparisons within the same evaluation data are statistically different from each other ($p \leq 0.05$) with the exception of mask_face and fsl_deface trained models evaluated on original data, and original and mask_face trained models evaluated on pydeface data. Statistical significance was measured with Wilcoxon signed-rank tests corrected with Benjamini-Hochberg procedure comparisons within evaluation data.

pydeface, respectively, 1 left level II lymph node for pydeface, and 2 left parotids for pydeface. The model evaluated on the original data performed significantly better than the models evaluated on the defaced data for all of the structures and both evaluation metrics except in the case of left submandibular gland DSC for fsl_deface which exhibited a non-significant difference. The full comparison of the model performance for each of the OARs is shown in Figure 4. Additional surface distance metrics are shown in Appendix A (Figure A2).

Discussion

This study has systematically investigated the impact of a variety of defacing algorithms on structures of interest used for radiotherapy treatment planning. This study demonstrated that the overall usability of segmentations is heavily dependent on the choice of the defacing algorithm. Moreover, the results indicate that several OARs have the potential to be negatively impacted by the defacing algorithms, which is shown by the decreased performance of auto-segmentation algorithms trained and evaluated on defaced data in comparison to algorithms trained and evaluated on non-defaced data.

Defacing for HNC applications should be deemed optimal if the method simultaneously removes all recognizable facial features from the image and no voxels from structures of interest are affected. In this study, eight commonly available defacing algorithms developed by the neuroimaging community were applied: afni_refacer, mri_deface, defacer, DeepDefacer, mask_face, fsl_deface, pydeface, and quickshear. Unfortunately, for the investigated CT data, no defacing method was able to yield successful removal of facial features while preserving the OARs. This is not necessarily surprising given that the methods investigated were developed primarily with MRI in mind; these results echo previous similar work using CT data (42). Importantly, even when applied to MRI data of HNC patients, many of these

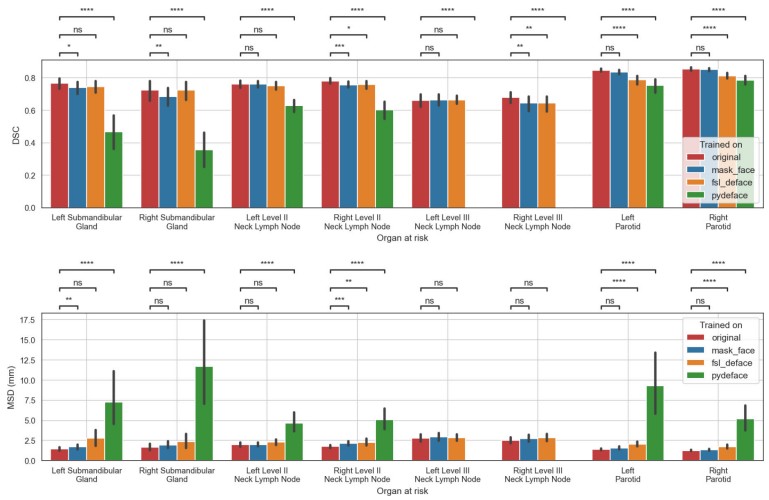


FIGURE 3
Performance of the models trained on original or defaced data and evaluated on the original data. The mean and standard deviation for each metric are represented as the center and extremes of the error bars, respectively. Statistical significance was determined using Wilcoxon signed-rank tests corrected with Benjamini-Hochberg procedure for all OARs and models. Comparison symbols: ns ($p > 0.05$), * ($p \leq 0.05$), ** ($p \leq 0.01$), *** ($p \leq$ 1e-4), **** ($p \leq$ 1e-5).

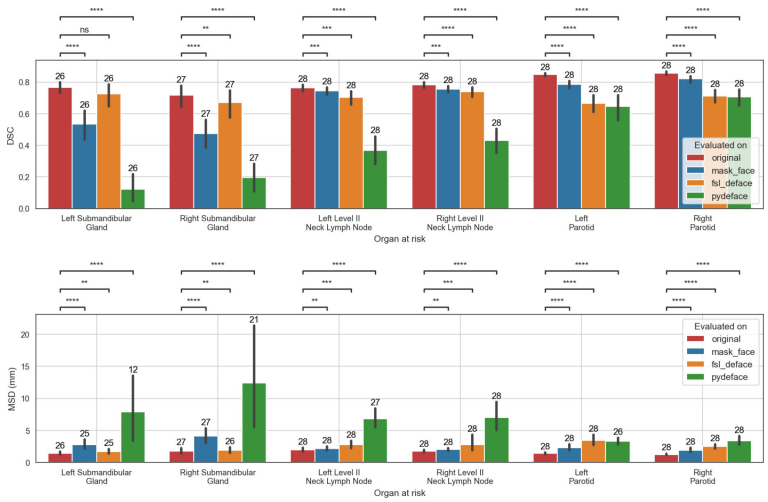


FIGURE 4
The performance of models trained on the original data when evaluated on the original, mask_face, fsl_deface, or pydeface data for the six organs at risk included in the analysis. Only cases that were available for all the methods were included: 28 segmentations were used for all structures except in the case of the left and right submandibular glands where 26 and 27 segmentations were used, respectively. In addition, for the MSD metric, empty model output segmentations were discarded, which resulted in a smaller number of evaluated structures. The number of evaluated structures is shown on top of the barplot. The mean and standard deviation for each metric are represented as the center and extremes of the error bars, respectively. Statistical significance was measured with Wilcoxon signed-rank tests corrected with Benjamini-Hochberg procedure for all OARs and models. Comparison symbols: ns ($p > 0.05$), * ($p \leq 0.05$), ** ($p \leq 0.01$), *** ($p \leq 1e\text{-}4$), **** ($p \leq 1e\text{-}5$).

defacing methods outright failed for most if not all patients. Therefore, despite extant studies demonstrating the acceptability of these methods to remove facial features from neuroimaging scans (16–21), these tools may not necessarily be robust to HNC-related imaging. Moreover, for those defacing algorithms that were able to successfully remove facial information in the MRI data, i.e. mask_face, fsl_deface, and pydeface, it was shown that regardless of the choice of the method, there was a loss of voxel-level information for all the OAR structures investigated. Importantly, pydeface leads to a greater number of lost voxels than mask_face and fsl_deface for all the OAR structures, with the exception of the parotid glands. While mask_face and fsl_deface lead to relatively minimal reduction of available voxels in many cases, the loss of topographic information in a radiotherapy workflow cannot be underscored enough. It is well known that even minor variations in the delineation of tumors and OARs can drastically alter the resulting radiotherapy dose delivered to a patient, which can impact important clinical outcomes such as toxicity and overall survival (43–46). Therefore, the loss of voxel-level information of OARs caused by the defacing algorithms, while potentially visibly imperceptible, can still affect downstream clinical workflows.

Relatively few studies have been conducted that determined the downstream analysis effects of defacing algorithms. For example, recent studies by Schwartz et al. (16) and Mikulan et al. (21) demonstrated that several defacing methods showed differences in specific neuroimaging applications, namely brain volume measurements and electroencephalography-related calculations. In this study, as a proxy for a clinically relevant task, an OAR auto-segmentation workflow was developed to investigate the impact of defacing-induced voxel-level information loss on downstream radiotherapy applications. As evident through both pooled analysis and investigation of individual OARs for auto-segmentation model training and evaluation, performance is often modestly decreased for fsl_deface and mask_face but greatly decreased for pydeface; these results were consistent with the overall voxel-level information loss. While pydeface has been shown to have favorable results for use with neuroimaging data (19, 21), its negative impact on HNC imaging is apparent. Therefore, in cases where defacing is unavoidable, mask_face or fsl_deface should likely be preferred for HNC image anonymization. Regardless, this study demonstrates existing approaches to anonymize facial data may not be sufficient for implementation on HNC-related datasets, particularly for deep learning model training and testing.

This study has several limitations. Firstly, to examine defacing methods as they are currently distributed ("out-of-the-box"), modifications to the templates or models utilized in any methods were not performed. Further preprocessing either of the CT and MRI data as well as subject specific settings could have helped some of the methods to better identify the face. In addition, more suitable templates for the HNC images (for both CT and MRI) would likely improve the defacing performance; for the registration-based methods, algorithms likely expected scans to cover the whole brain, while the field-of-view of the images for HNC mostly covered the neck and mouth, leaving the top of the brain excluded. Notably, additional deep learning model training schemes (i.e., transfer learning) may potentially allow for eventual implementation of existing deep learning methods on domain-specific datasets (i.e., HNC radiotherapy), but this negates the immediate interoperability of these tools. Furthermore, no additional image processing other than what was integrated into the defacing methods was implemented; it may be possible

alternative processing could change these results. Secondly, while a robust analysis utilizing multiple relevant metrics established in existing literature (38) was performed to evaluate OAR auto-segmentation, there is not always a perfect correlation between spatial similarity metrics and radiotherapy plan acceptability (39). This study has not tested the downstream effects of defacing on radiotherapy plan generation, which may lead to different results from what was observed for the OAR segmentation. Thirdly, this study was limited to public data with no modifications. Only structures that were already available in existing datasets were analyzed. Moreover, as an initial exploration of defacing methods for radiotherapy applications, only a single imaging modality on a relatively limited sample size, namely T2-weighted MRI, was investigated for auto-segmentation experiments, despite the HNC radiotherapy workflow commonly incorporating additional modalities (47). Thus, experiments on additional imaging modalities and larger diverse HNC patient populations should be the subject of future investigations. Fourthly, the current analysis does not thoroughly explore possible performance confounding related to phenotypical and individual variables such as sex, ethnicity, and age of the measured individuals. Finally, this study has focused on defacing methods as an avenue for public data sharing for training and evaluating machine learning models, but privacy-preserving modeling approaches, e.g., through federated learning (48), may also act as a potential alternative solution.

Conclusion

In summary, by using publicly available data, the effects of eight established defacing algorithms, afni_refacer, mask_face, mri_deface, defacer, DeepDefacer, quickshear, fsl_deface, and pydeface, have been systematically investigated for radiotherapy applications. Specifically, the impact of defacing directly on ground-truth HNC OARs was determined and a deep learning based OAR auto-segmentation workflow to investigate the use of defaced data for algorithmic training and evaluation was developed. All methods failed to properly remove facial features on the CT dataset investigated. Moreover, it was observed that only fsl_deface, mask_face, and pydeface yielded usable images from the MRI dataset, but still decreased the total number of voxels in OARs and negatively impacted the performance of OAR auto-segmentation, with pydeface having more severe negative effects than mask_face or fsl_deface. This study is an important step towards ensuring widespread privacy-preserving dissemination of HNC imaging data without endangering data usability. Given that current defacing methods remove critical data, future larger studies should investigate alternative approaches for anonymizing facial data that preserve radiotherapy-related structures. Moreover, studies on the impact of these methods on radiotherapy plan generation, the inclusion of a greater number of OARs and target structures, and the incorporation of additional imaging modalities are also warranted.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://wiki.cancerimagingarchive.net/display/Public/AAPM+RT-MAC+Grand+Challenge+2019.

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

Study concepts: all authors; Study design: JS, EG, and JJ; Data acquisition: KW, MN, and RH; Quality control of data and algorithms: JS and EG; Data analysis and interpretation: JS, KW, EG, JJ, BK, AM, and KK; Manuscript editing: JS, KW, EG, JJ, BK, AM, KK, and CF. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the National Institutes of Health (NIH)/National Cancer Institute (NCI) through a Cancer Center Support Grant (CCSG; P30CA016672-44). MN is supported by an NIH grant (R01DE028290-01). KW is supported by a training fellowship from The University of Texas Health Science Center at Houston Center for Clinical and Translational Sciences TL1 Program (TL1TR003169), the American Legion Auxiliary Fellowship in Cancer Research, and an NIH/National Institute for Dental and Craniofacial Research (NIDCR) F31 fellowship (1 F31DE031502-01). CF received funding from the NIH/NIDCR (1R01DE025248-01/R56DE025248); an NIH/NIDCR Academic-Industrial Partnership Award (R01DE028290); the National Science Foundation (NSF), Division of Mathematical Sciences, Joint NIH/NSF Initiative on Quantitative Approaches to Biomedical Big Data (QuBBD) Grant (NSF 1557679); the NIH Big Data to Knowledge (BD2K) Program of the NCI Early Stage Development of Technologies in Biomedical Computing, Informatics, and Big Data Science Award (1R01CA214825); the NCI Early Phase Clinical Trials in Imaging and Image-Guided Interventions Program (1R01CA218148); an NIH/NCI Pilot Research Program Award from the UT MD Anderson CCSG Radiation Oncology and Cancer Imaging Program (P30CA016672); an NIH/NCI Head and Neck Specialized Programs of Research Excellence (SPORE) Developmental Research Program Award (P50CA097007); and the National Institute of Biomedical Imaging and Bioengineering (NIBIB) Research Education Program (R25EB025787).

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.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2023.1120392/full#supplementary-material

References

1. Foster ED, Deardorff A. Open science framework (OSF). *J Med Libr Assoc JMLA* (2017) 105:203. doi: 10.5195/jmla.2017.88

2. Wilkinson MD, Dumontier M, IjJ A, Appleton G, Axton M, Baak A, et al. The FAIR guiding principles for scientific data management and stewardship. *Sci Data* (2016) 3:1–9. doi: 10.1038/sdata.2016.18

3. Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P, et al. The cancer imaging archive (TCIA): Maintaining and operating a public information repository. *J Digit Imaging* (2013) 26:1045–57. doi: 10.1007/s10278-013-9622-7

4. Wahid KA, Glerean E, Sahlsten J, Jaskari J, Kaski K, Naser MA, et al. Artificial intelligence for radiation oncology applications using public datasets. In: *Seminars in radiation oncology*. Elsevier (2022). p. 400–14.

5. Press RH, Shu H-KG, Shim H, Mountz JM, Kurland BF, Wahl RL, et al. The use of quantitative imaging in radiation oncology: A quantitative imaging network (QIN) perspective. *Int J Radiat Oncol Biol Phys* (2018) 102:1219–35. doi: 10.1016/j.ijrobp.2018.06.023

6. Beaton L, Bandula S, Gaze MN, Sharma RA. How rapid advances in imaging are defining the future of precision radiation oncology. *Br J Cancer* (2019) 120:779–90. doi: 10.1038/s41416-019-0412-y

7. Andrearczyk V, Oreiller V, Jreige M, Vallières M, Castelli J, Elhalawani H, et al. Overview of the HECKTOR challenge at MICCAI 2020: automatic head and neck tumor segmentation in PET/CT. In: *3D head and neck tumor segmentation in PET/CT challenge*. Springer (2020). p. 1–21.

8. Andrearczyk V, Oreiller V, Boughdad S, Rest CCL, Elhalawani H, Jreige M, et al. Overview of the HECKTOR challenge at MICCAI 2021: Automatic head and neck tumor segmentation and outcome prediction in PET/CT images. In: *3D head and neck tumor segmentation in PET/CT challenge*. (Strasbourg, France: Springer) (2021). p. 1–37.

9. Oreiller V, Andrearczyk V, Jreige M, Boughdad S, Elhalawani H, Castelli J, et al. Head and neck tumor segmentation in PET/CT: The HECKTOR challenge. *Med Image Anal* (2022) 77:102336. doi: 10.1016/j.media.2021.102336

10. Meystre SM, Friedlin FJ, South BR, Shen S, Samore MH. Automatic de-identification of textual documents in the electronic health record: A review of recent research. *BMC Med Res Methodol* (2010) 10:1–16. doi: 10.1186/1471-2288-10-70

11. Prior FW, Brunsden B, Hildebolt C, Nolan TS, Pringle M, Vaishnavi SN, et al. Facial recognition from volume-rendered magnetic resonance imaging data. *IEEE Trans Inf Technol BioMed* (2008) 13:5–9. doi: 10.1109/TITB.2008.2003335

12. Mazura JC, Juluru K, Chen JJ, Morgan TA, John M, Siegel EL. Facial recognition software success rates for the identification of 3D surface reconstructed facial images: Implications for patient privacy and security. *J Digit Imaging* (2012) 25:347–51. doi: 10.1007/s10278-011-9429-3

13. Schwarz CG, Kremers WK, Therneau TM, Sharp RR, Gunter JL, Vemuri P, et al. Identification of anonymous MRI research participants with face-recognition software. *N Engl J Med* (2019) 381:1684–6. doi: 10.1056/NEJMc1908881

14. Parks CL, Monson KL. Automated facial recognition of computed tomography-derived facial images: patient privacy implications. *J Digit Imaging* (2017) 30:204–14. doi: 10.1007/s10278-016-9932-7

15. Delbarre DJ, Santos L, Ganjgahi H, Horner N, McCoy A, Westerberg H, et al. Application of a convolutional neural network to the quality control of MRI defacing. *Comput Biol Med* (2022) 151:106211. doi: 10.1016/j.compbiomed.2022.106211

16. Schwarz CG, Kremers WK, Wiste HJ, Gunter JL, Vemuri P, Spychalla AJ, et al. Changing the face of neuroimaging research: Comparing a new MRI de-facing technique with popular alternatives. *NeuroImage* (2021) 231:117845. doi: 10.1016/j.neuroimage.2021.117845

17. Schimke N, Kuehler M, Hale J. Preserving privacy in structural neuroimages. In: *IFIP annual conference on data and applications security and privacy*. (Richmond, VA, USA: Springer) (2011). p. 301–8.

18. Bischoff-Grethe A, Ozyurt IB, Busa E, Quinn BT, Fennema-Notestine C, Clark CP, et al. A technique for the deidentification of structural brain MR images. *Hum Brain Mapp* (2007) 28:892–903. doi: 10.1002/hbm.20312

19. Theyers AE, Zamyadi M, O'Reilly M, Bartha R, Symons S, MacQueen GM, et al. Multisite comparison of MRI defacing software across multiple cohorts. *Front Psychiatry* (2021) 12:189. doi: 10.3389/fpsyt.2021.617997

20. De Sitter A, Visser M, Brouwer I, Cover K, van Schijndel R, Eijgelaar R, et al. Facing privacy in neuroimaging: removing facial features degrades performance of image analysis methods. *Eur Radiol* (2020) 30:1062–74. doi: 10.1007/s00330-019-06459-3

21. Mikulan E, Russo S, Zauli FM, d'Orio P, Parmigiani S, Favaro J, et al. A comparative study between state-of-the-art MRI deidentification and AnonyMI, a new method combining re-identification risk reduction and geometrical preservation. (Hoboken, USA: John Wiley & Sons, Inc.) (2021) 42(17):5523–34. doi: 10.1101/2021.07.30.454335

22. Cardenas CE, Mohamed ASR, Yang J, Gooding M, Veeraraghavan H, Kalpathy-Cramer J, et al. Head and neck cancer patient images for determining auto-segmentation accuracy in T2-weighted magnetic resonance imaging through expert manual segmentations. *Med Phys* (2020) 47:2317–22. doi: 10.1002/mp.13942

23. Alfaro-Almagro F, Jenkinson M, Bangerter NK, Andersson JL, Griffanti L, Douaud G, et al. Image processing and quality control for the first 10,000 brain imaging datasets from UK biobank. *Neuroimage* (2018) 166:400–24. doi: 10.1016/j.neuroimage.2017.10.034

24. Milchenko M, Marcus D. Obscuring surface anatomy in volumetric imaging data. *Neuroinformatics* (2013) 11:65–75. doi: 10.1007/s12021-012-9160-3

25. Gulban O, Nielson D, Poldrack R, Gorgolewski C. *Poldracklab/pydeface: v2.0.0* . Available at: https://github.com/poldracklab/pydeface.

26. Jeong YU, Yoo S, Kim Y-H, Shim WH. De-identification of facial features in magnetic resonance images: Software development using deep learning technology. *J Med Internet Res* (2020) 22:e22739. doi: 10.2196/22739

27. Khazane A, Hoachuck J, Gorgolewski KJ, Poldrack RA. DeepDefacer: Automatic removal of facial features *via* U-net image segmentation. *arXiv* (2022) arXiv:2205.15536. doi: 10.48550/arXiv.2205.15536

28. Wahid KA, Ahmed S, He R, van Dijk LV, Teuwen J, McDonald BA, et al. Evaluation of deep learning-based multiparametric MRI oropharyngeal primary tumor auto-segmentation and investigation of input channel effects: Results from a prospective imaging registry. *Clin Transl Radiat Oncol* (2022) 32:6–14. doi: 10.1016/j.ctro.2021.10.003

29. McDonald BA, Cardenas C, O'Connell N, Ahmed S, Naser MA, Wahid KA, et al. Investigation of autosegmentation techniques on T2-weighted MRI for off-line dose reconstruction in MR-linac adapt to position workflow for head and neck cancers. *medRxiv* (2021). doi: 10.1101/2021.09.30.21264327

30. Taku N, Wahid KA, van Dijk LV, Sahlsten J, Jaskari J, Kaski K, et al. Auto-detection and segmentation of involved lymph nodes in HPV-associated oropharyngeal cancer using a convolutional deep learning neural network. *Clin Transl Radiat Oncol* (2022) 36:47–55. doi: 10.1016/j.ctro.2022.06.007

31. Naser MA, van Dijk LV, He R, Wahid KA, Fuller CD. *Tumor segmentation in patients with head and neck cancers using deep learning based-on multi-modality PET/CT images*. Springer (2020) p. 85–98.

32. Naser MA, van Dijk LV, He R, Wahid KA, Fuller CD. Tumor segmentation in patients with head and neck cancers using deep learning based-on multi-modality PET/CT images. *In Head and Neck Tumor Segmentation: First Challenge, HECKTOR 2020, Held in Conjunction with MICCAI 2020, Proceedings 1 2021*; Lima, Peru: Springer International Publishing (2020) p. 85–98.

33. Naser MA, Wahid KA, Grossberg AJ, Olson B, Jain R, El-Habashy D, et al. Deep learning auto-segmentation of cervical skeletal muscle for sarcopenia analysis in

patients with head and neck cancer. *Front Oncol* (2022) 12. doi: 10.3389/fonc.2022.930432

34. Van Rossum G, Drake FLJr. Python Reference manual. *Centrum voor Wiskunde en Informatica Amsterdam* (1995).

35. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, et al. Pytorch: An imperative style, high-performance deep learning library. *Adv Neural Inf Process Syst* (2019) 32:8026–37.

36. *The MONAI consortium*. Project MONAI (2020). doi: 10.5281/zenodo.4323059

37. Pérez-García F, Sparks R, Ourselin S. TorchIO: a Python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning. *Comput Methods Programs BioMed* (2021) 208:106236. doi: 10.1016/j.cmpb.2021.106236

38. Taha AA, Hanbury A. Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool. *BMC Med Imaging* (2015) 15:1–28. doi: 10.1186/s12880-015-0068-x

39. Sherer MV, Lin D, Elguindi S, Duke S, Tan L-T, Cacicedo J, et al. Metrics to evaluate the performance of auto-segmentation for radiation treatment planning: A critical review. *Radiother Oncol* (2021) 160:185–91. doi: 10.1016/j.radonc.2021.05.003

40. Wilcoxon F. Individual comparisons by ranking methods. In: *Breakthroughs in statistics*. (New York, USA: Springer) (1992). p. 196–202.

41. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B Methodol* (1995) 57:289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x

42. Muschelli J. Recommendations for processing head CT data. *Front Neuroinformatics* (2019) 13:61. doi: 10.3389/fninf.2019.00061

43. Lin D, Lapen K, Sherer MV, Kantor J, Zhang Z, Boyce LM, et al. A systematic review of contouring guidelines in radiation oncology: Analysis of frequency, methodology, and delivery of consensus recommendations. *Int J Radiat Oncol Biol Phys* (2020) 107:827–35. doi: 10.1016/j.ijrobp.2020.04.011

44. Abrams RA, Winter KA, Regine WF, Safran H, Hoffman JP, Lustig R, et al. Failure to adhere to protocol specified radiation therapy guidelines was associated with decreased survival in RTOG 9704–a phase III trial of adjuvant chemotherapy and chemoradiotherapy for patients with resected adenocarcinoma of the pancreas. *Int J Radiat Oncol Biol Phys* (2012) 82:809–16. doi: 10.1016/j.ijrobp.2010.11.039

45. Peters LJ, O'Sullivan B, Giralt J, Fitzgerald TJ, Trotti A, Bernier J, et al. Critical impact of radiotherapy protocol compliance and quality in the treatment of advanced head and neck cancer: Results from TROG 02.02. *J Clin Oncol* (2010) 28:2996–3001. doi: 10.1200/JCO.2009.27.4498

46. Ohri N, Shen X, Dicker AP, Doyle LA, Harrison AS, Showalter TN. Radiotherapy protocol deviations and clinical outcomes: a meta-analysis of cooperative group clinical trials. *J Natl Cancer Inst* (2013) 105:387–93. doi: 10.1093/jnci/djt001

47. Salzillo TC, Taku N, Wahid KA, McDonald BA, Wang J, van Dijk LV, et al. Advances in imaging for HPV-related oropharyngeal cancer: Applications to radiation oncology. In: *Seminars in radiation oncology*. Elsevier (2021). p. 371–88.

48. Kaissis G, Ziller A, Passerat-Palmbach J, Ryffel T, Usynin D, Trask A, et al. End-to-end privacy preserving deep learning on multi-institutional medical imaging. *Nat Mach Intell* (2021) 3:473–84. doi: 10.1038/s42256-021-00337-8

frontiers | Frontiers in Oncology

TYPE Methods
PUBLISHED 02 March 2023
DOI 10.3389/fonc.2023.1112481

OPEN ACCESS

EDITED BY
Bahram Mohajer,
Johns Hopkins Medicine, United States

REVIEWED BY
Christian Bäumer,
West German Proton Therapy Centre Essen (WPE), Germany
Hong Qi Tan,
National Cancer Centre Singapore, Singapore

*CORRESPONDENCE
Saber Nankali
saber.nankali@rm.dk

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

RECEIVED 30 November 2022
ACCEPTED 13 February 2023
PUBLISHED 02 March 2023

CITATION
Nankali S, Worm ES, Thomsen JB, Stick LB, Bertholet J, Høyer M, Weber B, Mortensen HR and Poulsen PR (2023) Intrafraction tumor motion monitoring and dose reconstruction for liver pencil beam scanning proton therapy.
Front. Oncol. 13:1112481.
doi: 10.3389/fonc.2023.1112481

Intrafraction tumor motion monitoring and dose reconstruction for liver pencil beam scanning proton therapy

Saber Nankali[1,2*], Esben Schjødt Worm[3], Jakob Borup Thomsen[1], Line Bjerregaard Stick[1], Jenny Bertholet[4], Morten Høyer[1], Britta Weber[1,3], Hanna Rahbek Mortensen[1] and Per Rugaard Poulsen[1,3]

[1]Danish Centre for Particle Therapy, Aarhus University Hospital, Aarhus, Denmark, [2]Department of Clinical Medicine, Aarhus University, Aarhus, Denmark, [3]Department of Oncology, Aarhus University Hospital, Aarhus, Denmark, [4]Division of Medical Radiation Physics and Department of Radiation Oncology, Inselspital, Bern University Hospital, and University of Bern, Bern, Switzerland

Background: Pencil beam scanning (PBS) proton therapy can provide highly conformal target dose distributions and healthy tissue sparing. However, proton therapy of hepatocellular carcinoma (HCC) is prone to dosimetrical uncertainties induced by respiratory motion. This study aims to develop intra-treatment tumor motion monitoring during respiratory gated proton therapy and combine it with motion-including dose reconstruction to estimate the delivered tumor doses for individual HCC treatment fractions.

Methods: Three HCC-patients were planned to receive 58 GyRBE (n=2) or 67.5 GyRBE (n=1) of exhale respiratory gated PBS proton therapy in 15 fractions. The treatment planning was based on the exhale phase of a 4-dimensional CT scan. Daily setup was based on cone-beam CT (CBCT) imaging of three implanted fiducial markers. An external marker block (RPM) on the patient's abdomen was used for exhale gating in free breathing. This study was based on 5 fractions (patient 1), 1 fraction (patient 2) and 6 fractions (patient 3) where a post-treatment control CBCT was available. After treatment, segmented 2D marker positions in the post-treatment CBCT projections provided the estimated 3D motion trajectory during the CBCT by a probability-based method. An external-internal correlation model (ECM) that estimated the tumor motion from the RPM motion was built from the synchronized RPM signal and marker motion in the CBCT. The ECM was then used to estimate intra-treatment tumor motion. Finally, the motion-including CTV dose was estimated using a dose reconstruction method that emulates tumor motion in beam's eye view as lateral spot shifts and in-depth motion as changes in the proton beam energy. The CTV homogeneity index (HI) The CTV homogeneity index (HI) was calculated as $\frac{D2\% - D98\%}{D50\%} \times 100\%$.

Results: The tumor position during spot delivery had a root-mean-square error of 1.3 mm in left-right, 2.8 mm in cranio-caudal and 1.7 mm in anterior-posterior directions compared to the planned position. On average, the CTV HI was larger

than planned by 3.7%-points (range: 1.0-6.6%-points) for individual fractions and by 0.7%-points (range: 0.3-1.1%-points) for the average dose of 5 or 6 fractions.

Conclusions: A method to estimate internal tumor motion and reconstruct the motion-including fraction dose for PBS proton therapy of HCC was developed and demonstrated successfully clinically.

KEYWORDS

proton therapy, pencil beam scanning, dose reconstruction, liver cancer, motion management, respiratory gating, tumor motion monitoring (Min.5-Max. 8)

1 Introduction

Radiation therapy is a local treatment option for small hepatocellular carcinoma (HCC) tumors in inoperable patients with a good liver function (Child-Pugh A) (1). However, HCC patients often present with a considerable tumor burden and an underlying cirrhotic liver and even low doses of radiation to the liver leads to a high risk of developing radiation induced liver disease (RILD) in patients with cirrhosis (2). Since RILD is a severe condition that can lead to liver failure and death, it is crucial to minimize the dose to the normal liver tissue surrounding the tumor (3).

Compared to photon based radiation therapy, pencil beam scanning (PBS) proton therapy can often provide more conformal target dose distributions with less healthy tissue irradiation (4, 5). Proton therapy is therefore increasingly used in the treatment of HCC (6, 7). However, liver tumors often exhibit large and variable respiratory motion during treatment (8), which can cause considerable deviations between the delivered and planned doses. Due to interplay effects and a high sensitivity to water equivalent path length changes, PBS proton therapy is particularly prone to dosimetric uncertainties caused by target motion (9–11) and international guidelines underline the special need for motion management in PBS proton therapy (12, 13). Hence, respiratory gating, where the beam is only turned on during specific phases of the breathing cycle has been proposed and implemented in proton therapy to mitigate tumor motion effects (14–17). Still, residual motion within the gating window is of concern (16, 18).

Reconstruction of the actual delivered tumor dose at a fraction requires knowledge of the internal motion during treatment delivery and synchronization of this motion with the beam delivery. One method is to calculate the dynamic 4D dose (D4DD) by ascribing a specific phase of a 4DCT scan to each delivered spot, use this to calculate phase-specific doses in each 4DCT phase and accumulate these doses in a reference phase by deformable image registration (12, 13). This method has been implemented clinically for PBS carbon therapy (19) and proton therapy (20, 21) using a waist belt for respiratory monitoring during beam delivery. A limitation of the D4DD is that it neglects setup errors and assumes that the internal motion during treatment is well described by the respiration signal and identical to the motion in the 4DCT. However, liver motion is known to be highly variable and often poorly represented by 4DCT scans that by nature only capture one (random) respiratory cycle at each longitudinal position within the patient (22, 23). To overcome these limitations, Yamada et al. monitored the internal motion of implanted fiducial markers in the liver during gated proton PBS delivery by a gantry-mounted stereoscopic fluoroscopic x-ray imager (24). By combining the internal motion with the spot delivery timing in machine log files the authors reconstructed the tumor dose by a spot shift method that can account for arbitrary rigid motion (25). However, although many modern conventional proton facilities are equipped with dual x-ray imagers, these can typically not be used during treatment delivery. Consequently, target motion monitoring during treatment is normally not available even though it is recommended by international guidelines for proton therapy of moving targets (12, 13).

In this study, we introduce a method to overcome the limitations of conventional proton facilities in internal tumor motion monitoring during proton PBS treatment. The method uses an external-internal correlation model (ECM) to estimate the internal tumor motion from an external respiratory signal and combines the internal motion with spot delivery timing in machine log files to estimate the tumor position during each spot delivery. The motion is then combined with the spot shift dose reconstruction method to estimate the tumor doses for individual HCC treatment fractions.

2 Material and methods

2.1 Patients and treatment planning

Three patients with HCC underwent proton PBS in April-September 2022 in a national phase II clinical trial that allowed inclusion of both large tumors and Child-Pugh B patients. The trial was approved by the relevant ethics committees (ClinicalTrials.gov Identifier: NCT05203120). Three gold or platinum fiducial markers with dimensions of 0.75 mm × 5 mm (Visicoil^{TM}) were implanted near the tumor the day before planning CT scanning. An internal

clinical target volume (iCTV) was formed as the union of the CTV in the five phases of a 10-phase 4-dimensional CT scan (4DCT) that were closest to full exhale. It corresponded to exhale respiratory gating with approximately 50% duty cycle. A 3-field proton plan was created on the exhale phase of the 4DCT using a commercial treatment planning system (TPS, Eclipse 16.01.10, Varian, a Siemens Healthineers Company, Palo Alto, CA) and dose calculation algorithm (Varian Proton Convolution Superposition 16.1.0). Robust single field uniform dose (SFUD) optimization was performed with ±4.5% range uncertainty and ±5 mm shifts in the left-right (LR) and anterior-posterior (AP) directions and ±7 mm shifts in the cranio-caudal (CC) direction. The treatment plans used beam energies in the range 71-153 MeV, for which the spot size in air is 4-6mm (1 standard deviation). Each field had 528-2134 spots and a total of 6617-16047 monitor units. The prescribed mean iCTV dose was 58 GyRBE (Patient 1 and 3) for central tumors (≤2 cm from porta hepatis) or 67.5 GyRBE (Patient 2) for peripheral tumors (>2 cm from porta hepatis) in 15 fractions.

2.2 Treatment delivery and imaging

Daily patient setup was based on a free-breathing CBCT scan in which the estimated exhale positions of the motion-blurred fiducial markers were matched with the planning CT. CBCT scan was done using Standard ProBeam CBCT imaging system with Paxscan 4030CB flat panel detectors. The resolution of the image detector was 0.388 mm/pixel in both directions with source-to-imager distance (SID) of 3700 mm and source-to-axis distance (SAD) of 2700 mm. During the CBCT acquisition and throughout the whole treatment session the position of a marker block (Real-time Position Management System, RPM, Varian) on the patient's abdomen was recorded with a camera. During treatment the RPM signal was used for respiratory gating with a gating window adjusted before treatment to obtain a duty cycle of approximately 50% centered around the exhale phase in accordance with the iCTV construction. A post-treatment control CBCT scan was captured at 6, 1, and 7 fractions for patients 1, 2, and 3, respectively. The RPM log file was missing for one of these treatment fractions for patient 1 and patient 3. The analysis presented in this study requires a post-treatment CBCT and an RPM log file and was therefore only made for 5, 1 and 6 fractions for patients 1, 2, and 3, respectively.

2.3 Data analysis

After the treatments the fiducial markers were segmented in each raw 2D CBCT-projection (~1000 images per CBCT) using an automatic method (26) followed by manual inspection and semi-automatic correction of failed segmentations. The 3D motion trajectory of each marker during CBCT was estimated by a probability-based method (27) and the marker group centroid motion was used as a surrogate for the tumor motion. The exhale period was defined as the time within the 95^{th}-100^{th} percentile of the markers position in the CC direction for each CBCT. This was used to determine the exhale position in each direction of motion as the mean marker position during the exhale period. For the setup CBCT scans, the resulting exhale tumor position was used to determine the optimal setup couch correction for marker alignment with the planned marker positions in exhale. This is similar to the trajectory-based setup introduced for non-gated treatments in (28). The online registration error was then calculated as the difference between the retrospective trajectory-based patient setup and the actual couch correction based on online 3D/3D registration of the setup CBCT with the planning CT scan. Furthermore, the intrafraction baseline drift of the exhale position between the setup CBCT and the post-treatment CBCT was determined as the difference between their respective exhale positions.

The analysis in this study required synchronization of the RPM signal with the projection images of the post-treatment CBCT (to establish an ECM) and with the delivery time of each proton spot (to perform dose reconstruction). Synchronization between RPM and CBCT projections was obtained by placing a 3 mm diameter tungsten sphere on the RPM block such that it was visible in most of the CBCT projections (Figures 1A, B). After treatment the 3D motion of the tungsten sphere during the CBCT scan was estimated from its projected motion in the CBCT projections (27), and its AP motion was temporally aligned with the RPM motion in gating log files (Figure 1C). This synchronization provided the logged RPM position at the acquisition time of each CBCT projection.

The synchronization between RPM and spot delivery times was based on the gate-on signal in the gating log files. The logged gate-on signal specifies the time intervals when the RPM block is inside the gating window, but it does not account for the gate-on latency between entering the gating window and beam-on and the gate-off latency between exiting the gating window and beam-off. The gate-open times during which the beam could potentially be turned on were estimated from the logged gate-on signal by assuming a gate-on latency of 240 ms and a gate-off latency of 80 ms (Figure 1D). These latencies were measured using the method proposed by Worm et al. (29) and rounded to an integer number of gating log file samples (40ms resolution). Next, a comparison of the gate-open times with the actual spot delivery times in machine log files (30) provided the synchronization between RPM log files and spot delivery times (Figure 1E). While the machine log files specified the duration of each spot with microsecond resolution it did not directly specify the beam-off times occurring during larger spots shifts, energy shifts and gate-off periods. However, by using the logged number of User Datagram Protocol (UDP) messages received between each spot the beam-off times were estimated with a scaling factor uncertainty of a few percent. During the synchronization of the machine log files with the gating log files the beam-off times were scaled to fit the time scale in the gating log file.

To estimate the tumor motion during treatment delivery an augmented linear ECM that estimated the tumor motion during the post-treatment CBCT from the synchronized RPM motion (31) was built (Label 1 in Figure 2):

$$\mathrm{INT}(t) = \mathrm{A.EXT}(t) + \mathrm{B.EXT}(t\text{-}\tau) + \mathrm{C} \quad (1)$$

Here, INT and EXT are internal tumor and external RPM motion as a function of time (t). The coefficients A, B and C and the

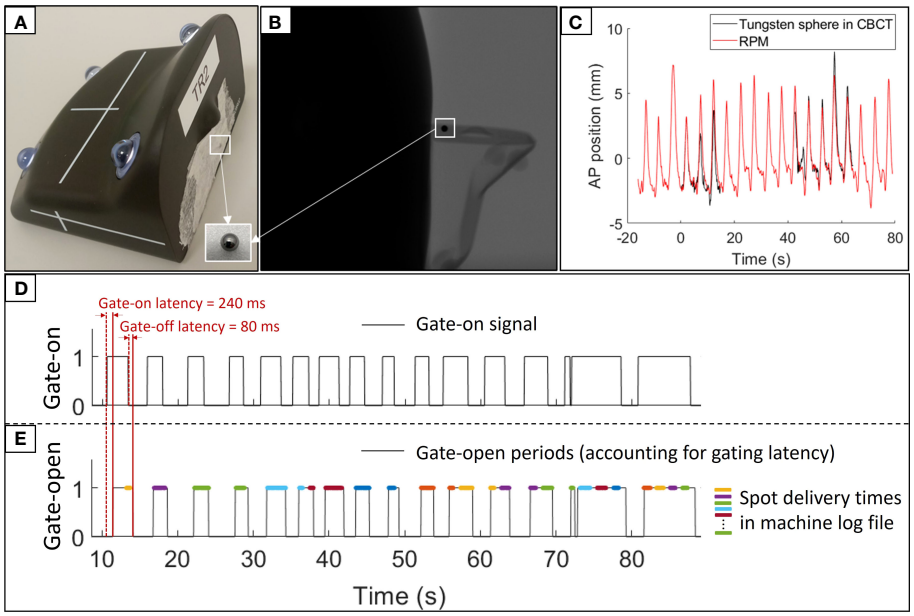


FIGURE 1
Synchronization of RPM log files. **(A)** Marker block with the tungsten sphere used for synchronization with the CBCT projections. **(B)** A CBCT projection showing the marker block and tungsten sphere. **(C)** Synchronized RPM signal (red) and anterior-posterior (AP) tungsten sphere trajectory extracted from the CBCT projections (black). **(D)** Gate-on signal in RPM log file used for synchronization with spot delivery times. **(E)** Gate-open times accounting for the gating latency (black) and synchronized spot delivery times from machine log files (colored lines, with different colors indicating different energy layers).

time delay τ are fitting parameters. The augmentation term B.EXT (t-τ) accounts for hysteresis and phase differences between internal and external motion (31). A, B and C were optimized individually for each motion direction with least-square fitting while the same value of τ was used for all three motion directions.

Next, The ECM was used to estimate the tumor motion throughout the treatment delivery from the intra-treatment RPM signal (Label 2). This synchronization resulted in the ECM estimated tumor position at the time of each spot delivery (Label 3). For each spot, the geometrical treatment accuracy was determined as the tumor position relative to the planned position. The tumor motion range was calculated for each fraction as the difference between the 98th and the 2nd percentiles of the tumor position over all three fields during beam-on periods and, for comparison, over the full field durations including both beam-on and beam-off periods.

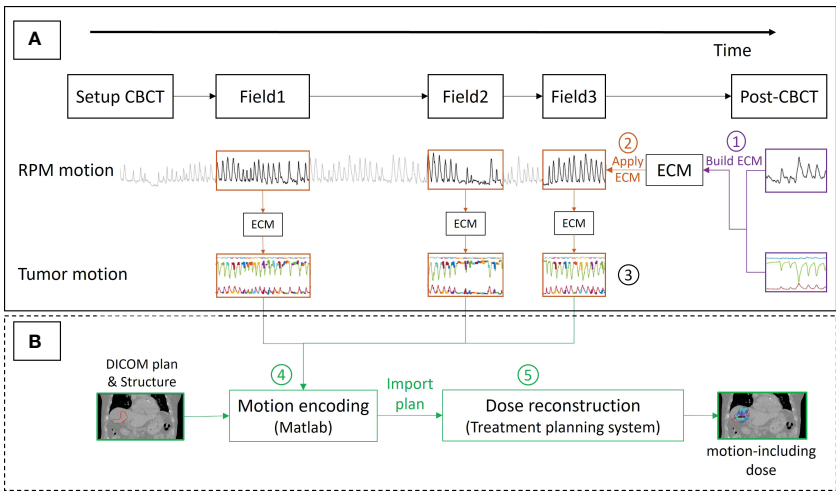


FIGURE 2
Workflow for **(A)** estimating the tumor position at the time of delivery of each spot by an external-internal correlation model (ECM) and **(B)** motion-including dose reconstruction. The numbers refer to the description in the text. The thick colored curves shown on the top of the tumor motion (Label 3) show the spot delivery times with different energy layers indicated with different colors.

Finally, the CTV dose and the dose to the healthy liver tissue of each analyzed fraction was estimated by a motion-including dose reconstruction method that emulates tumor motion in beam's eye view by shifting each spot in the opposite direction of the tumor motion and in-depth tumor motion as changes in the proton beam energy (25). A motion-encoded plan with these manipulations of the original spot positions and energies was created by an in-house developed Matlab program (Label 4) and imported into the TPS for recalculation (Label 5). The reconstructed CTV doses of each fraction and averaged over all analyzed fractions were compared with the planned doses using the metrics of CTV D98% and D2% (minimum dose received by 98% and 2% of the CTV volume) and the homogeneity index HI% = $\frac{D2\% - D98\%}{D50\%} \times 100\%$. Furthermore, the mean healthy liver tissue doses averaged over all analyzed fractions were compared with the planned doses.

3 Results

Figure 3 shows an example of the tumor motion trajectory during the setup CBCT and the post-treatment CBCT at one fraction for patient 3. At this fraction, the online registration errors were 0.4-0.8 mm (Figure 3). The mean online registration error was in general sub-millimeter for all patients (Table 1). The example case in Figure 3 had a small cranial and posterior drift of the tumor exhale position from setup CBCT to post-treatment CBCT (black arrows). Averaged over all fractions and patients a similar trend was seen with mean ± standard deviation (SD) drift motion of 0.0 mm ± 0.8 mm (LR), 1.3 mm ± 1.3 mm (CC), and -0.7 mm ± 1.0 mm (AP) (Table 1).

The example ECM presented in Figure 3 (blue curves) had an accuracy close to the mean RMS fit error for patient 3, while patients 1 and 2 had larger RMS fit errors up to 2.1 mm (Table 1). Over all patients, the mean (± SD) RMS fit error of the ECM was 0.5 mm ± 0.4 mm (LR), 1.5 mm ± 0.8 mm (CC), and 1.0 mm ± 0.6 mm (AP).

Figure 4 presents a typical example of the RPM signal and the ECM estimated tumor motion at a fraction, synchronized with the spot delivery times. Due to the gating latency the beam started 240 ms into the gating window and continued 80 ms after the RPM block moved out of the gating window (Figure 4C). The treatment error is reported in Table 1 for each patient. Over all delivered spots the RMS treatment error was 1.3 mm (LR), 2.8 mm (CC), and 1.7 mm (AP), while the mean (± SD) 3D treatment error per patient was 3.9 mm ± 1.9 mm (patient 1), 3.7 mm ± 0.6 mm (patient 2) and 2.6 mm ± 1.7 mm (patient 3).

The maximum tumor motion range during a fraction was 6.4 mm (LR), 27.9 mm (CC), and 19.2 mm (AP) during field delivery independent of beam-on status and 2.7 mm (LR), 10 mm (CC), and 7.1 mm (AP) during beam-on periods. The mean tumor motion range during a single fraction was usually more than halved with gating compared to the full motion range (Table 1).

Large dose deterioration occurred at single fractions due to interplay effects with D2% being up to 4.7%-points higher than planned and D98% up to 4.4%-points lower than planned (Figures 5, 6; Table 2). After 5-6 fractions the interplay effects tended to smear out due to averaging effects such that D2% and D98% converged towards the planned values (Figure 6). On average the CTV HI was larger than planned by 3.7%-points (range: 1.0-6.6%-points) for individual fractions and by 0.7%-points (range: 0.3-1.1%-points) for the average dose of 5 or 6 fractions (Table 2). The mean dose to the healthy liver tissue, averaged over all analyzed fractions, was different from the planned dose by 0.3%-points (patient 1), 1%-points (patient 2) and -0.1%-points (patient 3).

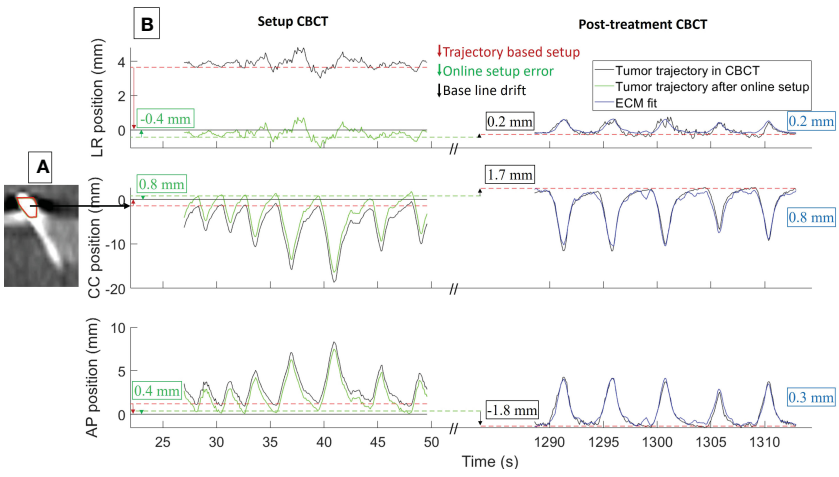


FIGURE 3
Example of **(A)** reconstructed CBCT and **(B)** tumor motion during CBCT (patient 3, fraction 3). **(A)** Blurred marker in the setup CBCT scan and the online registration to the planned exhale marker position (red contour). **(B)** Estimated 3D tumor trajectories during setup CBCT and post-treatment CBCT relative to the planned position (black curves), tumor trajectory in setup CBCT after online registration and couch correction (green curves), and external-internal correlation model (ECM) fit for the post-treatment CBCT (blue curves). The numbers show the online registration error (green), intrafraction baseline drift (black), and root mean square error of ECM fit (blue). LR, left-right; CC, cranio-caudal; AP, anterior-posterior.

TABLE 1 Mean ± standard deviation over all fractions of the online registration error, baseline drift, root-mean-square (RMS) ECM fitting error, tumor position error during spot delivery (treatment error), and the tumor motion range (2nd to 98th percentile motion) during beam-on and during field delivery regardless of beam-on status.

Patient	1			2			3		
Direction	LR (mm)	CC (mm)	AP (mm)	LR (mm)	CC (mm)	AP (mm)	LR (mm)	CC (mm)	AP (mm)
Online registration error	-0.7 ± 0.4	-0.3 ± 0.9	-0.4 ± 0.5	-0.9	-0.8	0.8	0.3 ± 0.6	1.0 ± 1.0	-0.2 ± 0.5
Baseline drift	-0.1 ± 0.5	1.4 ± 1.4	-0.8 ± 0.4	-1.2	2.0	1.7	0.3 ± 0.9	1.2 ± 1.3	-1.0 ± 0.8
RMS error of ECM fit	0.9 ± 0.3	2.1 ± 0.8	1.5 ± 0.6	0.7	1.4	0.7	0.2 ± 0.0	0.9 ± 0.2	0.5 ± 0.1
Treatment error during beam-on	-1.6 ± 1.1	-0.7 ± 3.0	0.2 ± 2.4	-1.7 ± 0.4	-0.6 ± 1.3	2.9 ± 0.5	0.1 ± 0.7	-0.7 ± 2.8	0.1 ± 1.2
Tumor motion range during beam-on periods	2.7 ± 0.3	10.9 ± 1.6	7.4 ± 1.0	1.8	5.0	2.1	0.4 ± 0.2	6.8 ± 1.3	3.5 ± 0.7
Full tumor motion range during beam-on and beam-off periods	5.1 ± 0.8	23.8 ± 2.7	15.9 ± 2.1	3.9	13.9	6.1	0.9 ± 0.5	15.9 ± 2.1	7.7 ± 1.1

4 Discussion

With the present study we have developed and clinically demonstrated a method to estimate the internal target motion and its consequence on dose delivery during proton therapy of liver cancer at a standard equipped proton facility. As pointed out by recent international guidelines such monitoring is important for PBS proton therapy of moving targets but typically not commercially available (12, 13). The motion estimation was based on an ECM that was constructed at each treatment fraction using external RPM motion synchronized with internal 3D tumor motion extracted from CBCT projections. An in-house developed method that has been validated on a group of ten patients (11, 25) was subsequently applied to reconstruct the motion-including CTV dose. Considerable interplay effects at single fractions tended to smear out after more fractions.

Additionally, we investigated the accuracy of the online CBCT match to determine the exhale marker positions. Due to motion smearing and motion artifacts, the manual online match is subjective and prone to human errors, while the estimated

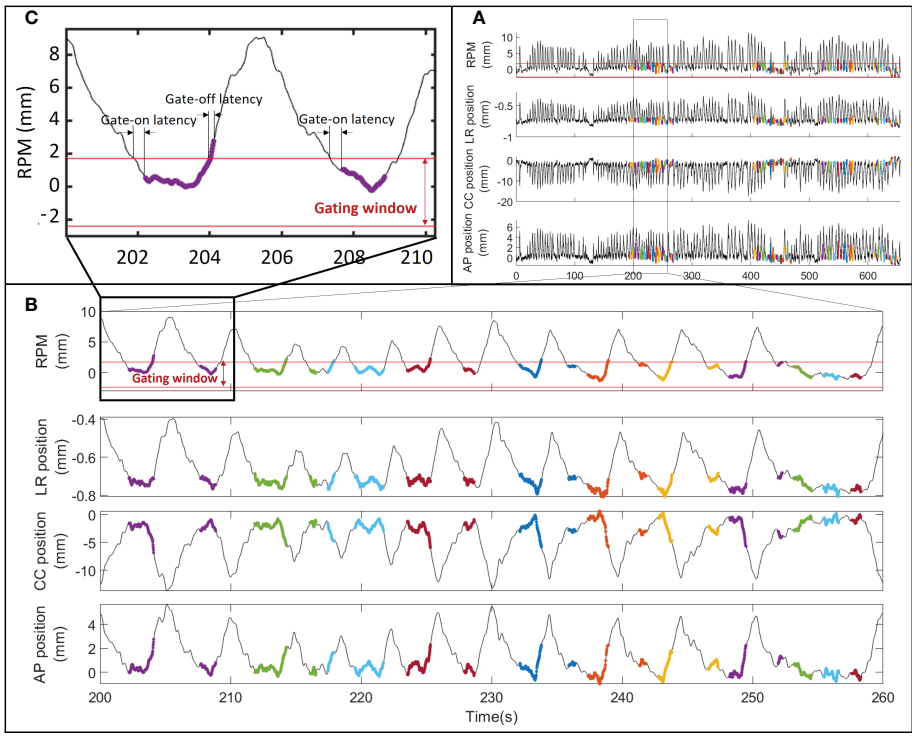


FIGURE 4
RPM signal and estimated internal tumor motion for fraction 1 of patient 3 during **(A)** the entire fraction, **(B)** a single field and **(C)** two breathing cycles. The thick colored curves show the spot delivery times with different energy layers indicated with different colors. The gating window and the gate-on and gate-off latencies are shown for the RPM signal.

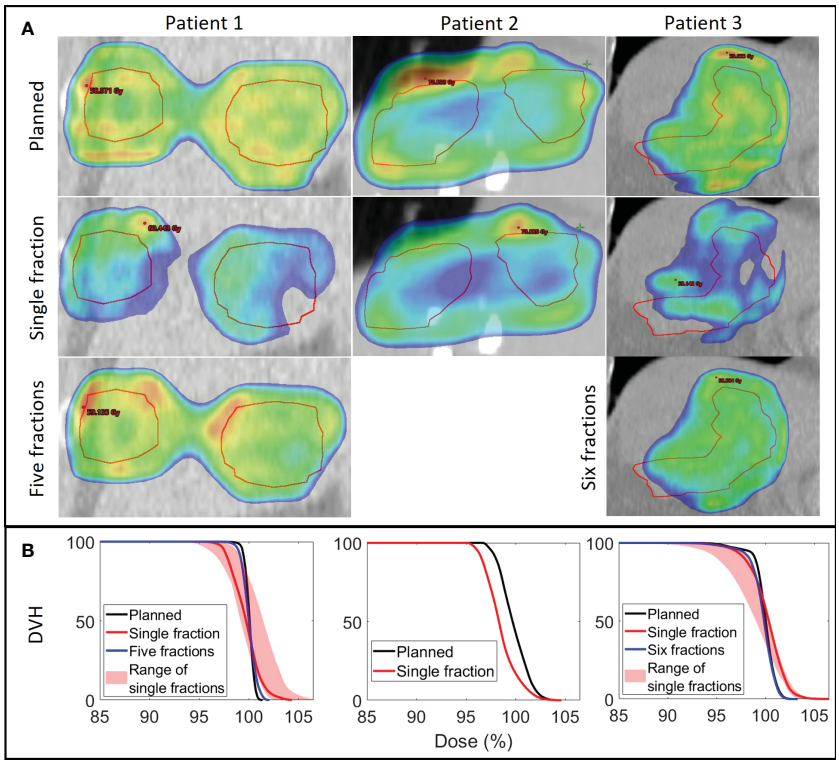


FIGURE 5
(A) Planned dose (1st row) and examples of reconstructed doses at a single fraction (2nd row) and averaged over all investigated fractions (3rd row) shown in a coronal plane through the center of the CTV (red contour) for each patient. Dose levels ≥ 95% are shown. Patient 2 only received one fraction. **(B)** Corresponding dose volume histograms (DVHs) for the CTV with the full range of single fraction DVHs indicated by the shaded area.

marker trajectories provided an objective measure of the exhale position during CBCT. However, with differences between the online manual registration and offline trajectory-based marker match close to the resolution of the CT scan (2mm CC, ~1mm in-plane), the online match accuracy was acceptable. Yet, at a few individual fractions, slightly larger discrepancies in the CC direction were observed (up to 3.2 mm). Offline inspection of the online match revealed that these discrepancies could be ascribed to the online procedure being prone to human subjectivity and performed under time pressure with the patient waiting for treatment.

During CBCT scan and treatment delivery, large motion variations between individual respiratory cycles and total motion amplitudes of 2-3 cm, most prominent in the CC direction, were observed (Figures 3, 4). Such motion is on par with previous studies of internal motion during radiotherapy of tumors in the liver (8, 32–35). Due to the extended time typically spent near the exhale phase of the respiratory cycle, limiting the treatment to an approximate 50% duty cycle around exhale more than halved the motion during beam-on and also reduced motion variation between treatment fractions (Table 1). The resulting mean geometrical treatment

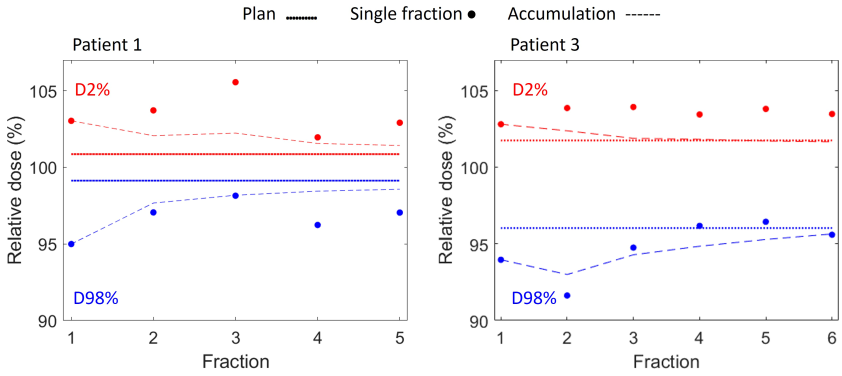


FIGURE 6
CTV D98% (blue) and D2% (red) for planned dose (solid lines), for reconstructed single fraction doses (dots) and for the cumulative reconstructed dose of 1-5 fractions (Patient 1, left) and 1-6 fractions (Patient 3, right) (dashed curves).

TABLE 2 D98%, D2% and homogeneity index (HI) for the CTV as planned and in the reconstructed doses for single fractions and averaged over all investigated fractions.

Patient	1			2			3		
Dose parameter	D98% (Gy)	D2% (Gy)	HI (%)	D98% (Gy)	D2% (Gy)	HI (%)	D98% (Gy)	D2% (Gy)	HI (%)
Planned	57.5	58.5	1.7	65.6	69.4	5.6	55.7	59.0	5.7
Single fraction (Mean ± SD)	56.1 ± 0.6	60.0 ± 0.7	6.7 ± 0.9	64.6	69.0	6.6	55.0 ± 0.9	60.0 ± 0.2	8.8 ± 1.7
5-6 fractions	57.2	58.8	2.9	–	–	–	55.4	58.9	6.0

Only one fraction was investigated for patient 2.

errors during proton spot delivery were thereby also limited to a few millimeters, including errors introduced by the rather small baseline shift of 0 - 3 mm typically observed between setup CBCTs and post-treatment CBCTs (Table 1). Notably, though not observed in the limited cohort of this study, baseline shifts and resulting treatment errors of higher magnitude can be expected during liver treatments of some patients (8, 33). Still, the initial findings of the present study confirmed the usability of externally guided respiratory gating, which has also been proposed by other groups for reducing the internal motion during liver proton therapy delivery (14–16).

Despite the use of respiraory gating, our dose reconstructions showed considereable dose deteriorations on a single-fraction level caused by interplay between proton spot delivery and target motion (Figures 5, 6). This finding clinically confirms the simulation results of Zhang et al. who concluded that respiratory gating alone was insufficient to mitigate interplay effects in PBS proton therapy (16). However, fractionation tends to reduce interplay effects by averaging out local over and under dosage over several treatment fractions (36–39). In the present study, the interplay effects were almost averaged out after 5-6 fractions (Figure 5). Nevertheless, for hypofractionation treatments, one may need to combine gating with repainting (16, 40).

A few previous clinical studies investigated the dosimetric consequences of respiratory motion during particle therapy. Richter et al. (19) and Meijers et al. (21) combined machine log files with the spot delivery with an external respiratory signal obtained during treatment and used this to calculate the D4DD by distributing each spot delivery into corresponding phases of 4DCT scans. These studies also found that fractionation effectively mitigated interplay effects. For tumors in the thoraic region, anatomical changes such as presence of fluid or tumor shrinkage, caused more severe dosimetric changes (21). Limitations of the D4DD include neglectance of setup errors, the assumption of identical anatomy and respiratory motion amplitude at treatment as in the 4DCT and the dependence on deformable image registration for dose accumulation in a reference 4DCT phase. An advantage of the 4DCT based dose reconstruction, compared to our method, is that it includes estimations of dose degration caused by internal anatomical changes between individual 4DCT phases. Since our spot shift dose reconstruction was based on a single phase of the planning 4DCT it only considers the effects of rigid intrafraction motion. A potential improvement could be a hybrid dose reconstruction method that extends phase specific dose calculations in each 4DCT phase with spot shifts that accounts for the tumor motion during treatment that goes beyond the motion in the 4DCT scan. Such an extension of our method to individual 4DCT phases would improve the D4DD reconstruction method recommended by international guidelines to also include actual intrafraction motion and not only motion observed in 4DCTs (12, 13). However, a persisting challenge with such 4DCT dose accumulation is the reliance on deformable image registration and dose warping between CT scans, which is a procedure with considerable uncertainty (41).

In a recent and closely related study by Yamada et al. at Hokkaido University, machine log files were combined with intra-treatment monitoring of internal fiducial markers by stereoscopic x-ray fluoroscopy during liver proton therapy (24). The same dose reconstruction method as applied in the present study was used (25). Respiratory exhale gating guided by direct internal tumor motion monitoring was a big advantage of the study by Yamada et al. compared to ours. Tight internal gating windows of ±2 mm along each direction may also explain why Yamada et al. found considerably smaller mean 3D tumor position errors during spot delivery (0.8-1.3 mm) than the present study (2.6-3.9 mm) where only external monitoring with a gating window corresponding to approximately 50% duty cycle was applied. For comparison, a previous study on liver SBRT with internal exhale gating based on implanted electromagnetic markers and gating windows of ±3 mm (LR/AP) and ±4 mm (CC) found mean 3D tumor position errors of 1.2-3.0 mm (8). Despite the larger treatment errors in the present study the estimated delivered CTV dose was close to the planned dose when averaged over 5-6 fractions. Hence, the robust SFUD planning approach combined with CBCT-based setup to internal markers and fractionated exhale gated treatments provided appropriate mitigation of intrafraction motion. Further studies including more patients and the effects of inter-fraction deformations (e.g., by dose reconstruction on weekly control 4DCT scans) are necessary to conclude on the overall treatment quality. For example, deformations may affect the proton range and the assumption that implanted markers serve as an accurate surrogate for the CTV position, especially if they are not implanted near the tumor (42). Ultimately, comprehensive all-inclusive fraction-specific dose reconstruction and dose accumulation could be used to trigger plan-adaptations in case of unacceptable dose coverage.

A limitation of the current study is the indirect estimation of the tumor motion during treatment by an ECM. Although an ECM of the day was built, the accuracy cannot be expected to be better than the ECM fit, which had mean RMS errors of almost half of the treatment errors and exceeding 2 mm in the CC direction for one patient (Table 1). Furthermore, high intrafraction stability of the ECM with the ability to detect internal baseline shift cannot in general be assumed (43–46). Optimally, the ECM should be built based on the setup CBCT and then validated by the post-treatment CBCT.

However, phantom tests showed discrepancies between couch shifts and changes in the RPM signal that hindered correct adjustment of an ECM from a setup CBCT to allow usage after setup couch corrections. For this reason, intra-treatment motion could only be investigated for fractions with a post-treatment CBCT in this study. The ECM stability issue may be addressed by capturing a series of stereoscopic images before each field delivery to confirm and update the ECM.

It is worth noting that the RPM system available for proton therapy clearly lags behind the optical monitoring system available from the same vendor for photon radiotherapy (TrueBeam, Varian) with a stereoscopic camera that allows adaptation of the ECM to couch shifts (47). Furthermore, the cumbersome manual synchronizations of the RPM log files with CBCT projections and treatment machine log files in the current study (Figure 1) would not be needed if the respiratory monitoring, imaging, and beam delivery systems were similarly well-integrated as on a TrueBeam linear accelerator. The manual synchronization of RPM log files with machine log files was only possible in this study because the beam pauses in the machine log files had a unique temporal pattern that could be matched with the RPM gating signal (Figure 1E). This synchronization would not be possible for non-gated treatments. Since the synchronization can only be performed post-treatment, it is currently a barrier for online real-time dose reconstruction, which has been demonstrated clinically for liver SBRT (48), but could be even more relevant for proton PBS.

In summary, dose reconstruction including the effects of setup errors, rigid motion, interplay effects and the smearing hereof after 5-6 fractions was performed for HCC proton therapy. For the included patients, it showed that our treatment strategy of exhale gating resulted in an acceptable CTV dose coverage. With a smoother workflow and automation this could be used to trigger a plan adaptation if the CTV dose coverage turned out to be unacceptable. Since CTV dose deficits could also be caused by interfractional changes, the motion-including dose reconstruction should ideally be extended to account for such changes, for example by applying it on the anatomy of weekly 4DCTs.

5 Conclusion

A method to estimate internal tumor motion and reconstruct the motion-including fraction dose for PBS proton therapy in the liver was developed and successfully demonstrated clinically at a conventional proton facility.

Data availability statement

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

Ethics statement

The studies involving human participants were reviewed and approved by the regional research ethics committee and the regional register of research projects (ref. number 742503, case number 1-16-02-320-21). The research was conducted in accordance with the principles of the Helsinki Declaration and local statutory requirements. The patients/participants provided their written informed consent to participate in this study.

Author contributions

EW, HM, BW, MH, JT, LS, PP designed the treatment and imaging protocol for the study. PP, SN, EW conceived, designed and drafted the manuscript and all authors revised the manuscript. JT, LS, SN, EW collected the data. Software was developed by JB (marker segmentation) and PP (segmentation correction, 3D trajectory estimation, data synchronization, dose reconstruction), and streamlined by SN who performed the data analysis. All authors contributed to the article and approved the submitted version.

Funding

This work was kindly supported by DCCC Radiotherapy – The Danish National Research Center for Radiotherapy, Danish Cancer Society [grants no. R269-A15843 and R191-A11526], Novo Nordisk Foundation [grant no. NNF20OC0065282] and Danish Comprehensive Cancer Center. Dr Worm received funding from a research foundation at the Department of Clinical Medicine, Aarhus University Hospital.

Conflict of interest

PP is co-inventor on a patent on the 3D motion estimation method applied in this study.

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

References

1. Benson AB, D'Angelica MI, Abbott DE, Abrams TA, Alberts SR, Anaya DA, et al. Nccn guidelines insights: Hepatobiliary cancers, version 1.2017. *J Natl Compr Cancer Network* (2017) 15(5):563–73. doi: 10.6004/jnccn.2017.0059

2. Cheng JC-H, Wu J-K, Lee PC-T, Liu H-S, Jian JJ-M, Lin Y-M, et al. Biologic susceptibility of hepatocellular carcinoma patients treated with radiotherapy to radiation-induced liver disease. *Int J Radiat Oncol Biol Phys* (2004) 60(5):1502–9. doi: 10.1016/j.ijrobp.2004.05.048

3. Miften M, Vinogradskiy Y, Moiseenko V, Grimm J, Yorke E, Jackson A, et al. Radiation dose-volume effects for liver sbrt. *Int J Radiat Oncol Biol Phys* (2021) 110 (1):196–205. doi: 10.1016/j.ijrobp.2017.12.290

4. Lomax AJ, Pedroni E, Rutz HP, Goitein G. The clinical potential of intensity modulated proton therapy. *Z für Medizinische Physik* (2004) 14(3):147–52. doi: 10.1078/0939-3889-00217

5. Yoo GS, Yu JI, Park HC. Proton therapy for hepatocellular carcinoma: Current knowledges and future perspectives. *World J Gastroenterol* (2018) 24(28):3090. doi: 10.3748/wjg.v24.i28.3090

6. Kobeissi JM, Hilal L, Simone CB2nd, Lin H, Crane CH, Hajj C. Proton therapy in the management of hepatocellular carcinoma. *Cancers* (2022) 14(12):2900. doi: 10.3390/cancers14122900

7. Spychalski P, Kobiela J, Antoszewska M, Błażyńska-Spychalska A, Jereczek-Fossa BA, Høyer M. Patient specific outcomes of charged particle therapy for hepatocellular carcinoma–a systematic review and quantitative analysis. *Radiotherapy Oncol* (2019) 132:127–34. doi: 10.1016/j.radonc.2018.12.012

8. Worm ES, Høyer M, Hansen R, Larsen LP, Weber B, Grau C, et al. A prospective cohort study of gated stereotactic liver radiation therapy using continuous internal electromagnetic motion monitoring. *Int J Radiat Oncol Biol Phys* (2018) 101(2):366–75. doi: 10.1016/j.ijrobp.2018.02.010

9. Kooy H, Grassberger C. Intensity modulated proton therapy. *Br J Radiol* (2015) 88(1051):20150195. doi: 10.1259/bjr.20150195

10. Rietzel E, Bert C. Respiratory motion management in particle therapy. *Med Phys* (2010) 37(2):449–60. doi: 10.1118/1.3250856

11. Worm ES, Hansen R, Høyer M, Weber B, Mortensen H, Poulsen PR. Uniform versus non-uniform dose prescription for proton stereotactic body radiotherapy of liver tumors investigated by extensive motion-including treatment simulations. *Phys Med Biol* (2021) 66(20):205009. doi: 10.1088/1361-6560/ac2880

12. Li H, Dong L, Bert C, Chang J, Flampouri S, Jee KW, et al. AAPM task group report 290: Respiratory motion management for particle therapy. *Med Phys* (2022) 49 (4):e50–81. doi: 10.1002/mp.15470

13. Chang JY, Zhang X, Knopf A, Li H, Mori S, Dong L, et al. Consensus guidelines for implementing pencil-beam scanning proton therapy for thoracic malignancies on behalf of the ptcog thoracic and lymphoma subcommittee. *Int J Radiat Oncol Biol Phys* (2017) 99(1):41–50. doi: 10.1016/j.ijrobp.2017.05.014

14. Lu HM, Brett R, Sharp G, Safai S, Jiang S, Flanz J, et al. A respiratory-gated treatment system for proton therapy. *Med Phys* (2007) 34(8):3273–8. doi: 10.1118/1.2756602

15. Gelover E, Deisher AJ, Herman MG, Johnson JE, Kruse JJ, Tryggestad EJ. Clinical implementation of respiratory-gated spot-scanning proton therapy: An efficiency analysis of active motion management. *J Appl Clin Med Phys* (2019) 20 (5):99–108. doi: 10.1002/acm2.12584

16. Zhang Y, Huth I, Weber DC, Lomax AJ. A statistical comparison of motion mitigation performances and robustness of various pencil beam scanned proton systems for liver tumour treatments. *Radiotherapy Oncol* (2018) 128(1):182–8. doi: 10.1016/j.radonc.2018.01.019

17. Mizuhata M, Takamatsu S, Shibata S, Bou S, Sato Y, Kawamura M, et al. Respiratory-gated proton beam therapy for hepatocellular carcinoma adjacent to the gastrointestinal tract without fiducial markers. *Cancers* (2018) 10(2):58. doi: 10.3390/cancers10020058

18. Sharp GC, Lu HM, Trofimov A, Tang X, Jiang SB, Turcotte J, et al. Assessing residual motion for gated proton-beam radiotherapy. *J Radiat Res* (2007) 48(Suppl_A): A55–A9. doi: 10.1269/jrr.48.A55

19. Richter D, Saito N, Chaudhri N, Härtig M, Ellerbrock M, Jäkel O, et al. Four-dimensional patient dose reconstruction for scanned ion beam therapy of moving liver tumors. *Int J Radiat Oncol Biol Phys* (2014) 89(1):175–81. doi: 10.1016/j.ijrobp.2014.01.043

20. Meijers A, Jakobi A, Stützer K, Guterres Marmitt G, Both S, Langendijk J, et al. Log file-based dose reconstruction and accumulation for 4D adaptive pencil beam scanned proton therapy in a clinical treatment planning system: Implementation and proof-of-Concept. *Med Phys* (2019) 46(3):1140–9. doi: 10.1002/mp.13371

21. Meijers A, Knopf A-C, Crijns AP, Ubbels JF, Niezink AG, Langendijk JA, et al. Evaluation of interplay and organ motion effects by means of 4D dose reconstruction and accumulation. *Radiotherapy Oncol* (2020) 150:268–74. doi: 10.1016/j.radonc.2020.07.055

22. Worm ES, Høyer M, Fledelius W, Hansen AT, Poulsen PR. Variations in magnitude and directionality of respiratory target motion throughout full treatment courses of stereotactic body radiotherapy for tumors in the liver. *Acta Oncol* (2013) 52 (7):1437–44. doi: 10.3109/0284186X.2013.813638

23. Rankine L, Wan H, Parikh P, Maughan N, Poulsen P, DeWees T, et al. Cone-beam computed tomography internal motion tracking should be used to validate 4-dimensional computed tomography for abdominal radiation therapy patients. *Int J Radiat Oncol Biol Phys* (2016) 95(2):818–26. doi: 10.1016/j.ijrobp.2016.01.047

24. Yamada T, Takao S, Koyano H, Nihongi H, Fujii Y, Hirayama S, et al. Validation of dose distribution for liver tumors treated with real-Time-Image gated spot-scanning proton therapy by log data based dose reconstruction. *J Radiat Res* (2021) 62(4):626–33. doi: 10.1093/jrr/rrab024

25. Colvill E, Petersen JB, Hansen R, Worm E, Skouboe S, Høyer M, et al. Validation of fast motion-including dose reconstruction for proton scanning therapy in the liver. *Phys Med Biol* (2018) 63(22):225021. doi: 10.1088/1361-6560/aaeae9

26. Bertholet J, Wan H, Toftegaard J, Schmidt M, Chotard F, Parikh P, et al. Fully automatic segmentation of arbitrarily shaped fiducial markers in cone-beam CT projections. *Phys Med Biol* (2017) 62(4):1327. doi: 10.1088/1361-6560/aa52f7

27. Poulsen PR, Cho B, Keall PJ. A method to estimate mean position, motion magnitude, motion correlation, and trajectory of a tumor from cone-beam CT projections for image-guided radiotherapy. *Int J Radiat Oncol Biol Phys* (2008) 72 (5):1587–96. doi: 10.1016/j.ijrobp.2008.07.037

28. Worm ES, Høyer M, Fledelius W, Nielsen JE, Larsen LP, Poulsen PR. On-line use of three-dimensional marker trajectory estimation from cone-beam computed tomography projections for precise setup in radiotherapy for targets with respiratory motion. *Int J Radiat Oncol Biol Phys* (2012) 83(1):e145–51. doi: 10.1016/j.ijrobp.2011.12.007

29. Worm E, Thomsen J, Johansen J, Poulsen P. Oc-0040 gating latencies and resulting geometrical errors at clinical proton and photon accelerators. *Radiotherapy Oncol* (2022) 170:S13–S5. doi: 10.1016/S0167-8140(22)02459-8

30. Poulsen PR, Eley J, Langner U, Simone CBII, Langen K. Efficient interplay effect mitigation for proton pencil beam scanning by spot-adapted layered repainting evenly spread out over the full breathing cycle. *Int J Radiat Oncol Biol Phys* (2018) 100(1):226–34. doi: 10.1016/j.ijrobp.2017.09.043

31. Ruan D, Fessler JA, Balter JM, Berbeco R, Nishioka S, Shirato H. Inference of hysteretic respiratory tumor motion from external surrogates: A state augmentation approach. *Phys Med Biol* (2008) 53(11):2923. doi: 10.1088/0031-9155/53/11/011

32. Park JC, Park SH, Kim JH, Yoon SM, Song SY, Liu Z, et al. Liver motion during cone beam computed tomography guided stereotactic body radiation therapy. *Med Phys* (2012) 39(10):6431–42. doi: 10.1118/1.4754658

33. Worm ES, Høyer M, Fledelius W, Poulsen PR. Three-dimensional, time-resolved, intrafraction motion monitoring throughout stereotactic liver radiation therapy on a conventional linear accelerator. *Int J Radiat Oncol Biol Phys* (2013) 86 (1):190–7. doi: 10.1016/j.ijrobp.2012.12.017

34. Poulsen PR, Worm ES, Petersen JB, Grau C, Fledelius W, Høyer M. Kilovoltage intrafraction motion monitoring and target dose reconstruction for stereotactic volumetric modulated arc therapy of tumors in the liver. *Radiotherapy Oncol* (2014) 111(3):424–30. doi: 10.1016/j.radonc.2014.05.007

35. Xu Q, Hanna G, Grimm J, Kubicek G, Pahlajani N, Asbell S, et al. Quantifying rigid and nonrigid motion of liver tumors during stereotactic body radiation therapy. *Int J Radiat Oncol Biol Phys* (2014) 90(1):94–101. doi: 10.1016/j.ijrobp.2014.05.007

36. Dolde K, Zhang Y, Chaudhri N, Dávid C, Kachelrieß M, Lomax AJ, et al. 4DMRI-based investigation on the interplay effect for pencil beam scanning proton therapy of pancreatic cancer patients. *Radiat Oncol* (2019) 14(1):1–13. doi: 10.1186/s13014-019-1231-2

37. Grassberger C, Dowdell S, Lomax A, Sharp G, Shackleford J, Choi N, et al. Motion interplay as a function of patient parameters and spot size in spot scanning proton therapy for lung cancer. *Int J Radiat Oncol Biol Phys* (2013) 86(2):380–6. doi: 10.1016/j.ijrobp.2013.01.024

38. Dowdell S, Grassberger C, Sharp G, Paganetti H. Interplay effects in proton scanning for lung: A 4D Monte Carlo study assessing the impact of tumor and beam delivery parameters. *Phys Med Biol* (2013) 58(12):4137. doi: 10.1088/0031-9155/58/12/4137

39. Li H, Li Y, Zhang X, Li X, Liu W, Gillin MT, et al. Dynamically accumulated dose and 4D accumulated dose for moving tumors. *Med Phys* (2012) 39(12):7359–67. doi: 10.1118/1.4766434

40. Furukawa T, Inaniwa T, Sato S, Shirai T, Mori S, Takeshita E, et al. Moving target irradiation with fast rescanning and gating in particle therapy. *Med Phys* (2010) 37(9):4874–9. doi: 10.1118/1.3481512

41. Ribeiro CO, Knopf A, Langendijk JA, Weber DC, Lomax AJ, Zhang Y. Assessment of dosimetric errors induced by deformable image registration methods in 4D pencil beam scanned proton treatment planning for liver tumours. *Radiotherapy Oncol* (2018) 128(1):174–81. doi: 10.1016/j.radonc.2018.03.001

42. Wunderink W, Romero AM, Seppenwoolde Y, De Boer H, Levendag P, Heijmen B. Potentials and limitations of guiding liver stereotactic body radiation therapy set-up on liver-implanted fiducial markers. *Int J Radiat Oncol Biol Phys* (2010) 77(5):1573–83. doi: 10.1016/j.ijrobp.2009.10.040

43. Ge J, Santanam L, Yang D, Parikh PJ. Accuracy and consistency of respiratory gating in abdominal cancer patients. *Int J Radiat Oncol Biol Phys* (2013) 85(3):854–61. doi: 10.1016/j.ijrobp.2012.05.006

44. Pettersson N, Oderinde OM, Murphy J, Simpson D, Cerviño LI. Intrafractional relationship changes between an external breathing signal and fiducial marker positions in pancreatic cancer patients. *J Appl Clin Med Phys* (2020) 21(3):153–61. doi: 10.1002/acm2.12841

45. Takao S, Miyamoto N, Matsuura T, Onimaru R, Katoh N, Inoue T, et al. Intrafractional baseline shift or drift of lung tumor motion during gated radiation therapy with a real-time tumor-tracking system. *Int J Radiat Oncol Biol Phys* (2016) 94 (1):172–80. doi: 10.1016/j.ijrobp.2015.09.024

46. Ren Q, Nishioka S, Shirato H, Berbeco R. Adaptive external gating based on the updating method of Internal/External correlation and gating window before each beam delivery. *Phys Med Biol* (2012) 57(9):N145. doi: 10.1088/0031-9155/57/9/N145

47. Bertholet J, Toftegaard J, Hansen R, Worm ES, Wan H, Parikh PJ, et al. Automatic online and real-time tumour motion monitoring during stereotactic liver treatments on a conventional linac by combined optical and sparse monoscopic imaging with kilovoltage X-rays (Cosmik). *Phys Med Biol* (2018) 63(5):055012. doi: 10.1088/1361-6560/aaae8b

48. Skouboe S, Ravkilde T, Bertholet J, Hansen R, Worm ES, Muurholm CG, et al. First clinical real-time motion-including tumor dose reconstruction during radiotherapy delivery. *Radiotherapy Oncol* (2019) 139:66–71. doi: 10.1016/j.radonc.2019.07.007

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 16 March 2023
DOI 10.3389/fonc.2023.988784

OPEN ACCESS

EDITED BY
Jasper Nijkamp,
Aarhus University, Denmark

REVIEWED BY
Stephen Raskin,
Sheba Medical Center, Israel
Bo Zhou,
Yale University, United States

*CORRESPONDENCE
Hubert Beaumont
hubertbeaumont@hotmail.com

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

RECEIVED 07 July 2022
ACCEPTED 03 February 2023
PUBLISHED 16 March 2023

CITATION
Iannessi A and Beaumont H (2023) Breaking down the RECIST 1.1 double read variability in lung trials: What do baseline assessments tell us?
Front. Oncol. 13:988784.
doi: 10.3389/fonc.2023.988784

Breaking down the RECIST 1.1 double read variability in lung trials: What do baseline assessments tell us?

Antoine Iannessi and Hubert Beaumont*

Imaging Lab, Median Technologies, Valbonne, France

Background: In clinical trials with imaging, Blinded Independent Central Review (BICR) with double reads ensures data blinding and reduces bias in drug evaluations. As double reads can cause discrepancies, evaluations require close monitoring which substantially increases clinical trial costs. We sought to document the variability of double reads at baseline, and variabilities across individual readers and lung trials.

Material and methods: We retrospectively analyzed data from five BICR clinical trials evaluating 1720 lung cancer patients treated with immunotherapy or targeted therapy. Fifteen radiologists were involved. The variability was analyzed using a set of 71 features derived from tumor selection, measurements, and disease location. We selected a subset of readers that evaluated ≥50 patients in ≥two trials, to compare individual reader's selections. Finally, we evaluated inter-trial homogeneity using a subset of patients for whom both readers assessed the exact same disease locations. Significance level was 0.05. Multiple pair-wise comparisons of continuous variables and proportions were performed using one-way ANOVA and Marascuilo procedure, respectively.

Results: Across trials, on average per patient, target lesion (TL) number ranged 1.9 to 3.0, sum of tumor diameter (SOD) 57.1 to 91.9 mm. MeanSOD=83.7 mm. In four trials, MeanSOD of double reads was significantly different. Less than 10% of patients had TLs selected in completely different organs and 43.5% had at least one selected in different organs. Discrepancies in disease locations happened mainly in lymph nodes (20.1%) and bones (12.2%). Discrepancies in measurable disease happened mainly in lung (19.6%). Between individual readers, the MeanSOD and disease selection were significantly different ($p<0.001$). In inter-trials comparisons, on average per patient, the number of selected TLs ranged 2.1 to 2.8, MeanSOD 61.0 to 92.4 mm. Trials were significantly different in MeanSOD ($p<0.0001$) and average number of selected TLs ($p=0.007$). The proportion of patients having one of the top diseases was significantly different only between two

trials for lung. Significant differences were observed for all other disease locations ($p<0.05$).

Conclusions: We found significant double read variabilities at baseline, evidence of reading patterns and a means to compare trials. Clinical trial reliability is influenced by the interplay of readers, patients and trial design.

KEYWORDS

clinical trial, variability, RECIST, computed tomography, lung cancer

1 Highlights

- In RECIST BICR trials with double reads there is large variability in tumor measurement and localization.
- Individual reader's assessments are significantly different
- Advanced lung cancer trials with similar treatments can be significantly different in terms of baseline assessments.

2 Background

Since 2004 (1, 2), Blinded Independent Central Review (BICR) with double reads has been promoted in clinical trials with imaging to ensure data blinding and to reduce bias (3). A direct consequence of double reads is inter-reader variability. Because of these variabilities, discrepancies in the evaluation of treatment response during trials with double reads need to be monitored and, eventually, be adjudicated by a third reader (4). This directly impacts the quality and the cost of clinical trials that all stakeholders strive to mitigate. A better understanding of the root causes of the variability is needed. The ability to trigger warnings as early as after baseline evaluations would help reduce inter-reader variabilities during trial monitoring.

In clinical trials for drug development, the discrepancy rate of the treatment response assessment is the preferred indicator that summarizes the reliability of treatment evaluation (5). However, the discrepancy rate is a high-level indicator that encompasses all possible root causes of variability including the technical variability of image acquisition (6) and the interpretation of images (aka. reader variability). To manage variability, standardized reading rules are applied to radiology assessments that quantify the response: the response evaluation criteria, i.e. Response Evaluation Criteria in Solid Tumors (RECIST).

Several criteria-derived variability factors have been documented (7, 8). A large proportion of discrepant responses originate due to the subjectivity of the baseline assessment (9, 10) when using RECIST (11). Indeed, many imaging response criteria are based on the relative modifications from baseline, therefore it is logical that the initial definition of the disease has an impact on the response.

When endorsing an omnibus reference value of discrepancy rates based on the literature, the underlining assumption is that variabilities in reads, and the magnitudes of these variabilities, are consistent across "comparable" trials (12) however, this assumption has not been clearly confirmed. Indeed, less attention was given to the variability caused by the initial disease presentation and the heterogeneity of recruited readers across so called "comparable" trials.

In this paper, we consider similar lung clinical trials and focus on the baseline analysis of imaging data. We describe the distribution of double read variabilities, and compare the specificities of assessments between readers and, for individual readers across trials.

3 Methods

3.1 Study data

We aimed at minimizing the heterogeneity of our data by selecting studies with "comparable" indication and inclusion criteria. Our retrospective analysis included assessment data from five BICR clinical trials (Trials 1-5) that evaluated immunotherapy or targeted therapy for lung cancer. The selected BICR trials were conducted between 2017 and 2021 and used double reads with adjudication based on RECIST 1.1 guidelines. All data were fully blinded for sponsor data, study protocol number, therapeutic agent, subject demographics, and randomization. For these five trials, a total of 1720 patients were evaluated by 15 Board Certified US and Europe, 10y+ Senior Radiologist with previous experience in central RECIST 1.1 assessment (Reader R1-R15) (Table 1). The central reads were all performed using the same radiological reading platform (LMS; Median Technologies, France) ensuring automatic data extraction for analysis.

3.2 Independent central review

The pool of 15 independent radiologists reading across the five selected trials were trained on the RECIST 1.1 criteria and study protocol inclusion criteria regarding brain metastasis to perform a BICR of each baseline image and to determine the radiologic timepoint response in accordance with these read rules. In each

Abbreviations: ANOVA, Analysis of Variance; BICR, Blinded Independent Central Review; IR, Independent Reader; KPI, Key Performance Indicator; NSCLC, Non-Small Cell Lung Cancer; NTL, Non-Target Lesion; R, Reader; RECIST, Response Evaluation Criteria in Solid Tumor; SOD, Sum of Diameter; TL, Target Lesion.

TABLE 1 Description of included trials.

Trial ID	Indication	Phase	Therapy	Study specific criteria (protocol and read rules)	Readers ID
Trial 1	Metastatic NSCLC	III	Immune checkpoints + chemotherapy *vs.* Chemotherapy + placebo	Measurable disease No central eligibility Brain metastases can only be non-targe lesions	R2, R4, R5
Trial 2	Metastatic NSCLC	III	Immune checkpoints + chemotherapy *vs.* Chemotherapy + placebo	Measurable disease Central eligibility process Brain metastases can only be non-target lesions	R4, R5, R6
Trial 3	Metastatic NSCLC	II	Tyrosine kinase inhibitors	Measurable disease No central eligibility Brain metastases can be target lesion	R1, R2, R6, R7
Trial 4	Metastatic NSCLC	III	Tyrosine kinase inhibitors	Measurable disease No central eligibility Brain metastases can be target lesion	R1, R3, R5, R7
Trial 5	Metastatic NSCLC	III	Immune checkpoints + chemotherapy *vs.* Chemotherapy + placebo	Measurable disease No central eligibility Brain metastases can be target lesion	R2, R3, R5, R7

Primary study endpoints were: Progression Free Survival (PFS) and Overall Response Rate (ORR). Patients were treated for Metastatic Non-Small Cell Lung Cancer (NSCLC).

trial, the radiologist roles (i.e., independent reader IR1 or IR2) were randomly assigned to a Reader ID (R1 to R15) at the onset of the trial to create a double reading paradigm. All images and readers annotations underwent a quality control (e.g., checking the conformance with RECIST guideline and to the review protocol of the study) using software and operated by dedicated staff before the patient response to be evaluated. To improve the reliability of evaluations, the double reading paradigm involve a third reader when readers disagree on patient response, even at the early steps of eligibility.

Following RECIST 1.1 criteria, the tumor burden is quantified by the sum of diameter (SOD) as the sum of the largest lesions selected as targets lesions (TLs) within the "measurable" disease. To be measurable and qualify for a TL, the finding must measure at least 1cm for solid tumor or 1.5 cm for lymph-nodes. To be representative of the metastatic disease extent, the selection should be distributed across all involved organs and avoid the priorly irradiated areas. This prior therapy information was provided to the central readers. In total, a maximum of five TLs, maximum two per organ, are selected at baseline. Then, any additional lesions, smaller lesions and truly non-measurable lesions (e.g., blastic bone lesions) are represented by selecting Non-Target Lesions (NTL) which are only qualitatively assessed. For diffuse disease, the NTL lesions can be grouped instead of itemizing each one of the metastases.

3.3 RECIST 1.1 assessment analysis

3.3.1 Initial bivariate analysis

The study plan is depicted in Figure 1. Initially, to ensure the validity of our subsequent results, we first considered a subset of readers who participated in the same subset of trials, then we checked that the variability between the readers was not linked to the studies in which they were involved. Also, that the variability between studies was not related to the readers who carried them out. We effectively measured the bi-factorial impact of reader~trial interaction on the variability of SOD and the number of selected TLs at baseline through a two-way analysis of variance.

3.3.2 Core analysis: variability according to three perspectives

The RECIST 1.1 baseline assessment provides quantitative and qualitative information on disease extent and its spread throughout organs. Accordingly, we based our variability analysis on features designed to report on the extent and spread of the disease.

As detailed below, our analysis was divided into three parts:

3.3.2.1 Double read variability

We investigated inter-reader variability in the original double read setting through a set of predefined disease-related quantitative and location features (double read-derived features shown in Table 2, disease locations listed in Annex A). These features describe the two readers' selection of patients' tumors (TLs and NTLs) at baseline as illustrated in Figure 2. We documented the distribution of these features and compared them across trials. We provided typical values.

3.3.2.2 Reader's variability

To further compare reader's selections, we selected a subset of readers that evaluated 50 patients or more in two or more trials (i.e. more than 100 evaluations by an individual reader), therefore lowering the weight of the "trial" and "patient" covariates. For each reader, we computed the average number of TLs and SOD per patient (single reader-derived features shown in Table 3) and the proportion of patients in the most represented discrepant disease location (TopDisLocDisease, see Table 2) as previously determined in our double read variability analysis. We compared these baseline selection features with each other.

3.3.2.3 Trial's variability

The five trials selected were deemed "comparable" as they evaluated advanced lung cancer treated with similar therapeutics. The third part of our study checked the validity of our assumption by analyzing inter-trial homogeneity using the subset of patients for whom both readers selected the same disease locations (either TLs or NTLs). We assumed that, as both readers agreed on disease location (as in Figure 2), the derived findings would be more reliable, therefore allowing a more relevant inter-trial comparison. For this purpose, for each of the three quantitative

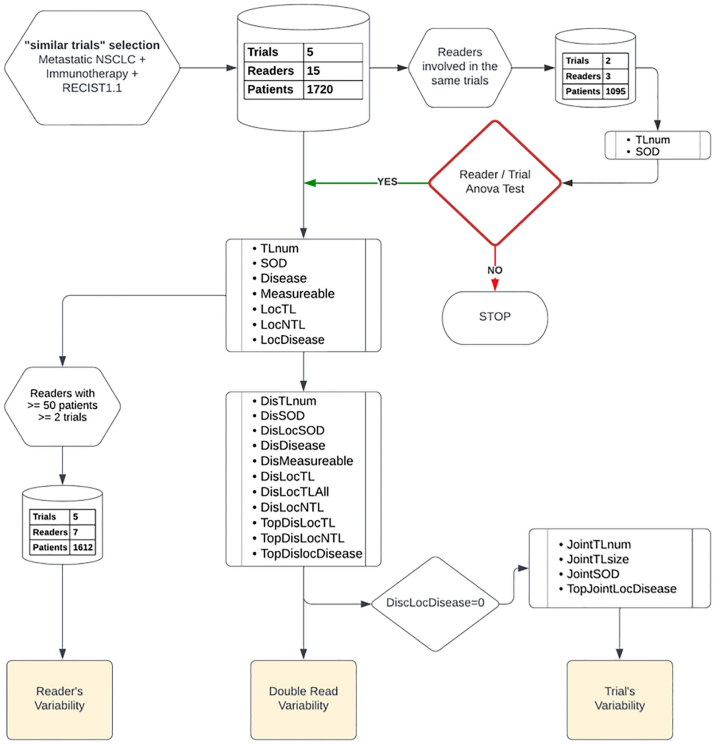


FIGURE 1
Data analysis plan: Top down, we started by a two-way factor analysis of Reader and Trial, then we documented readers, double-read, and trials' variabilities. We report, from top to bottom, the type of features and data preparation involved in each step of our analysis.

TABLE 2 Double read-extracted features.

Discrepancy analysis (averaged per double reads)	
DisTLnum	Difference in the number of TLs recorded during double reads
DisSOD	Difference of SOD recorded during double reads (proportional difference of double read SOD in %)
DisLocSOD	SOD of TL belonging to discrepant selected organs in double reads (ratio is derived by dividing by the reader total SOD value in %). Computed when one of the readers reported measurable disease
DisDisease	Proportion of patients reported with no disease at all by one of the readers (no TL and no NTL at all for one reader)
DisMeas	Proportion of patients reported with disease by both readers but one of the readers reported no measurable disease (No TL but at least one NTL)
DisLocTL	Proportion of patients for which readers targeted at least one TL at a different disease location
DisLocTLAll	Proportion of patients for which readers targeted all their TLs in totally different locations
DisLocNTL	Proportion of patients for which readers targeted at least one NTL at a different disease location
TopDisLocTL	List of the most represented discrepant TL locations
TopDisLocNTL	List of the most represented discrepant NTL locations
TopDisLocDisease	List of the most represented discrepant disease locations
Measurements derived from jointly selected organs (averaged per trial) when DisLocDisease = 0	
JointTLnum	Number of TL recorded in non-discrepant diseased organs
JointTLSize	Size of the TL recorded in non-discrepant diseased organs
JointSOD	Tumor burden recorded in non-discrepant diseased organs
TopJointLocDisease	List of the most represented non-discrepant diseased location (TL or NTL)

Feature acronyms and definitions that were used for the description of disease selection variability at baseline.

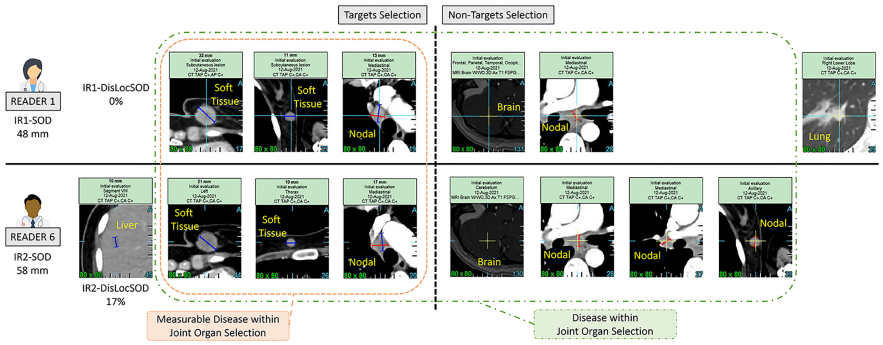


FIGURE 2
Example of a double read baseline disease assessment by RECIST 1.1. One patient in Trial 3. Each reader's selection is illustrated and grouped inside a Venn diagram illustrating the common organ selection of the disease and its measurable part. The discrepant selection within the measurable SOD illustrates the meaning and calculation of the double read DisLocSOD feature as (IR1-DisLocSOD + IR2-DisLocSOD)/2. The reader 6 identified disease in the liver and measured a centimetric liver nodule that could have been considered equivocal by reader 1.

features related to tumor burden (Table 2), we averaged the joint double read measurements. Similarly, we computed the proportion of patients for whom the top five metastatic locations were reported in agreement during double reads.

3.4 Statistics

All statistics were performed using base version and packages from R CRAN freeware. Level of significance was set to 5%. Continuous variables were analyzed using a paired two sample non-parametric Wilcoxon test. The confidence interval of the mean difference was computed using the "misty" package. A violin plot was used to display the difference of measurements. Multiple pair-wise comparisons between the five trials and between the subset of seven readers were performed. For comparison of continuous variables, when assumptions for homoscedasticity (Levene's test, "base" package) and normality (Jarque–Bera test, "lawstat" package) were met, one-way analysis of variance (ANOVA) ("base" package) was used. When these assumptions were not met, multiple comparison was performed using the Kruskal-Wallis test. Multiple comparison for proportions was performed according to Marascuilo procedure (13).

Type III two-way ANOVA ("car" package) was performed after BoxCox transformation of the measure and homoscedasticity and normality of residuals checked using Bartlett's and Shapiro-Wilk's test, respectively.

TABLE 3 Single read-extracted features.

Quantitative features	
TLNum	Number of TLs recorded per patient
TLSize	SOD/TLNum per patient (in mm)
SOD	Tumor burden as the Sum of tumor Diameter per patient (in mm)
Qualitative features	
Disease	Proportion of patients recorded with disease (at least one TL or NTL)
Meas	Proportion of patients recorded with measurable disease
LocTL	Proportion of patients recorded with TL located in a specific organ
LocNTL	Proportion of patients recorded with NTL located in a specific organ
LocDisease	Proportion of patients recorded with TL *or* NTL located in a specific organ

SOD, Sum of Diamters; TLNum, Number of target lesion; TL, Target lesion; NTL, Non target lesion.
Feature acronyms and definitions that were used for the description of disease selection at baseline per reader.

4 Results

As a preliminary analysis, we considered three readers that were all involved in the same two trials with a total of 1095 patients (see Figure 1). We confirmed that reader and trial are both factors contributing to the variability of SOD and the number of selected TLs ($p<0.0001$). The interaction between the two main factors was not significant for SOD ($p=0.24$) or the number of selected TLs ($p=0.67$), meaning that, for our data, inter-reader variability had no effect on the measurement of inter-trial variability and vice versa.

4.1 Double read variability

The differences between reader's measurements are summarized in the Table 4.

Regarding eligibility, our analysis showed a very low discrepancy rate for disease detection at baseline with an overall discrepancy rate <0.1%. The measurement of tumor burden was more variable with an overall discrepancy rate of approximately 6% for studies without a centralized eligibility process.

In all trials the two pools of readers (IR1 and IR2) selected a statistically significantly different average number of TLs per patient, ranging from 1.9 to 3.0 across trials (median values being either 2 or 3). The average SOD per patient ranged from 57.1 mm to 91.9 mm across trials with an overall average SOD of approximately 84 mm. The difference in the number of TLs during double reads (DisTLnum) was often higher than 2 and could be as high as 4 (Figure 3A). The

TABLE 4 Double-read measurements (test of differences).

	DisDisease	DisMeas	Read-TLnum	Read-SOD (mm)	Read-DisLocSOD (%)
Trial 1 ***(N=333)***	0%	4.8% (16/333)	IR1-TLNum=2.1 ** IR2-TLNum=2.3	IR1-SOD=77.7 ** IR2-SOD=86.7	IR1-SPropSOD=19.2 ** IR2-SPropSOD=12.6
Trial 2 ***(N=493)***	0%	0.4% (2/493)	IR1-TLNum=1.9 ** IR2-TLNum=2.3	IR1-SOD=89.4 IR2-SOD=91.9	IR1-SPropSOD=10.7 ** IR2-SPropSOD=15.6
Trial 3 ***(N=240)***	0.8% (2/240)	23.1% (55/238)	IR1-TLNum=2.1 * IR2-TLNum=2.4	IR1-SOD=57.1 ** IR2-SOD=75.4	IR1-SPropSOD=14.9 * IR2-SPropSOD=18.7
Trial 4 ***(N=276)***	0%	6.1% (17/276)	IR1-TLNum=2.4 ** IR2-TLNum=3.0	IR1-SOD=71.3 ** IR2-SOD=78.4	IR1-SPropSOD=15.3 ** IR2-SPropSOD=21.4
Trial 5 ***(N=378)***	0%	1.3% (5/378)	IR1-TLNum=2.4 ** IR2-TLNum=2.6	IR1-SOD=91.0 IR2-SOD=89.9	IR1-SPropSOD=10.4 IR2-SPropSOD=10.9
Average (N=1720)	0.1% [0.01; 0.4]	5.5% [4.5; 6.7]	2.34 [2.29; 2.40]	83.7 [81.6; 85.8]	14.4% [13.1; 15.6]

We documented double read features (displayed by column) for the five clinical trials (displayed by row). The two left-most columns display discrepancy features. Only for patients reported as having measurable diseases, the three right-most columns display the means of each reader's measurements (independent reader [IR]1 and 2), the p-value of the corresponding two-sample test of difference is indicated by asterisks: **, p<0.001; *, p<0.05; no asterisk means no statistically significant difference. The last row is the average overall measurements of both R1 and R2 with corresponding confidence intervals.

difference in SOD (The absolute difference of SOD divided by the average of the double reads SOD in %, DisSOD) reached more than 100% in all trials (Figure 3B).

In all trials, except Trial 5, average tumor burden was statistically significantly different between the two readers.

The average specific proportional SOD (DisLocSOD) ranged from 10.4% to 21.4%. In all trials, except Trial 5, DisLocSOD was statistically significantly different between the two readers.

The distribution of double read discrepancies in TL measurements are depicted in Figure 3 as violin plots, which confirm the findings in Table 4. Figure 3A shows that the DisTLnum was different across trials (median value for DisTLnum was 1 for Trial 1 and 5 and was 2 for the other trials). Per patient, the same number of TLs were selected in 49.6%, 46.1%, 27.3%, 33.4% and 53.0% of the Trial 1 to 5, respectively. These proportions were significantly different across trials (Marascuilo procedure, p<0.05). Figure 3B shows that for all trials, the DisSOD was higher than 100%, reaching 150% in all except Trial 5. The mean absolute DisSOD was 34.6%, 27.4, 41.2%, 40.6% and 26.8% in Trial 1 to 5, respectively. Therefore, three trials had an absolute difference significantly higher than 33%. Figure 3C shows that for all trials the average DisLocSOD value can reach 100%, which is confirmed by the proportion of patients for which readers targeted all TLs in totally different locations (DisLocTLAll) being different from zero for all trials (see Table 5). For Trials 1 to 5, 75% of their tumor burden had an average DisLocSOD value less than 22.2%, 15.8%, 19.9%, 26.3% and 12.2%, respectively. Therefore, the average DisLocSOD value of the third quartile in a trial can be two times higher than in another trial.

The distribution of reader's discrepancies according to disease location is summarized in Table 5. For all trials, a non-null proportion of patients had TLs selected in completely different organs, however, this proportion concerned less than 10% of patients. Overall, the readers targeted at least one different organ (TLs) in 43.5% of patients, ranging from 36.0% to 57.9% across trials. The organs with the highest risk of discrepancies were the lymph nodes (20.1%) and bones (12.2%) (see Figure 4). The discrepancies in detection of measurable disease occurred mainly in the lungs (19.6%).

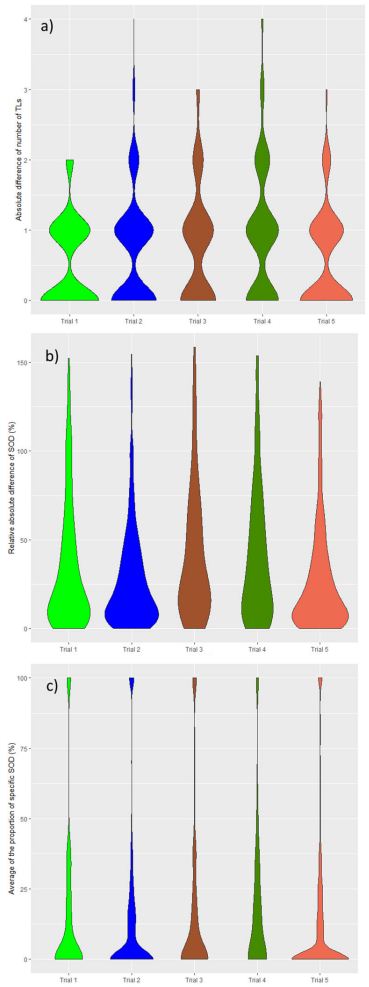


FIGURE 3
Distribution of double read measurement discrepancies. Distributions of double read measurement discrepancies are presented by color, representing each trial. Left to right; **(A)** the difference in the number of selected TLs (DisTLnum), **(B)** the proportional difference in SOD measurement (DisSOD), **(C)** the average of the inter-reader mean of readers (DisLocSOD).

TABLE 5 Distribution of discrepancies in disease locations.

Trial ID	Measurable disease (%)			Non-measurable disease (%)		Disease (%)
	DisLocTL	DisLocTLAll	TopDisLocTL	DisNTLLoc	TopDisLocNTL	TopDisLocDisease
Trial 1 (N=333)	42.9 (136/317)	7.5 (24/317)	Lung: 25.2 LN: 19.2 Pleura: 3.9 Bone: 3.3 Liver: 0.9	61.6 (205/333)	LN: 34.8 Lung: 31.8 Bone: 7.2 Pleura: 4.8 Liver: 3.6	LN: 21.9 Lung: 10.8 Bone: 8.1 Pleura: 6.3 Adrenal:2.7 Liver: 1.8
Trial 2 (N=493)	40.9 (201/491)	6.3 (31/491)	LN: 24.1 Lung: 13.2 Pleura: 3.1 Adrenal: 3.1	66.9 (330/493)	Lung: 32.0 LN: 30.0 Bone: 13.2 Pleura: 12.2 Liver: 3.0	LN: 25.8 Pleura: 12.6 Bone: 12.6 Lung: 8.1 Adrenal: 4.0
Trial 3 (N=240)	46.9 (86/183)	8.2 (15/183)	Lung: 22.1 LN: 19.6 Brain: 11.2 Pleura: 6.7	64.2 (154/240)	Lung: 30.4 LN: 21.7 Pleura: 15.4 Bone: 11.25 Liver: 6.6	LN: 10.6 Pleura: 15.8 Lung: 14.2 Bone: 11.2 Liver: 6.7
Trial 4 (N=276)	57.9 (150/259)	5.8 (15/259)	Lung: 27.2 LN: 21.7 Misc: 16.7 Brain: 9.8 Liver: 6.5	73.5 203/276	LN: 33.7. Lung: 28.9 Misc: 19.2 Bone: 18.1 Brain: 12.3	LN: 23.2 Misc: 21.0 Bone: 17.4 Lung: 13.4 Brain: 10.4
Trial 5 (N=378)	36 (133/373)	4 (15/373)	Lung: 14.0 LN: 12.4 Pleura: 6.6 Bone: 4.8	60.5% (229/378)	Lung: 38.6 LN: 20.6 Bone: 10.8 Pleura: 4.8	LN: 14.8 Bone: 12.4 Lung: 6.9 Pleura: 6.1
Overall (N= 1720)	43.5 [41.0; 45.9] (706/1623)	6.1 [5.0; 7.4] (100/1623)	LN: 19.6 Lung:19.2 Pleura: 4.0 Misc: 3.3	65.2 [62.9; 67.4] (1121/1720)	Lung: 32.7 LN: 28.3 Bone: 12.0 Pleura: 7.6	LN: 20.1 Bone: 12.2 Lung: 10.1 Pleura: 8.4

LN, Lymph node; TL, Target lesion; NTL, Non target lesion.
For the five clinical trials (displayed in rows) we computed: The proportion of patients for which readers targeted all their TLs in totally different locations (DisLocTLAll); the proportion of patients for which readers targeted at least one TL (DisTLLoc) or NTL (DisLocNTL) at a different disease location; the top proportion of discrepancies in TL (TopDisLocTL), NTL (TopDisLocNTL) and disease locations (TopDisLocDisease) (in % of patients concerned).

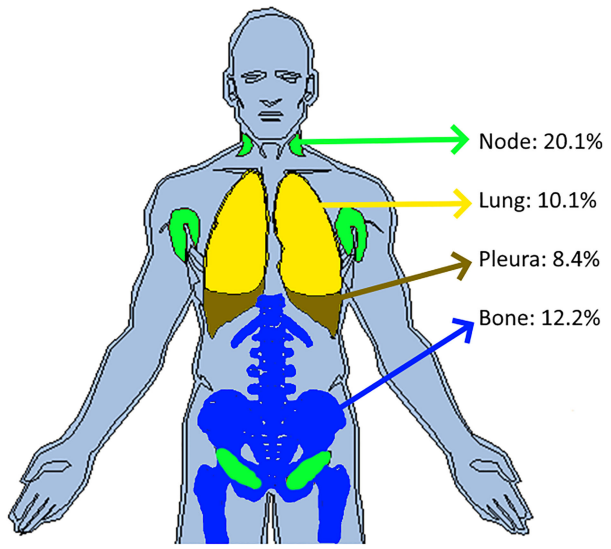


FIGURE 4
Inter-reader discrepancies in location of the disease. In pooling data from five comparable trials, we colored the top four organs where inter-reader discrepancy occurred in associating the corresponding proportion of patient concerned. Top four organs were the lymph nodes (20.1%), bones (12.2%), lungs (10.1%) and pleura (8.4%).

4.2 Reader's variability

Readers' selections and measurement data from the five trials were pooled and are summarized in Table 6.

Derived from the data in Table 6, and graphically confirmed in Figure 5, the distribution of readers' SODs and TLNum were statistically significantly different (Kruskal-Wallis, $p<0.001$). To be noted: for one patient in Trial 3, two readers (R1 and R6) did not find any measurable disease.

A 21 pair-wise comparison of the seven readers showed that 14, 13, 13, 10 and 10 pairs of readers (out of 21) significantly differed in the proportion of patients for whom diseases were selected in nodal, bones, infrequent (14) (see Annex A for definition), lung and pleura disease locations, respectively (Marascuilo procedure). These differences in proportion are depicted in Figure 6.

4.3 Trial's variability

The following results are for the subset of patients for whom the two readers documented the same disease locations (Table 7). On average per patient, the number of selected TLs ranged from 2.1 to

TABLE 6 Distribution of readers' selections across trials.

Readers ID	Nb Pat	Quantification		Proportion of patients in top discrepant disease locations				
		TLNum	SOD (mm)	Nodal (%)	Bone (%)	Lung (%)	Pleura (%)	Infrequent (%)
Reader 1	253	1.6 [1.5; 1.8]	47.5 [42.2; 52.7]	52.6 [46.2; 58.9]	33.6 [27.8; 39.8]	79.8 [74.3; 84.6]	12.3 [8.5; 16.9]	20.6 [15.7; 26.0]
Reader 2	304	1.9 [1.8; 2.0]	60.3 [55.7; 64.9]	70.4 [64.9; 75.5]	5.9 [3.5; 9.2]	93.1 [89.6; 95.7]	3.6 [1.8; 6.4]	9.5 [6.5; 13.4]
Reader 3	414	2.5 [2.4; 2.7]	80.9 [76.5; 85.3]	82.1 [78.1; 85.6]	34.3 [29.7; 39.1]	88.9 [85.4; 91.7]	0 [NA; NA]	28.5 [24.2; 33.1]
Reader 4	106	2.2 [2.0; 2.4]	74.4 [67.4; 81.3]	88.7 [81.0; 94.0]	5.7 [2.1; 11.9]	71.7 [62.1; 80.0]	5.7 [2.1; 11.9]	0.0 [NA; NA]
Reader 5	734	2.7 [2.6; 2.8]	94.0 [90.6; 97.3]	74.9 [71.6; 78.0]	18.0 [15.3; 20.9]	96.2 [94.5; 97.4]	3.7 [2.4; 5.3]	8.8 [6.9; 11.1]
Reader 6	365	1.7 [1.6; 1.8]	79.6 [74.7; 84.4]	46.6 [41.3; 51.8]	20.8 [16.8; 25.3]	94.5 [91.7; 96.7]	15.3 [11.8; 19.4]	12.9 [9.6; 16.7]
Reader 7	423	2.4 [2.3; 2.6]	77.8 [82.4; 73.1]	65.2 [60.5; 69.8]	25.1 [21.0; 29.5]	87.0 [83.4; 90.0]	0.2 [0.0; 1.3]	21.7 [17.9; 26.0]

SOD, Sum of Diamters; TLNum, Number of target lesion.
For seven readers (displayed in rows) involved in two or more trials, we reported in column 1) the number of assessed patients at baseline, 2) the average number of TLs selected in patients, 3) the average measured SOD, 4) the proportion of patients for whom nodal, bone, lung, pleura and infrequent disease were evaluated. Confidence intervals are provided in brackets.

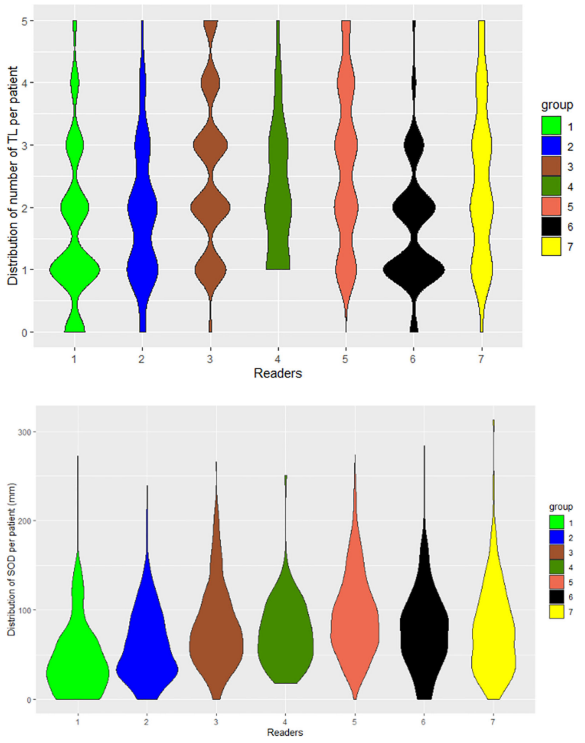


FIGURE 5
Distribution of readers' measurements. The seven readers are represented by violin plots for the number of selected TLs (TLNum) per patient (left) and SOD measurements (right).

FIGURE 6
Distribution of reader's selections. We display per reader, the proportion of patients (as percentage) in which the readers collected the disease in one of the top four discrepant locations (derived from Table 6) and in infrequent locations.

2.8, SOD ranged from 61.0 mm to 92.4 mm and average TL size ranged from 28.0 mm to 44.9 mm. Multiple comparisons showed that trials differed in average SOD, the number of selected TLs and the average size of selected TLs (Kruskal-Wallis, $p<0.0001$). The

TABLE 7 Trial features for double read with joint organ selection.

Trial ID	JointTLNum	JointTLSize (mm)	JointSOD (mm)	TopJointLocDisease
Trial 1 ***(N=184/333)***	2.2 [2.1; 2.4]	40.3 [37.4; 43.3]	82.8 [76.9; 88.8]	Lung=94% LN=73.9% Liver=7.1% Brain=2.2% Bone=1.1%
Trial 2 ***(N=224/493)***	2.2 [2.1; 2.4]	44.9 [42.4; 47.5]	92.4 [86.9; 97.9]	Lung=99.1% LN=69.2% Liver=5.8% Brain=2.2% Bone=6.7%
Trial 3 ***(N=102/240)***	2.1 [1.9; 2.4]	29.4 [26.6; 32.2]	61.0 [52.8; 69.1]	Lung=87.2% LN=45.1% Liver=2.9% Brain=22.5% Bone=21.6%
Trial 4 ***(N=100/276)***	2.8 [2.6; 3.1]	28 [25.8; 30.3]	78.4 [69.6; 87.1]	Lung=93% LN=75% Bone=36% Brain=22% Liver=14%
Trial 5 ***(N=223/378)***	2.6 [2.4; 2.7]	38.4 [36.2; 40.6]	91.3 [86.4; 96.2]	Lung=97.3% LN=83.8% Liver=7.6% Brain=2.7% Bone=6.3%

LN, Lymph node.
For each of the five trials (number of patients shown), we computed, per patient, the average number of TLs selected by the two readers (JointTLNum), the average TLs' diameter (JointTLSize) in mm, the average sum of TLs' diameter (JointSOD) in mm and the proportion of patients where the top five diseased locations were assessed. Averaged values are displayed with corresponding 95% confidence intervals.

proportion of patients having one of the top disease locations was statistically significantly different for lung between Trial 2 and 3 only. Multiple statistically significant differences were measured for all other disease locations (Marascuilo procedure, $p<0.05$). Figure 7 displays the top jointly selected diseased organs.

5 Discussion

We found that reader and trial were significant factors of variability (without interaction) for SOD and number of TLs estimated at baseline. This outcome allowed the development of the following discussion.

5.1 Variability between readers in disease detection and its measurability

The first variabilities in reading images lies in the identification of the disease. In clinical trials, it is key to include patients with relevant disease and a quantifiable tumor burden as required for proper treatment response assessment (11). For this reason, except for adjuvant evaluation setting, trials evaluating a treatment response endpoint usually require RECIST 1.1 "measurable" disease at baseline for eligibility, meaning that the radiologist should identify at least one TL in the recorded lesions at baseline. The eligibility evaluation is

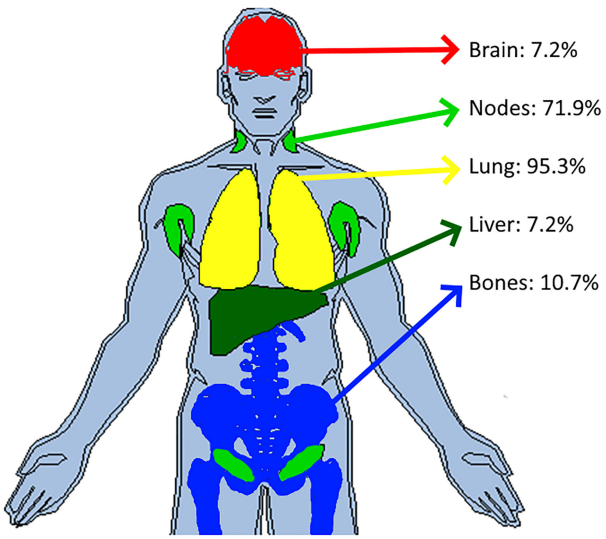


FIGURE 7
Top jointly selected organs. In pooling data from five comparable trials, we colored the top four jointly selected organs. With attached proportion of patient concerned, top jointly selected organs were the lungs (95.3%), lymph nodes (71.9%), bones (10.7%), brain (7.2%) and liver (7.2%).

usually performed on site before submitting baseline examinations of screened patients for central review. To mitigate the risks of disagreement between the screening site and central review, another option is to perform eligibility evaluation centrally by involving three readers. The two concordant readers are then kept for the RECIST 1.1 assessment. Central eligibility review was performed for Trial 2, which explains the low rate of discrepancies at baseline in double reads. It is interesting to note that if Trial 2 is excluded, there is still a low average discrepancy rate (<1% regarding non-detection of disease and <10% regarding measurability status of the disease at baseline). However, the magnitude of discrepancies fluctuates widely across the trials with measurable *vs*. non-measurable disease discrepancies reaching up to 23% in Trial 3. This may be partly explained by Trial 3 having patients with the smallest tumor burden average reported by the readers.

5.2 Variability between readers in disease burden measurements

To be representative of metastatic disease, baseline selection evaluations may involve quantifying tumor burden by number of TLs and related SOD.

For, respectively, the number of TLs and SOD, three and four out of the five trials, had significant double reads differences ($p<0.001$) (Table 4).

Approximately 50% of patients had a different number of TLs recorded during double reads. This value is similar to the 59% reported by Kuhl et al. (15). Regarding the distribution of this quantitative discrepancy, violin plots showed readers had most often recorded one (rarely two) TL more or less than the paired reader. The discrepancy in SOD was in average higher than 25%, with violin plots showing a large range of variation, up to 150% difference for all trials. There are several reasons for readers' measurement discrepancies and our results shows a larger variability on lung organ in respect to lymph-node. Indeed, lung metastasis delineation can be subjective if the reader needs to separate adjacent atelectasis.

The magnitude of differences in SOD raises questions regarding the impact this has on follow-up and RECIST 1.1 response thresholds (11). We know from Sharma et al. (16) that the variability in SOD at baseline is a risk for discrepant responses but, while threshold values for detecting significant longitudinal SOD changes (in follow up) are proposed (17), none are proposed for critical values for differences in SOD with double reads at baseline. Darkeh et al. (18) showed the impact of the number of TLs selected on discordances. If we assume a direct correlation between variability in the number of selected TLs and SOD, the conclusions of Darkeh et al. and Sharma et al. are consistent. Baseline SOD is also reported as an independent prognostic biomarker, however, the magnitude of variability questions the reliability its use (19).

In documenting the extent of the SOD variability, we confirmed previous works, notably the permissiveness of RECIST (20) in the selection of lesions to include as TLs. For selecting a TL, its size is not the only criteria, conspicuity, vicinity and the number of other candidate TLs are some other numerous factors that are left to readers appreciation. This subjective choice can easily explain a 150% discrepancy in the SOD between two observers.

5.3 Variability in assessments of disease distribution

As metastatic patients have multi-organ disease, RECIST 1.1 recommends a representative selection of TLs across all involved organs to capture the extent of the disease. In this analysis, we introduced a new quantitative feature (DisLocDisease) to represent the proportion of disease burden measured in organs by only one of the two paired readers. At baseline, due to the central review setting, historical data are censored; therefore readers may subjectively select, more or less equivocal lesions, such as the small liver nodule in Figure 2 (10).

On average, the DisLocSOD represented up to 20% of the SOD per trial per reader. The violin plots of DisLocSOD displayed bimodal distributions where the second local maxima in the probability density function at 100% corresponds to the 6% of the patients assessed with zero common disease locations during double reading (DisLocTL). This discrepancy in disease location/measurement mainly involved assessments of lung and lymph node disease. The latter can be explained due to the "size" related threshold (short axis >=1.5 cm) of a measurable adenopathy according to RECIST 1.1. Specifically for our indication of interest, in the mediastinum of smoker patients, it is not uncommon to observe centimetric nodes which are otherwise non-specific (21) and not captured as TLs by the readers.

For double read assessments of NTLs, the bones were identified as the third most at risk location for discrepant metastatic disease identification. Indeed, bone metastases are almost always recorded as NTL as blastic lesions are truly non-measurable and even when they are measurable, RECIST 1.1 rules consider them as a secondary choice. We suspect that identification of bone metastasis demonstrates a variability during double reads for conspicuity reasons. The same detection errors have been documented during follow-up (8).

The variability in the selection of the diseased organ was greater for the NTL than for the TL, in 40% and 60% of patients, respectively. This is concordant with the literature (22). This greater variability demonstrates that the NTL category contains more ambiguous findings with respect to the TL lesions, which agrees with the literature.

Even when readers consider the same organ, classifying tumors as TL or NTL is of importance as some studies (15, 23) showed possible difference in malignancy, which consequently led to discordance in the evaluation of treatment response (22). A limitation in the RECIST 1.1 rules (20) may explain the origin of such discordances because, unlike the TL, the NTL category is designed to record the smallest measurable lesions and non-measurable lesions under 1 **cm** (or short axis < 1.5 **cm** for lymph node). Typically, a lung micro nodule may be considered as NTL by one reader while the paired reader may not consider the finding significant (see example in Figure 8) which could potentially lead to a discrepancy in evaluating the extent of disease. Indeed, during the follow-up, some of the differences between readers in capturing disease

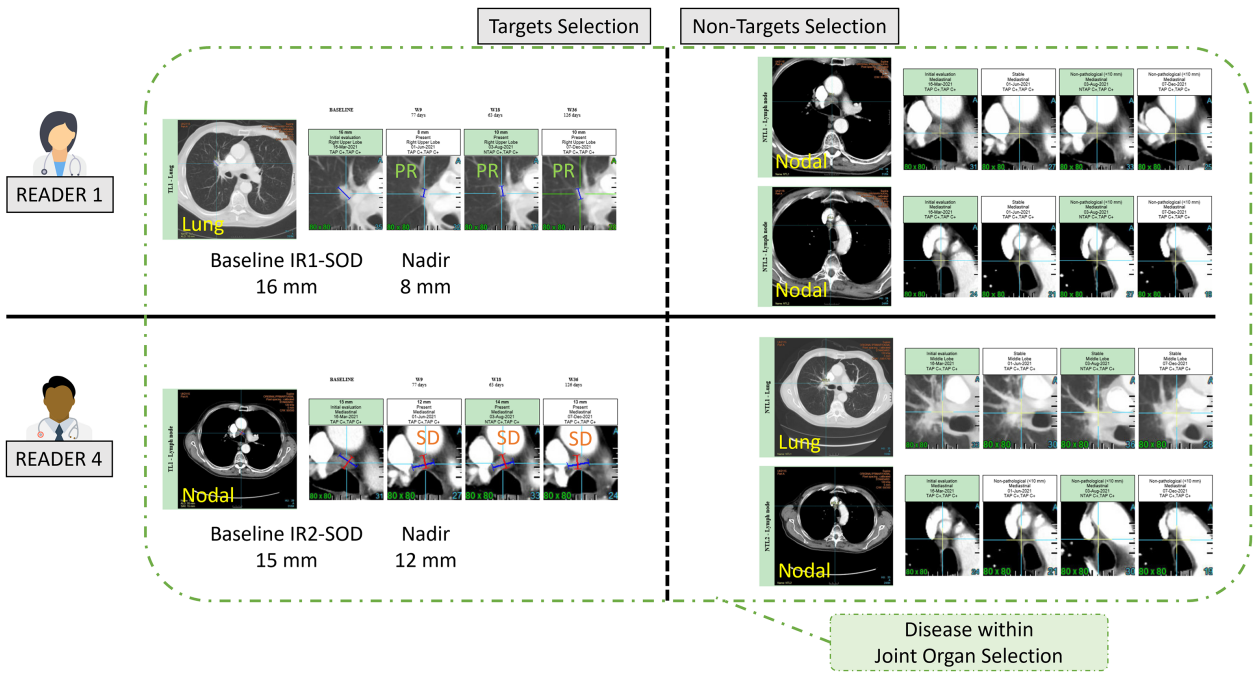


FIGURE 8
Example of inter-reader variability in classifying target and non-target lesions. In this example two readers (R1 and R4) selected the same lesions with different classification of target and non-target leading to discrepant responses. This discrepancy hinges on the measurement of the lymph node (15mm for Reader 4, and unstated for Reader 1) as well as the subjective opinion of Readers 1 and 4 about the measurability of the lung lesion.

progression have been explained by possible dissociation of the response i.e., when tumors selected from different disease locations respond differently to treatment (24, 25). In trials evaluating the efficacy of immunotherapy, a dissociated response has been reported in 30% of patients for our indication (26). Discrepancies in baseline selection may increase the risk of discrepancies in double read evaluations if the patient experiences a dissociated response (27, 28).

5.4 Typology of radiologist readings

One of the challenges of BICR monitoring is to identify an "outlier" radiologist likely to increase the rate of discrepancies. The preferred key performance indicator (KPI) for this is based on follow-up radiological response timepoint (29). However, the performance of a reader who tends to under- or miss-diagnose the disease as early as baseline, could also be represented by specific performance indicators.

According to RECIST 1.1 criteria, the radiologist should collect up to five TLs with a maximum of two selected per organ. Our analysis showed that radiologists tended to select less than three TLs on average. This suggests that the representation of measurable disease of metastatic lung cancer is predominantly unifocal or bifocal. This is confirmed by the analysis in joint organ selection where the average number of TLs selected by both readers was also less than three for all trials (Table 7).

In addition, the violin plots in Figure 5 shows that two radiologists (R1 and R6) tended to collect fewer TLs than their colleagues. In contrast, we observed that two readers (R3 and R5) tended to systematically select more TLs than other readers. As readers were involved in several studies, this demonstrates the existence of a specific reading pattern with a potential impact on double reading. The impact of a reader's behavior on SOD is not straightforward. The distribution of tumor burden seemed globally similar for all the readers except two (R1 and R2) who had lower SOD measurements.

Comparing the selection profile of each reader helps to highlight the disease locations that are the most difficult to characterize, and which lead to greater inter-reader variability. We found there was a greater dispersion between readers on the recording of metastatic bone locations that are sometimes difficult to see, forgotten or ambiguous. Infrequent and therefore unexpected locations are also not surprisingly more dispersed and probably linked to their detection rate. The variability of reader's evaluations in the context of lymph node disease remains important despite the greater frequency of this location. Two readers (R1 and R6) tended to record less disease in this location. This may again suggest the existence of a reading pattern specific to a radiologist.

Geijer et al. (30) documented the variability between two readers when they have differing experience, background, or interpretation of guidelines. As stated by Schmid et al. (5), "The greatest contributing factor of inter-reader variability originates from a radiologist's own expertise". In our study we observed a large inter-reader variability

that would substantially contribute to the magnitude of the double read variability. The relationship between inter-reader and double read variabilities would require further investigation with the aim to optimize reader's pairing.

5.5 Homogeneity of trials

We analyzed inter-trial homogeneity in a subset of patients for whom both readers selected the same disease locations for a more robust approach. Our analysis showed that disease presentation across trials differed significantly in terms of average size of TLs (ranging 28mm to 45mm) and tumor burden indicators. We reached same conclusions in focusing on the subgroup of lung and nodal TLs which are the most frequent targeted tumor locations. This variability is therefore a limitation of generalizing our KPI with the aim of benchmarking "comparable" metastatic non-small cell lung cancer (NSCLC) trials.

However, in concordant patients, the distribution of metastatic disease was comparable to the literature; readers primarily targeted the lung, lymph nodes, and more rarely, the liver and bone (14). The frequency of lymph node and bone disease was the most variable disease location across the trials, even when concordant evaluations were considered. Again, this shows that despite selecting patients according to relatively similar criteria, the presentation of the disease can differ greatly and may partly explain the differing rates of discordance in double readings found across the available literature for the same indication. The limitations in generalizing results across similar studies are related to the well-documented representativeness issues of the study population (31).

Although we endeavored to evaluate "comparable" trials, the patient population at each site may have had slightly varying characteristics (e.g., stage of the disease, treatment line). Our measurements showed inter-trial differences that can partially be explained by variable inclusion criteria. Liu et al. (32) did show that broadening restrictive inclusion criteria in advanced NSCLC trials had little impact on the trial hazard ratios, but little remains known about the impact of inclusion criteria and readers reliability (4).

5.6 Limitations

Firstly, our analysis of the reader~trial interaction was a partial analysis. We measured the variability of only two features (SOD and TLNum) as no recognized statistics were available to analyze the interaction within the measuring proportion (e.g. TopDisLocDisease). We were also limited by our data as all readers were not involved in all five trials and not all readers measurements were applicable to the different steps of our analysis.

Secondly, our raw data were blinded from tumor coordinates therefore, unlike Kuhl et al. (22), it was impossible to identify when the exact same finding was selected by both readers. The highest level of localization was at organ level.

Thirdly, as we were blinded from randomization, we were not able to refine our analysis by treatment or control. All trials included consisted of two sub-cohorts.

Fourthly, our two-way analysis considered only two features; the average number of TLs and SOD.

Finally, the analysis focused on a specific metastatic cancer indication thus limiting the generalization to another type of primary cancer. Indeed, our variability root cause analysis demonstrates that variability depends greatly on the metastasis locations known to be related to the primary cancer.

5.7 Perspectives

We proposed an innovative method that can be applied to clinical trials that use RECIST 1.1 to explore the initial disease presentation assessment and the variability of these assessments.

Bearing in mind the caution against generalization, our baseline variability analysis may help with detecting a deviation from an expected variability rate and lead to early investigation into the origin of the deviation. In the context of BICR, our baseline variability analysis can contribute to the quality control of reads. The double and single read-derived features should be further investigated for this purpose.

The initial investigation should consist of evaluating the correlation between the variabilities of feature values at baseline and at the corresponding therapeutic response evaluation timepoint. A significant correlation would allow to build a predictive model for the reliability of the evaluation using minimal data, therefore triggering early corrective actions or adaptations to trial sample size.

The second investigation should focus on features derived from single radiologist assessments. The existence of patterns attached to a set of radiologists would allow optimal pairing of radiologists for double reading.

The last application of our features applies to the core annotations (annotations performed on the same diseases by double reads); to confirm that clinical trials expected to be "similar" really are.

With the emerging use of synthetic arms (33), it has become very attractive to pool several "similar" control arms together to design a single synthetic larger one.

Ultimately, considering variabilities and discrepancies only as "an event to avoid" is probably not an optimal strategy. The baseline variability assessment is not purely noise. From a patient benefit perspective, a second opinion still means a higher chance of a correct diagnosis. The discrepancy event with the use of the proposed framework of analysis can help us detect both pseudo-lesions (34) at the baseline disease assessment and real metastasis missed by single reading.

As we can expect more information from combining readers' annotations into logical sets and applying the advanced algebra of our features, it may be possible to detect dissociated responses or to improve our understanding of the disease prognostic and drug mechanism of action.

6 Conclusion

Variability in baseline disease selection is known to be one of the major contributors to RECIST 1.1 inter-reader variability and is largely documented in previous papers.

Our analysis focused on the discrepancy between radiologists in disease selection. We provided an innovative method for quantifying discrepant tumor burden evaluations and for qualifying discrepant tumor distribution evaluations.

Considering our dataset extracted from five trials in metastatic NSCLC, we found that approximately 15% of patient tumor burden was measured in discrepant locations. The locations with the highest risk of discrepancies in disease identification were the lymph nodes and bone metastasis.

Our figures showed a greater uncertainty on the selection of the disease in the NTL category compared to the TL category.

The baseline lesion selection criteria in the RECIST guidelines leaves room for subjective assessments, potentially causing some of the observed differences in the chosen target or non-target lesions.

By analyzing the reader's baseline assessments, we observed the existence of a reader's specific pattern of assessment. This explains in part the observed inter-reader variability and could lead to possible pairing optimization to decrease discrepancies between readers.

In addition, we demonstrated that even though lung trials may be comparable in terms of the patient population, the indication, inclusion criteria and the primary tumor, that does not necessarily ensure their comparability in terms of disease presentation. Therefore, literature-based benchmarks for discrepancy KPIs should be used with caution.

Data availability statement

The data analyzed in this study is subject to the following licenses/restrictions: Datasets are clinical trials sponsors properties. Requests to access these datasets should be directed to hubert.beaumont@mediantechnologies.com.

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

HB: Conceptualization, methodology, data curation, formal analysis, original draft, writing, review, and editing. AI: Conceptualization, methodology, formal analysis, project administration, original draft, review, and editing. All authors contributed to the article and approved the submitted version.

Conflict of interest

The authors of this manuscript, HB and AI, declare a relationship with the following company: Median Technologies.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2023.988784/full#supplementary-material

References

1. FDA, U.S. Department of Health and Human Services, Center for Drug Evaluation and Research and Center for Biologics Evaluations and Research. *Guidance for industry developing medical imaging drug and biologic products. part 3: Design, analysis, and interpretation of clinical studies*. (2004).

2. FDA, US Department of Health and Human Services, Center for Drug Evaluation and Research and Center for Biologics Evaluations and Research. *Clinical trial imaging endpoints process standards guidance for industry draft*. (2015).

3. US Food and Drug Administration. *Clinical trial imaging endpoint process standards: Guidance for industry*. (2018). pp. 18571–3.

4. Beaumont H, Iannessi A, Wang Y, Voyton CM, Cillario J, Liu Y. Blinded independent central review (BICR) in new therapeutic lung cancer trials. *Cancers (Basel)* (2021) 13:4533. doi: 10.3390/cancers13184533

5. Schmid AM, Raunig DL, Miller CG, Walovitch RC, Ford RW, O'Connor M, et al. Radiologists and clinical trials: Part 1 the truth about reader disagreements. *Ther Innov Regul Sci* (2021). doi: 10.1007/s43441-021-00316-6

6. Raunig DL, McShane LM, Pennello G, Gatsonis C, Carson PL, Voyvodic JT, et al. Quantitative imaging biomarkers: A review of statistical methods for technical performance assessment. *Stat Methods Med Res* (2015) 24:27–67. doi: 10.1177/0962280214537344

7. Zhao B, Tan Y, Bell D, Marley S. Intra-and inter-reader variability in uni-dimensional, bi-dimensional, and volumetric measurements of solid tumors on CT scans reconstructed at different slice intervals. *Eur J* (2013) 82:1–22. doi: 10.1016/j.ejrad.2013.02.018.Exploring

8. Beaumont H, Evans TL, Klifa C, Guermazi A, Hong SR, Chadjaa M, et al. Discrepancies of assessments in a RECIST 1.1 phase II clinical trial – association between adjudication rate and variability in images and tumors selection. *Cancer Imaging* (2018) 18:50. doi: 10.1186/s40644-018-0186-0

9. Fournier L, de Geus-Oei LF, Regge D, Oprea-Lager DE, D'Anastasi M, Bidaut L, et al. Twenty years on: RECIST as a biomarker of response in solid tumours an EORTC imaging group – ESOI joint paper. *Front Oncol* (2022) 11:800547. doi: 10.3389/fonc.2021.800547

10. Iannessi A, Beaumont H, Liu Y, Bertrand AS. RECIST 1.1 and lesion selection: How to deal with ambiguity at baseline? *Insights Imaging* (2021) 12. doi: 10.1186/s13244-021-00976-w

11. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). *Eur J Cancer* (2009) 45:228–47. doi: 10.1016/j.ejca.2008.10.026

12. Ford R, O' Neal M, Moskowitz S, Fraunberger J. Adjudication rates between readers in blinded independent central review of oncology studies. *J Clin Trials* (2016) 06. doi: 10.4172/2167-0870.1000289

13. Marascuilo LA. Extensions of the significance test for one-parameter signal detection hypotheses. *Psychometrika* (1970) 35:237–43. doi: 10.1007/BF02291265

14. Niu FY, Zhou Q, Yang JJ, Zhong WZ, Chen ZH, Deng W, et al. Distribution and prognosis of uncommon metastases from non-small cell lung cancer. *BMC Cancer* (2016) 16:1. doi: 10.1186/s12885-016-2169-5

15. Kuhl CK. RECIST needs revision: A wake-up call for radiologists. *Radiology* (2019) 292:110–1. doi: 10.1148/radiol.2019190785

16. Sharma M, Singareddy A, Bajpai S, Narang J, O'Connor M, Jarecha R. To determine correlation of inter reader variability in sum of diameters using RECIST 1.1 with end point assessment in lung cancer. *J Clin Oncol* (2021) 39:e13557–7. doi: 10.1200/JCO.2021.39.15_suppl.e13557

17. Yoon SH, Kim KW, Goo JM, Kim D-W, Hahn S. Observer variability in RECIST-based tumour burden measurements: A meta-analysis. *Eur J Cancer* (2016) 53:5–15. doi: 10.1016/j.ejca.2015.10.014

18. Darkeh MHSE, Suzuki C, Torkzad MR. The minimum number of target lesions that need to be measured to be representative of the total number of target lesions (according to RECIST). *Br J Radiol* (2009) 82:681–6. doi: 10.1259/bjr/72829563

19. Machida N, Yoshino T, Boku N, Hironaka S, Onozawa Y, Fukutomi A, et al. Impact of baseline sum of longest diameter in target lesions by RECIST on survival of patients with metastatic colorectal cancer. *Jpn J Clin Oncol* (2008) 38:689–94. doi: 10.1093/jjco/hyn086

20. Morse B, Jeong D, Ihnat G, Silva AC. Pearls and pitfalls of response evaluation criteria in solid tumors (RECIST) v1.1 non-target lesion assessment. *Abdom Radiol* (2019) 44:766–74. doi: 10.1007/s00261-018-1752-4

21. Libshitz HI, McKenna RJ. Mediastinal lymph node size in lung cancer. *AJR Am J Roentgenol* (1984) 143:715–8. doi: 10.2214/ajr.143.4.715

22. Kuhl CK, Alparslan Y, Schmoee J, Sequeira B, Keulers A, Brümmendorf TH, et al. Validity of RECIST version 1.1 for response assessment in metastatic cancer: A prospective, multireader study. *Radiology* (2019) 290:349–56. doi: 10.1148/radiol.2018180648

23. Coy HJ, Douek ML, Ruchalski K, Kim HJ, Gutierrez A, Patel M, et al. Components of radiologic progressive disease defined by RECIST 1.1 in patients with metastatic clear cell renal cell carcinoma. *Radiology* (2019) 292:103–9. doi: 10.1148/radiol.2019182922

24. Tozuka T, Kitazono S, Sakamoto H, Yoshida H, Amino Y, Uematsu S, et al. Dissociated responses at initial computed tomography evaluation is a good prognostic factor in non-small cell lung cancer patients treated with anti-programmed cell death-1/ligand 1 inhibitors. *BMC Cancer* (2020) 20:1–8. doi: 10.1186/s12885-020-6704-z

25. Chen DT, Chan W, Thompson ZJ, Thapa R, Beg AA, Saltos AN, et al. Utilization of target lesion heterogeneity for treatment efficacy assessment in late stage lung cancer. *PloS One* (2021) 16:1–15. doi: 10.1371/journal.pone.0252041

26. Humbert O, Chardin D. Dissociated response in metastatic cancer: An atypical pattern brought into the spotlight with immunotherapy. *Front Oncol* (2020) 10:566297. doi: 10.3389/fonc.2020.566297

27. Beaumont H, Faye N, Iannessi A, Chamorey E, Kliffa C, Hsieh C-Y. Differences in sensitivity to new therapies between primary and metastatic breast cancer: A need to stratify the tumor response? (2020), 1–21. doi: 10.21203/rs.3.rs-92580/v1

28. Carter BW, Bhosale PR, Yang WT. Immunotherapy and the role of imaging. *Cancer* (2018) 124:2906–22. doi: 10.1002/cncr.31349

29. Raunig DL, Schmid AM, Miller CG, Walovitch RC, O'Connor M, Noever K, et al. Radiologists and clinical trials: Part 2: Practical statistical methods for understanding and monitoring independent reader performance. *Ther Innov Regul Sci* (2021) 55:1122–38. doi: 10.1007/s43441-021-00317-5

30. Geijer H, Geijer M. Added value of double reading in diagnostic radiology, a systematic review. *Insights Imaging* (2018) 9:287–301. doi: 10.1007/s13244-018-0599-0

31. Kennedy-Martin T, Curtis S, Faries D, Robinson S, Johnston J. A literature review on the representativeness of randomized controlled trial samples and implications for the external validity of trial results. *Trials* (2015) 16:1–14. doi: 10.1186/s13063-015-1023-4

32. Liu R, Rizzo S, Whipple S, Pal N, Pineda AL, Lu M, et al. Evaluating eligibility criteria of oncology trials using real-world data and AI. *Nature* (2021) 592:629–33. doi: 10.1038/s41586-021-03430-5

33. Thorlund K, Dron L, Park JJH, Mills EJ. Synthetic and external controls in clinical trials – a primer for researchers. *Clin Epidemiol* (2020) 12:457–67. doi: 10.2147/CLEP.S242097

34. Teslenko I, Belotserkovsky M. Common pitfalls of RECIST 1.1 application in clinical trials. *Eur J Cancer* (2015) 51:S132. doi: 10.1016/S0959-8049(16)30387-2

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 21 March 2023
DOI 10.3389/fonc.2023.1115485

OPEN ACCESS

EDITED BY
Bahram Mohajer,
Johns Hopkins Medicine, United States

REVIEWED BY
Deepak Kumar,
Shoolini University, India
Asli Suner,
Ege University, Türkiye

*CORRESPONDENCE
Rui Tian
rui.tian0312@gmail.com

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

RECEIVED 04 December 2022
ACCEPTED 27 February 2023
PUBLISHED 21 March 2023

CITATION
Zhang X, Lu Y, Huang K, Pan Q, Jia Y, Cui B, Yin P, Li J, Ju J, Fan X and Tian R (2023) The synergized diagnostic value of VTQ with chemokine CXCL13 in lung tumors. *Front. Oncol.* 13:1115485. doi: 10.3389/fonc.2023.1115485

The synergized diagnostic value of VTQ with chemokine CXCL13 in lung tumors

Xu Zhang[1], Yejian Lu[2], Kenan Huang[2], Qingfang Pan[2], Youchao Jia[3], Baoshuan Cui[2], Peipei Yin[2], Jianhui Li[1], Junping Ju[4], Xiangyu Fan[5] and Rui Tian[2]*

[1]Department of Ultrasound, Affiliated Hospital of Hebei University, Baoding, Hebei, China, [2]Department of Oncology, Hospital of the People's Liberation Army: 82nd Group Army, Baoding, China, [3]Department of Oncology, Affiliated Hospital of Hebei University, Baoding, Hebei, China, [4]Department of Radiology, Affiliated Hospital of Hebei University, Baoding, Hebei, China, [5]Department of Pathology, Affiliated Hospital of Hebei University, Baoding, Hebei, China

Virtual Touch Tissue Quantification (VTQ) offers several advantages in the diagnosis of various lung diseases. Chemokine expression levels, such as CXCL13, play a vital role in the occurrence and development of tumors and aid in the diagnosis process. The purpose of this study was to evaluate the combined value of VTQ and changes in CXCL13 expression levels for the diagnosis of lung tumors. A total of 60 patients with thoracic nodules and pleural effusion were included, with 30 of them having malignant pleural effusion (based on pathology) and the remaining 30 having benign thoracic nodules and pleural effusion. The relative expression level of CXCL13 was measured in the collected pleural effusions using Enzyme-Linked Immunosorbent Assay (ELISA). The relationship between CXCL13 expression levels and various clinical features was analyzed. A Receiver Operating Characteristic (ROC) curve analysis was conducted on the VTQ results and relative expression levels of CXCL13, and the areas under the curve, critical values, sensitivity, and specificity were calculated. Multivariate analysis incorporating multiple indicators was performed to determine the accuracy of lung tumor diagnosis. The results showed that the expression levels of CXCL13 and VTQ were significantly higher in the lung cancer group compared to the control group ($P < 0.05$). In the Non-Small Cell Lung Cancer (NSCLC) group, CXCL13 expression levels increased with later TNM staging and poorer tumor differentiation. The expression level of CXCL13 in adenocarcinoma was higher than that in squamous cell carcinoma. The ROC curve analysis revealed that CXCL13 had an area under the curve (AUC) of 0.74 (0.61, 0.86) with an optimal cut-off value of 777.82 pg/ml for diagnosing lung tumors. The ROC curve analysis of VTQ showed an AUC of 0.67 (0.53, 0.82) with a sensitivity of 60.0% and a specificity of 83.3%, and an optimal diagnostic cut-off of 3.33 m/s. The combination of CXCL13 and VTQ for diagnosing thoracic tumors had an AUC of 0.842 (0.74, 0.94), which was significantly higher than either factor alone. The results of the study demonstrate the strong potential of combining VTQ results with chemokine CXCL13 expression levels for lung tumor diagnosis. Additionally, the findings suggest that elevated relative expression of CXCL13 in

cases of malignant pleural effusion caused by non-small cell lung cancer may indicate a poor prognosis. This provides promising potential for using CXCL13 as a screening tool and prognostic indicator for patients with advanced lung cancer complicated by malignant pleural effusion.

KEYWORDS

ARFI, VTQ, lung tumor, malignant pleural effusion, CXCL13

Introduction

Lung cancer is one of the leading causes of cancer-related deaths worldwide. According to the World Health Organization, lung cancer accounts for approximately 1.76 million deaths each year. The incidence of lung cancer is higher in men than in women, and it is more commonly diagnosed in older adults. Tobacco use is the leading cause of lung cancer, with long-term exposure to tobacco smoke increasing the risk of developing the disease. Other risk factors for lung cancer include exposure to air pollution, radon, asbestos, and certain genetic mutations. The early detection of lung cancer is crucial for improving survival rates, but due to the lack of symptoms in early stages and the difficulty in diagnosing the disease, the survival rate for lung cancer remains low.

Lung cancer is posing a serious threat to human health, especially in China (1). Unfortunately, patients with early-stage lung cancer often show no distinctive symptoms and are only detected when they reach an advanced stage, leading to a short survival time (1). The gold standard for diagnosing malignant pleural effusion is histopathological examination, but the low probability of finding tumor cells and the high risk of invasive examination for elderly patients with underlying diseases make the diagnosis challenging. In comparison, the detection of tumor markers is a convenient and fast alternative that can not only reflect the occurrence and development of tumors but also assist in diagnosis and prognosis.

Recent research has shown that chemokines play a critical role in the development of tumors (2–7). However, as a single diagnostic indicator, chemokines have limitations in terms of sensitivity and specificity, and false negative and false positive results may occur in some patients. Thus, it is necessary to combine chemokine analysis with other auxiliary examinations for joint diagnosis.

In recent years, lung ultrasound has gained increasing popularity in clinical practice due to the continuous improvement of ultrasound diagnostic equipment. It is easy to use, affordable, reproducible, and does not involve radiation, making it particularly suitable for emergency patients who cannot undergo invasive operations (8–14). Virtual touch tissue quantification (VTQ) has been confirmed to have several advantages in diagnosing various lung diseases.

In light of these developments, this study aims to investigate the diagnostic value of combining VTQ technology and chemokine analysis for lung tumor diagnosis. The researchers hope to find evidence that combining the results of VTQ and chemokine CXCL13 expression could have a high potential value in lung tumor diagnosis. The results also suggest that a higher relative expression of CXCL13 in malignant pleural effusion caused by non-small cell lung cancer may indicate a poor prognosis and could be used as an indicator for screening and prognosis in patients with advanced lung cancer complicated by malignant pleural effusion.

Materials and methods

Priori analysis

The study aimed at evaluating the diagnostic accuracy of combining Virtual Touch Tissue Quantification (VTQ) results with chemokine expression levels of CXCL13 in the diagnosis of lung tumors. To ensure that the study was adequately powered to detect the hypothesized effects, *a priori* power analysis was performed. The priori power analysis involved estimating the effect size, determining the sample size required to achieve a desired level of statistical power, and selecting the appropriate statistical test.

The sample size was determined based on the number of patients with malignant pleural effusion (n=30) and the number of patients with benign thoracic nodules and pleural effusion (n=30). The desired level of statistical power was set at 0.8, meaning that the study had an 80% chance of detecting a significant difference between the two groups if such a difference existed. The effect size was estimated based on previous research studies that investigated the diagnostic accuracy of VTQ and CXCL13 expression levels in lung tumor diagnosis.

The statistical test used for the priori power analysis was the two-sample t-test, which is appropriate for comparing the means of two independent groups. The analysis was performed using a statistical software program, such as R or SAS, which allows for the calculation of sample size and power based on the desired effect size, significance level, and level of power.

The results of the priori power analysis determined the sample size required to achieve the desired level of statistical power and provided confidence in the study's ability to detect a significant difference between the two groups if such a difference existed. By conducting the priori power analysis, the researchers were able to ensure that the study was designed to have sufficient statistical power to detect meaningful differences and reduce the risk of type II errors, or false negatives.

Research objects

Thirty pleural effusion specimens were collected from newly diagnosed patients with pulmonary malignant tumors and pleural effusion. These patients were admitted to the 82nd Army Group Military Hospital and the Affiliated Hospital of Hebei University from November 2021 to July 2022 and all of the specimens showed the presence of tumor cells. The collection process was performed under sterile conditions, where 10 ml of pleural effusion was collected from each patient. The supernatant was then separated and collected after 10 minutes of centrifugation at 2000 rpm and 4°C. The collected supernatant was stored in -80°C in 1.5 ml eppendorf tubes until further treatment.

The pathological types of the specimens were classified according to the World Health Organization's histological classification standards for malignant tumors and the staging was performed following the UICC eighth edition TNM staging system. The inclusion criteria for the specimens included the following:

1. Patients' personal information such as name, gender, age, marital status, race, smoking history, and family disease history
2. Patients who have not received any form of anti-tumor therapy such as radiotherapy, chemotherapy, immunotherapy, or targeted therapy before admission
3. 30 additional specimens, used as the control group, were collected from patients with benign lung tumors and pleural effusion who had no history of malignancy
4. All participants signed an informed consent form approved by the ethics committee of the hospital. The general characteristics of the participants are listed in Table 1.

Main instruments and reagents

The following laboratory equipment and supplies were used in this study:

- Siemens ACUSON S2000 Color Ultrasonic Diagnostic Instrument
- US Biotek Microplate Reader (Gen5)
- US Thermo Tabletop Low Temperature High-Speed Centrifuge (CENTRIFUGE PK 121R)
- Thermo -80°C Freezer
- AISITE SPX-150BIII Biochemistry Incubator
- Eppendorf Pipettes
- R&D Systems Inc. Human B-Lymphocyte Chemoattractant (BLC-1/CXCL13) ELISA Kit

Experimental methods

The Siemens ACUSON S2000 color ultrasonic diagnostic instrument, equipped with VTQ (Virtual Touch Quantification) technology, was utilized in the investigation of lung tumors. A convex array probe was selected for the experiment, and the frequency was set to a range of 3.5-5 MHz. The patients were asked to remain seated in a quiet environment, and the VTQ function was activated to capture the relevant data.

To thoroughly examine the lung tumors, various parameters such as size, margins, internal parenchymal echoes, and blood flow were analyzed using the VTQ technology. A total of five successful measurements were taken, and to reduce the impact of outliers, the maximum and minimum values were removed. The average value of the remaining measurements was then calculated and recorded for statistical analysis.

In addition to the VTQ results, the expression level of chemokine CXCL13 was determined through ELISA (Enzyme-Linked Immunosorbent Assay). To obtain an accurate measurement, the pleural effusion specimens were taken out from the -80 oC freezer, thawed to room temperature, and analyzed. The relative expression of chemokine CXCL13 was determined through ELISA, which provides a quantitative measurement of the protein's concentration in the sample. Figure 1 summarizes the workflow.

TABLE 1 General characteristics of the participants.

Characteristics		Lung cancer group (n)	Control group (n)
Gender	Male	17 (56.7%)	18 (60.0%)
	Female	13 (43.3%)	12 (40.0%)
Age	≥60	16 (53.3%)	20 (66.7%)
	<60	14 (46.7%)	10 (33.3%)
Smoking history	Yes	21 (70.0%)	17 (56.7%)
	No	9 (30.0%)	13 (43.3%)

There was no significant statistical difference observed in the gender, age, and smoking history of the participants (P>0.05). Specifically, the results showed that the gender had a P-value of 0.79, age had a P-value of 0.29, and smoking history had a P-value of 0.28. The percentage represents the proportion of each characteristic within the group.

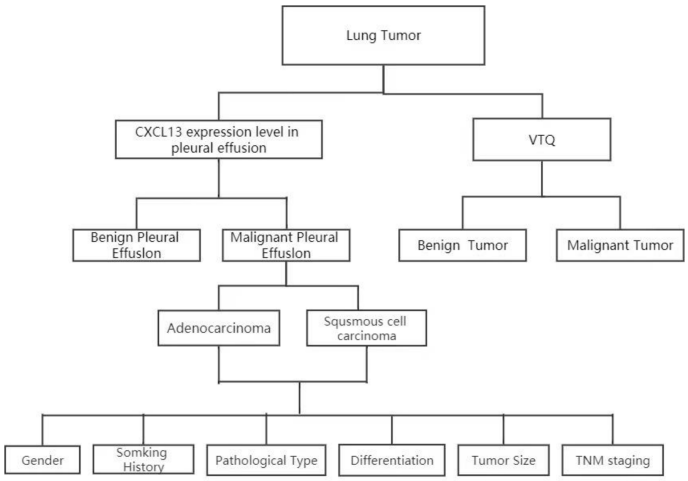


FIGURE 1
Analysis of CXCL13 and VTQ levels in patients with thoracic tumors and pleural effusion. Patient sample collection: Patients with thoracic tumors and pleural effusion were recruited; CXCL13 concentration measurement: The concentration of CXCL13 in pleural effusion was determined; VTQ measurement: Tissue quantification technology (VTQ) was applied to measure the numerical value of thoracic tumors; CXCL13 and clinical feature analysis: The relationship between CXCL13 and various clinical features was further analyzed; ROC curve analysis using SPSS software: The ROC curve analysis was performed using the multivariate observation values (CXCL13 and VTQ) in SPSS software.

Statistical methods

The statistical software SPSS19.0 was used to process the data collected in the study. The data was expressed in different ways depending on whether it followed a normal distribution or not. Measurement data that followed a normal distribution were expressed as the mean ± standard deviation (± s), while non-normally distributed data were presented as the median (interquartile range).

The non-parametric rank-sum test, also known as the Mann-Whitney Test, was used to compare two independent samples of non-normally distributed measurement data. Unlike the t-test, which assumes that the data follows a normal distribution, the Mann-Whitney Test does not make this assumption, making it a suitable test for comparing two samples of non-normally distributed data.

To examine the relationship between the relative expression level of the chemokine CXCL13 and various clinical characteristics in the non-small cell lung cancer (NSCLC) group, the Mann-Whitney Test was employed. The receiver operating characteristic (ROC) curve was used to determine the optimal diagnostic threshold and corresponding sensitivity and specificity of tumor markers in pleural effusion. The ROC curve plots the true positive rate against the false positive rate for various diagnostic thresholds, allowing for the selection of the threshold that provides the best balance between sensitivity and specificity.

The multivariate ROC was utilized to assess the significance of combined diagnoses based on two or more different indicators. This type of analysis allows for the evaluation of the combined diagnostic power of multiple markers, which may provide a more accurate diagnosis than relying on a single marker alone. A p-value of less than 0.05 was considered statistically significant, meaning that the results observed were unlikely to have occurred by chance.

Results

Comparison of relative expression levels of CXCL13 and VTQ results between groups

Our study found that the relative expression levels of CXCL13 and the VTQ values in the lung cancer group were significantly higher compared to those in the control group ($P < 0.05$). Figure 2 displays the results of the VTQ measurements.

The differences between the two groups were analyzed and the results are presented in Table 2. The table summarizes the mean values, standard deviations, and statistical significance of the differences in CXCL13 and VTQ values between the lung cancer group and the control group. By comparing these values, we can understand the relationship between the expression levels of CXCL13 and the VTQ results, and how they may be used as indicators of lung cancer.

Investigating the relationship between the relative expression level of CXCL13 and clinical characteristics in NSCLC group

In our study of the Non-Small Cell Lung Cancer (NSCLC) group, we aimed to examine the relationship between the relative expression level of CXCL13, a chemokine, and various clinical characteristics of the patients. To achieve this, we analyzed the expression levels of CXCL13 and compared them to the TNM staging and tumor differentiation of each patient. Our findings indicated that as the TNM staging of the cancer progressed, meaning the cancer became more advanced, and as the tumor differentiation worsened, meaning the cancer cells were becoming more abnormal, the expression level of CXCL13 increased.

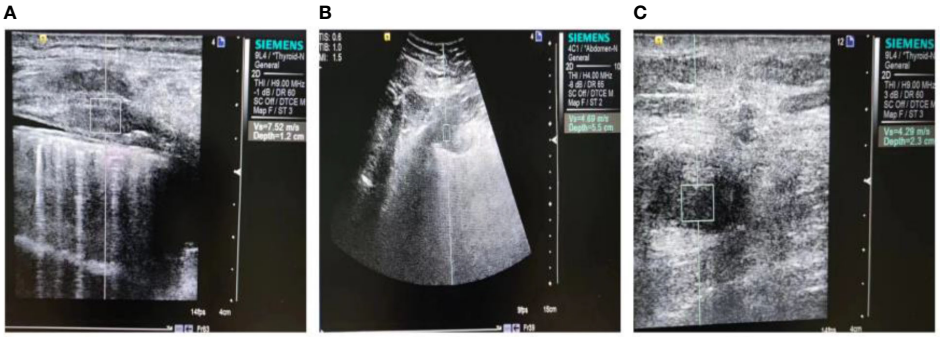


FIGURE 2
Ultrasound images of malignant chest wall and peripheral non-small cell lung cancers. **(A)** A solid hypoechoic nodule is displayed, with poorly defined borders and an irregular shape. The measurement depth is 1.2 cm and the VTQ value is 7.52 m/s. **(B)** A heterogeneous hypoechoic mass is shown, with unclear boundaries and an irregular shape. The measurement depth is 5.5 cm and the VTQ value is 4.69 m/s. **(C)** Another heterogeneous hypoechoic mass is displayed, with unclear borders, an irregular shape, and a burr-like appearance. The measurement depth is 2.3 cm and the VTQ value is 4.29 m/s.

Furthermore, our results showed that the expression level of CXCL13 was higher in patients diagnosed with adenocarcinoma compared to those with squamous cell carcinoma. These findings suggest that the expression level of CXCL13 may play a role in the progression of NSCLC and can provide valuable information for understanding the underlying mechanisms of this type of cancer. All of these results are presented in Table 3 for easy reference and further analysis.

Both the VTQ and CXCL13 biomarkers can be used individually for diagnosing lung tumors

The performance of the two biomarkers was evaluated using receiver operating characteristic (ROC) curve analysis, which is a commonly used method to assess the diagnostic accuracy of biomarkers. The ROC curve plots the true positive rate against the false positive rate at various threshold values, and the area under the curve (AUC) provides a single measure of the diagnostic performance.

In this study, the AUC for CXCL13 in diagnosing lung tumors was found to be 0.74 (95% confidence interval [CI]: 0.61-0.86). This indicates that CXCL13 has a good ability to differentiate between patients with lung tumors and those without. The sensitivity and specificity of CXCL13 were 60% and 80%, respectively, which means that 60% of the lung tumor cases were correctly identified, and 80% of the non-tumor cases were correctly identified. The optimal cut-off value, which is the threshold value that provides the best balance between sensitivity and specificity, was found to be 777.82 pg/ml. Similarly, the AUC for VTQ in diagnosing lung tumors was found to be 0.67 (95% CI: 0.53-0.82). This indicates that VTQ also has good diagnostic performance. The sensitivity and specificity of VTQ were 60% and 83.3%, respectively, with an optimal diagnostic cut-off of 3.33 m/s. These results are shown in Figure 3 and Table 4.

In conclusion, either the VTQ biomarker or the CXCL13 biomarker can serve as an effective individual diagnostic tool for lung tumors.

Using VTQ and CXCL13 in combination has been demonstrated to have a synergistic effect for diagnosing lung tumors

The study included 60 patients, with 30 cases of lung tumors complicated by malignant pleural effusion in the case group and 30 cases of benign pleural effusion in the control group. The gold standard for diagnosis was based on pathological examination, where malignant cases were coded as 1 and benign cases were coded as 0. Two indicators, VTQ and CXCL13, were measured in each patient and were represented as X1 and X2, respectively (where malignant cases were coded as 1, benign cases were coded as -1, and unknown cases were coded as 0). This is summarized in Table 5.

SPSS was used to calculate the individual prediction rate (pre-1) of covariate X1 and the combination of both covariates X1 and X2 (pre-2). The ROC curve analysis was used to evaluate the performance of the variables (pre-1, pre-2) and state variables

TABLE 2 The CXCL13 expression results and VTQ results from the two groups.

	Lung cancer group	Control Group	P
CXCL13 (pg/ml)	879.99 (619.51, 1223.58)	649.12 (525.78, 768.60)	0.002
VTQ (m/s)	3.47 (2.14, 4.93)	2.45 (1.99,3.21)	0.02

As the data is non-normally distributed, it is represented by quartiles: the median (50%), the lower quartile (25%), and the upper quartile (75%). The numbers were computed using the SPSS software.

10.3389/fonc.2023.1115485

TABLE 3 The correlation between the relative expression level of CXCL13 and various clinical characteristics in the NSCLC group.

Characteristics	n	CXCL13 (pg/ml)		
		P50 (P25, P75)	Z	p
Gender				
Male	17	735.11 (610.52, 1225.11)	1.07	0.29
Female	13	960.40 (698.65, 1276.66)		
Age				
≥ 60	16	720.99 (605.67, 1164.01)	1.54	0.12
< 60	14	1016.87 (739.26, 1378.14)		
Smoking history				
Yes	19	777.88 (620.20, 1228.18)	0.37	0.72
No	11	888.14 (617.43,1160.92)		
Pathological type				
squamous cell carcinoma	13	706.86 (588.43,933.24)	2.49*	0.01
adenocarcinoma	17	1160.92 (708.37,1385.22)		
Differentiation				
Low	13	1228.18 (920.91, 1521.25)	3.41*	0.00
Medium and high	17	706.86 (588.43, 882.35)		
Tumor Size				
≥ 4cm	16	778.88 (605.67, 1323.58)	0.33	0.74
< 4cm	14	924.27 (684.50,1166.25)		
TNM staging				
IV a	15	735.11 (600.83, 960.40)	2.43*	0.02
IV b	15	876.82 (636.87,1399.38)		

*Indicates a p-value of less than 0.05. There is a statistical difference with $P < 0.05$.

(group). The results showed that the AUC for covariate X1 (pre-1) in diagnosing malignant pleural effusion was 0.674 (95% confidence interval [CI]: 0.53-0.82). However, the combination of both covariates (pre-2) for the diagnosis of thoracic tumors had a much higher AUC of 0.842 (95% CI: 0.74-0.94). This result is summarised in Table 6 and Figure 4.

In conclusion, the ROC curve analysis using SPSS demonstrated that the combination of VTQ and CXCL13 is a more effective diagnostic tool for lung tumors complicated by malignant pleural effusion compared to the use of either marker alone.

Posteriori power analysis

A posteriori power analysis was conducted to evaluate the sample size and statistical power of the study. The results of the study showed that both the expression levels of CXCL13 and VTQ were significantly higher in the lung cancer group compared to the control group ($P < 0.05$), which suggests that the sample size was sufficient to detect a significant difference between the two groups. However, a larger sample size may have increased the statistical power of the study and improved the precision of the results.

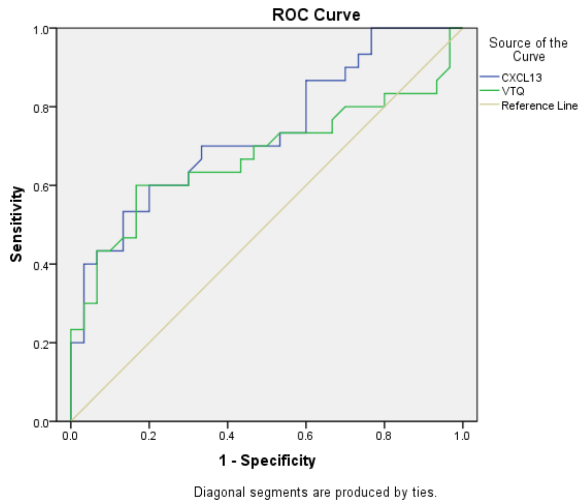


FIGURE 3
Receiver operating characteristic (ROC) curve analysis of the VTQ result in pleural effusion for diagnosing thoracic tumors. The ROC curve was used to evaluate the diagnostic accuracy of the VTQ result in pleural effusion for diagnosing thoracic tumors. The x-axis represents the false positive rate (1-specificity), and the y-axis represents the true positive rate (sensitivity). The diagonal line represents a random guess. The closer the curve is to the top left corner, the better the diagnostic accuracy. The area under the ROC curve (AUC) was 0.67 (95% confidence interval [CI]: 0.53-0.82). The sensitivity and specificity were 60.0% and 83.3%, respectively, and the optimal diagnostic cutoff was 3.33 m/s. The results indicate that the VTQ result in pleural effusion has a moderate diagnostic accuracy for thoracic tumors.

The ROC curve analysis of CXCL13 showed an AUC of 0.74 (0.61, 0.86) with an optimal cut-off value of 777.82 pg/ml, indicating that the test has moderate accuracy for diagnosing lung tumors. The ROC curve analysis of VTQ showed an AUC of 0.67 (0.53, 0.82) with a sensitivity of 60.0% and a specificity of 83.3%, which suggests that the test has limited accuracy for diagnosing lung tumors.

The combination of CXCL13 and VTQ for diagnosing thoracic tumors had an AUC of 0.842 (0.74, 0.94), which was significantly higher than either factor alone. This suggests that the combination of the two tests may be a more accurate tool for diagnosing lung tumors compared to using either test alone.

In conclusion, the sample size of the study was sufficient to detect a significant difference between the two groups, but a larger sample size may have increased the statistical power and precision of the results.

Discussion

Chemokines are a family of small proteins that play a crucial role in various biological processes, including the growth, adhesion, and directional migration of tumor cells. Over the years, an increasing number of studies have highlighted the significant role of chemokines, particularly CXCL13, in tumorigenesis. For instance, CXCL13 has been found to be overexpressed in various solid tumors, such as squamous cell carcinoma and adenocarcinoma, and has been suggested to play a role in regulating the migration and metastasis of cancer cells (2, 3, 5–7).

Studies have shown that the expression level of CXCL13 is positively correlated with the differentiation and stage of lung tumors (2, 3). This has led to the suggestion that CXCL13 may serve as a molecular marker for the diagnosis and prognosis of lung tumors. Although CXCL13 was found to be associated with the differentiation and staging of tumors, it cannot be fully considered as a prognostic indicator without follow-up on changes in CXCL13 levels in pleural effusions before and after treatment and on patient progression-free survival (PFS) and overall survival (OS). In future studies, we will further explore the relationship between changes in CXCL13 levels before and after treatment and overall survival of patients.

To determine the diagnostic value of CXCL13 for lung cancer, the present study established a receiver operating characteristic (ROC) curve, and the area under the curve (AUC) was calculated to be 0.74. Although this suggests that CXCL13 can be used as an independent indicator for lung tumor diagnosis, the study

TABLE 4 Areas under the ROC curves, sensitivity, specificity, and optimal threshold value for both the two techniques.

	n	AUC and 95%CI	sensitivity	specificity	optimal diagnostic cut-off value	p
CXCL13	60	0.74(0.61, 0.86)	60.0%	80.0%	777.82 pg/ml	0.002
VTQ	60	0.67(0.53, 0.82)	60.0%	83.3%	3.33 m/s	0.02

The numbers in parenthesis indicate the corresponding 95% confidence interval.

TABLE 5 Results of binary logistic regression using SPSS.

covariate	regression coefficient	standard error	p-value
X1	0.807	0.302	0.008
X2	0.005	0.002	0.002
Constant	6.07	1.698	0.000

The table displays the logistic regression coefficients, standard errors, and p-values for the covariates X1 and X2, obtained using the Binary Logistic procedure in SPSS. The logistic regression equation was used to generate a new variable in the working data table, which contains the predicted probabilities for each individual.

TABLE 6 Results of ROC curve analysis using SPSS.

Test variable	AUC and 95% confidence interval	standard error	p-value
Pre-1	0.674(0.53, 0.82)	0.073	0.02
Pre-2	0.842(0.74, 0.94)	0.051	0.00

The table shows the results of ROC curve analysis conducted using the ROC Curve function in the SPSS software. The test variable (TestVariable) was used in conjunction with the diagnosis results of the gold standard as the state variable (StateVariable). The variables Pre-1 and Pre-2 were analyzed to determine the area under the curve (AUC), standard error, p-value, and 95% confidence interval.

acknowledges that false negative and false positive results may occur.

Alongside traditional diagnostic techniques, such as lung ultrasound, acoustic palpation tissue quantification (VTQ) technology has emerged as a promising tool for evaluating lung tumors. VTQ is based on acoustic radiation force pulse technology, which uses low-pressure pulses to induce local micro-deformation and measure the velocity of shear waves generated by transverse deformation to quantify tissue hardness (8, 15, 16). This study applied VTQ technology to the dynamic and quantitative evaluation of lung tumors and found that the AUC of VTQ for diagnosing lung tumors was 0.67, with the optimal diagnostic cutoff being 3.33 m/s. Although VTQ showed good diagnostic accuracy, the study acknowledges that factors such as thick abdominal fat, breathing problems, and narrow intercostal spaces may affect the accuracy of the measurements (17–19).

To enhance the diagnostic value of VTQ for lung tumors, the present study combined VTQ with CXCL13 evaluation and found that the AUC of the joint diagnosis was significantly higher (0.842 (0.74, 0.94))than that of a single indicator. Furthermore, the study determined one or two diagnostic points through linear interpolation and divided positive and negative patients into three categories: positive, suspected positive, and negative. By combining VTQ with CXCL13 evaluation, the study demonstrated that this approach has high diagnostic value in differentiating benign and malignant pleural effusions.

In conclusion, the present study highlights the importance of CXCL13 as a potential molecular marker for the diagnosis and prognosis of lung tumors. Furthermore, the combination of VTQ and CXCL13 evaluation demonstrated high diagnostic value in differentiating benign and malignant pleural effusions. The relative expression of CXCL13 in malignant pleural effusions caused by lung tumors may indicate a poor prognosis, and it is expected to become a new indicator for screening and prognosis of patients with advanced lung cancer. Further studies are necessary to validate these findings and establish the clinical relevance of this diagnostic approach.

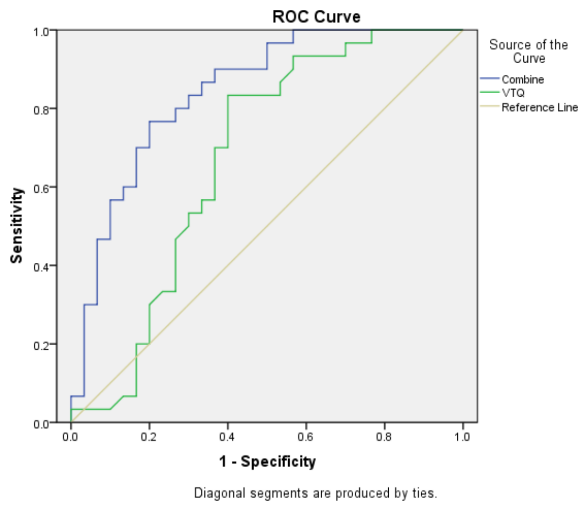


FIGURE 4
The area under the ROC curve (AUC) for the combined use of CXCL13 and VTQ was 0.842 (0.74, 0.94), which was significantly higher than that of VTQ alone (AUC = 0.674, 95% CI: 0.53, 0.82; P < 0.05), indicating a statistically significant difference between the two methods.

Data availability statement

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

Ethics statement

The studies involving human participants were reviewed and approved by The Ethics Committee of the Chinese People's Liberation Army 252 Hospital. The patients/participants provided their written informed consent to participate in this study. Adequate measures were taken to protect the participants' confidentiality and privacy. Additionally, the study was conducted in accordance with the Declaration of Helsinki and Good Clinical Practice guidelines. The researchers took care to minimize any potential harm to the participants and to ensure that their rights and welfare were protected at all times. It is important to emphasize that the high ethical standards of this study were maintained in order to ensure the validity and reliability of the results, as well as to demonstrate respect for the participants and their contributions to the advancement of scientific knowledge.

Author contributions

XZ collected and analyzed the data. XZ wrote the original manuscript. KH, QP, and YJ analyzed the data and edited the manuscript. RT supervised the project, reviewed and edited the manuscript. All authors contributed to the article and approved the submitted version.

Acknowledgments

We thank all the colleagues from Affiliated Hospital of Hebei University and Hospital of the People's liberation Army: 82nd Group Army for their helpful discussion.

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. Xia C, Zheng R, Zhou M, Lin C, Zeng H, Zhang S, et al. Disparities by province, age, and sex in site-specific cancer burden attributable to 23 potentially modifiable risk factors in china: A comparative risk assessment. *Lancet Glob Health* (2019) 7:e257–69. doi: 10.1016/S2214-109X(18)30488-1

2. Singh R, Gupta P, Kloecker GH, Singh S, Lillard JWJ. Expression and clinical significance of cxcr5/cxcl13 in human nonsmall cell lung carcinoma. *Int J Oncol* (2014) 45:2232–40. doi: 10.3892/ijo.2014.2688

3. Qi X-W, Xia S-H, Yin Y, Jin L-F, Pu Y, Hua D, et al. Expression features of cxcr5 and its ligand, cxcl13 associated with poor prognosis of advanced colorectal cancer. *Eur Rev Med Pharmacol Sci* (2014) 18:1916–24.

4. Chao C-C, Lee W-F, Wang S-W, chun Chen P, Yamamoto A, Chang T-M, et al. Cxc chemokine ligand-13 promotes metastasis *via* cxcr5-dependent signaling pathway in non-small cell lung cancer. *J Cell Mol Med* (2021) 25:9128–40. doi: 10.1111/jcmm.16743

5. Zhu D, Ye W, Jiang J. Clinical significance of cxcl13/cxcr5 axis in human cancers. *Trans Cancer Res* (2018) 7:1737–42. doi: 10.21037/tcr.2018.11.26

6. Hussain M, Adah D, Tariq M, Lu Y, Zhang J, Liu J. Cxcl13/cxcr5 signaling axis in cancer. *Life Sci* (2019) 227:175–86. doi: 10.1016/j.lfs.2019.04.053

7. Hsieh C, Jian C, Lin L, Low G, Ou P, Hsu C, et al. Potential role of cxcl13/cxcr5 signaling in immune checkpoint inhibitor treatment in cancer. *Cancers (Basel)* (2022) 14(2):294. doi: 10.3390/cancers14020294

8. Wei H, Lu Y, Ji Q, Zhou H, Zhou X. The application of conventional us and transthoracic ultrasound elastography in evaluating peripheral pulmonary lesions. *Exp Ther Med* (2018) 16:1203–8. doi: 10.3892/etm.2018.6335

9. Zhi X, Wang L, Chen J, Zheng X, Li Y, Sun J. Scoring model of convex probe endobronchial ultrasound multimodal imaging in differentiating benign and malignant lung lesions. *J Thorac Dis* (2020) 12:7645–55. doi: 10.21037/jtd-2020-abpd-005

10. Ozgokce M, Yavuz A, Akbudak I, Durmaz F, Uney I, Aydin Y, et al. Usability of transthoracic shear wave elastography in differentiation of subpleural solid masses. *Ultrasound Q* (2018) 34:233–7. doi: 10.1097/RUQ.0000000000000374

11. Liu Y, Zhen Y, Zhang X, Gao F, Lu X. Application of transthoracic shear wave elastography in evaluating subpleural pulmonary lesions. *Eur J Radiol Open* (2021) 8:100364. doi: 10.1016/j.ejro.2021.100364

12. Caroli G, Dell'Amore A, Cassanelli N, Dolci G, Pipitone E, Asadi N, et al. Accuracy of transthoracic ultrasound for the prediction of chest wall infiltration by lung cancer and of lung infiltration by chest wall tumours. *Heart Lung Circ* (2015) 24:1020–6. doi: 10.1016/j.hlc.2015.03.018

13. Lim C-K, Chung C-L, Lin Y-T, Chang C-H, Lai Y-C, Wang H-C, et al. Transthoracic ultrasound elastography in pulmonary lesions and diseases. *Ultrasound Med Biol* (2017) 43:145–52. doi: 10.1016/j.ultrasmedbio.2016.08.028

14. Sperandeo M, Trovato FM, Dimitri L, Catalano D, Simeone A, Martines GF, et al. Lung transthoracic ultrasound elastography imaging and guided biopsies of subpleural cancer: A preliminary report. *Acta Radiol* (2015) 56:798–805. doi: 10.1177/0284185114538424

15. Xu J-M, Chen Y-J, Dang Y-Y, Chen M. Association between preoperative us, elastography features and prognostic factors of papillary thyroid cancer with braf (v600e) mutation. *Front Endocrinol (Lausanne)* (2020) 10:902. doi: 10.3389/fendo.2019.00902

16. Zhong L-C, Yang T, Gu L-P, Ma F. The diagnostic performance of shear wave velocity ratio for the differential diagnosis of benign and malignant breast lesions: Compared with vtq, and mammography. *Clin Hemorheol Microcirc* (2021) 77:123–31. doi: 10.3233/CH-200813

17. Nakano C, Nishimura T, Tada T, Yoshida M, Takashima T, Aizawa N, et al. Severity of liver fibrosis using shear wave elastography is influenced by hepatic necroinflammation in chronic hepatitis patients, but not in cirrhotic patients. *Hepatol Res* (2021) 51:436–44. doi: 10.1111/hepr.13617

18. Shimada S, Kamiyama T, Kakisaka T, Orimo T, Nagatsu A, Asahi Y, et al. The impact of elastography with virtual touch quantification of future remnant liver before major hepatectomy. *Quant Imag Med Surg* (2021) 11:2572–85. doi: 10.21037/qims-20-1073

19. Takuma Y, Morimoto Y, Takabatake H, Tomokuni J, Sahara A, Matsueda K, et al. Changes in liver and spleen stiffness by virtual touch quantification technique after balloon-occluded retrograde transvenous obliteration of gastric varices and exacerbation of esophageal varices: A preliminary study. *Ultraschall Med* (2020) 41:157–66. doi: 10.1055/a-0731-0137

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 19 April 2023
DOI 10.3389/fonc.2023.1095353

OPEN ACCESS

EDITED BY
Bahram Mohajer,
Johns Hopkins Medicine, United States

REVIEWED BY
Xiangyuan Ma,
Shantou University, China
Enrico Checcucci,
IRCCS Candiolo Cancer Institute, Italy

*CORRESPONDENCE
Ming Zhang
zmmri@163.com

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

RECEIVED 11 November 2022
ACCEPTED 30 March 2023
PUBLISHED 19 April 2023

CITATION
Ren H, Ren C, Guo Z, Zhang G, Luo X, Ren Z, Tian H, Li W, Yuan H, Hao L, Wang J and Zhang M (2023) A novel approach for automatic segmentation of prostate and its lesion regions on magnetic resonance imaging.
Front. Oncol. 13:1095353.
doi: 10.3389/fonc.2023.1095353

A novel approach for automatic segmentation of prostate and its lesion regions on magnetic resonance imaging

Huipeng Ren[1,2], Chengjuan Ren[3], Ziyu Guo[4], Guangnan Zhang[5], Xiaohui Luo[6], Zhuanqin Ren[2], Hongzhe Tian[2], Wei Li[2], Hao Yuan[5], Lele Hao[5], Jiacheng Wang[5] and Ming Zhang[1]*

[1]Department of Medical Imaging, First Affiliated Hospital of Xi'an Jiaotong University, Xi'an, China, [2]Department of Medical Imaging, Baoji Central Hospital, Baoji, China, [3]Department of Language Intelligence, Sichuan International Studies University, Chongqing, China, [4]Department of Computer Science & Engineering, The Chinese University of Hong Kong, Hong Kong, Hong Kong SAR, China, [5]Department of Computer Science, Baoji University of Arts and Sciences, Baoji, China, [6]Department of Urology, Baoji Central Hospital, Baoji, China

Objective: To develop an accurate and automatic segmentation model based on convolution neural network to segment the prostate and its lesion regions.

Methods: Of all 180 subjects, 122 healthy individuals and 58 patients with prostate cancer were included. For each subject, all slices of the prostate were comprised in the DWIs. A novel DCNN is proposed to automatically segment the prostate and its lesion regions. This model is inspired by the U-Net model with the encoding-decoding path as the backbone, importing dense block, attention mechanism techniques, and group norm-Atrous Spatial Pyramidal Pooling. Data augmentation was used to avoid overfitting in training. In the experimental phase, the data set was randomly divided into a training (70%), testing set (30%). four-fold cross-validation methods were used to obtain results for each metric.

Results: The proposed model achieved in terms of Iou, Dice score, accuracy, sensitivity, 95% Hausdorff Distance, 86.82%,93.90%, 94.11%, 93.8%,7.84 for the prostate, 79.2%, 89.51%, 88.43%,89.31%,8.39 for lesion region in segmentation. Compared to the state-of-the-art models, FCN, U-Net, U-Net++, and ResU-Net, the segmentation model achieved more promising results.

Conclusion: The proposed model yielded excellent performance in accurate and automatic segmentation of the prostate and lesion regions, revealing that the novel deep convolutional neural network could be used in clinical disease treatment and diagnosis.

KEYWORDS

prostate cancer, convolution neural network, dense block, attention mechanism, U-Net

Introduction

Prostate cancer (PCa) is a significantly prevalent cancer among men, accounting for a half about of cancer diagnoses and the fifth highest cause of mortality (1). The age at which prostate cancer typically appears is relatively late, with most cases occurring after the age of 55, and the incidence gradually increases thereafter, peaking between the ages of 70 and 80. However, in cases of familial hereditary prostate cancer, the age of onset tends to be slightly earlier, with 43% of affected individuals developing the disease before the age of 55 (2). Prostate cancer can be diagnosed, treated, and monitored using several imaging modalities, including transrectal ultrasound (TRUS), magnetic resonance imaging (MRI), and computed tomography (CT). These imaging techniques are commonly used in clinical practice to aid in the diagnosis and management of prostate cancer.

In recent years, MRI technology has advanced significantly, providing high-level spatial resolution and soft tissue conspicuity, which makes MRI a suitable imaging technique for prostate segmentaion, staging and volum calculation of prostate cancer. The high-level spatial resolution and soft tissue conspicuity of MRI make it appropriate for prostate segmentation, staging and volume calculation of prostate cancer (3). In the same way, dynamic contrast-enhanced MRI (DCE-MRI) can be used to recognize malignant structures according to the spread rate of contrast agents (4–6), and magnetic resonance spectroscopy can discriminate malignant tissues in terms of the relative intensities of different metabolites (e.g., citric acid, choline, and creatine).

Traditional segmentation approaches of prostate or lesion area include contour and shape-based approaches and region-based techniques, and some hybrid methods. The prostate edge or boundary is applied to prostate segmentation. For example, Zwiggelaar et al. (7) employed a system of first and second-order Lindeberg directional derivatives (8) coordinates in polar coordinates to discern edges. To obtain the prostate boundary, the inverse shift of the longest curve was chosen after non-extreme on the disconnected curve in the vertical dimension. Flores-Tapia et al. (9) traced the boundary using *a priori* form information of the prostate by shifting a little filter mask over a feature space that was constructed from the Haar wavelet in the multiresolution structure. Klein et al. (10) adopted a multi-atlas method to segment the prostate. The training data was aligned to with the test data by affine alignment and subsequent non-rigid alignment with three b-spline bars in the framework. Gao et al. (11) developed the training set shape as a point cloud. The shape prior and local image statistics were integrated into the energy function to minimize the energy function for prostate segmentation in a level-set format. Manual segmentation remains the most widely utilized method for achieving accurate segmentation of the prostate and lesion region. it is not only a very time-consuming task and is subject to tissue variations. Additionally, it also heavily depends on the level of manual expertise and experience, which can lead to low reproducibility and higher observer variation.

In order to overcome those issues, there is an urgent need for reliable automatic segmentation of the prostate and lesion region in daily clinical practice. In 2012, Hilton's team participated in the ImageNet image recognition competition for the first time. AlexNet (12) was awarded the championship and crushed the second classification performance of support vector machine. As a result, the development of deep learning technology was advancing by leaps and bounds, and it was applied in many directions. CNN continuously extracts features from all layers, from local to overall features. CNNs are obtaining a concern in the medical image field due to the state-of-the-art scores on plentiful image identification and segmentation tasks. One of the outstanding representative works is that Ronneberger et al. (13) proposed the U-Net model and made full use of the limited and valuable training set to boost segmentation performance. The U-shaped structure makes the localization accurate for medical images instead of simple binary classification. The overall process of U-Net includes encoding and decoding, with only a convolutional layer and no fully connected layer. Currently, U-Net is arguably an even hotter segmentation network. Lai et al. (14) proposed a network for automatic segmentation for prostate zone and cancer RoI by Segnet. They considered different sequences into three channels of an image and used PROSTATEx dataset to train the network. At last, the T2W + DWI + ADC scheme obtained the best grade with a Dice similarity coefficient of 90.45%. Wang et al. (15) presented a 3D CNN model and used the attention mechanism to fully mining more useful features encoded in the network for prostate segmentation. To enhance local prostate cancer control, Chen et al. (16) proposed three-branch U-Net to distinguish different targets for segmentation in MRI. Deep monitoring policies were combined into the network to accelerate convergence and boost network capabilities. To reduce the loss of structural and spatial information, Orlando et al. (17) designed a 3D segmentation model based on 2D U-Net for the prostate. The novel model can offer a quick and effective segmentation compared to other methods.

It is very challenging to get an automatic segmentation model with high performance for the prostate and its lesion region. The ambiguity of each tissue boundary inside the image makes it difficult to distinguish it from the heterogeneous tissue within the surrounding prostate, further resulting in under-segmentation or over-segmentation. Additionally, the varying sizes and shapes of prostate glands among individuals pose challenges in modeling pervasive learning. The above reasons make regional resection of prostate cancer difficult and challenging. To address these challenges, we propose a new network for the automatic segmentation of prostate and prostate cancer regions. Our network is inspired by U-Net and utilizes a simple but effective attention module, which could be broadly used to improve the capability of CNN. In short, the attention network is in charge of focusing attention on certain important features of an image which improves the segmentation quality. Dense block also is employed to mitigate gradient disappearance and enhance the propagation of features in the model. Additionally, the dense block is employed to mitigate gradient disappearance and enhance the propagation of features, resulting in more abstracted interested features. In the data preparation phase, data augmentation is utilized to solve the problem of overfitting the model due to limited amount of date. The main contributions of this work are as follows.

Firstly, to fulfill the pixel-wise segmentation, a novel CNN model is proposed in this study. The model uses lengthy skip connections between the relative stages of the encoder and decoder and facilitates end-to-end training. To expand the perceptual field of the convolution kernel without loss of resolution (no down-sampling), group norm- Atrous Spatial Pyramidal Pooling is introduced in our model.

Secondly, to stabilize parameter updating and keep a more effective image feature, the dense block is incorporated to utilize short skip connections between different convolutional layers.

Thirdly, the introduction of CBAM is used to make the network more seneitive to the characteristics of both channel and space dimensions. In this study, CBAM allows the model to focus more features on the prostate and its lesion areas from space and channels. Thus, the model helps the flow of information within the network by learning which information should be emphasized and suppressed.

Finally, we evaluate the proposed model on a real dataset and show its effectiveness by outperforming state-of-the-art segmentation models on multiple evaluation indices.

Materials and methods

Datasets

The data were conducted with MRI from 180 patients (122 healthy individuals and 58 patients with prostate cancer from pathology report). Data were acquired using the GE3.0T 750 MR between January 2018 and May 2021. Informed consent was obtained from all patients. The input sequence was the DWI format. DWI: TE82ms, TR 6000ms, Thickness 3.5 mm, Scan Matrix 128 ×128, b value 0, 500, 1000, 1500mm^2/s. To ensure that the ground truth segmentation was as correct as possible, six experienced prostate clinicians participated in analyzing and annotating the prostate MR images. Three clinicians spent 3 months demarcating all masks with the monitoring interface (Labelme). To compensate for inaccurate label borders that may be caused by subjective physician judgment. The other three experts reviewed and revised the annotating masks. The overlapping part of the two outlined areas was finally considered as the labeled target.

Data pre-processing

The primary pre-processing stages used in this approach are data enhancement and image normalization. Data augmentation is used to address overfitting issues in the raw data. It involves following operations: image rotation by a variable number of degrees (-10, 0, 45,60); shifting the image up, down, left, and right; and resizing the image 0.9 and 1.1 times. Finally, the number of prostate and lesion region samples is 1936 and 514. Deep learning models are required to normalize their input data to ensure an adequate convergence point. Normalization can be achieved using different strategies, such as min-max norm and the linear function which converts the input data to the range [0, 1]. This operation achieves equal scaling of the original images. Another approach is 0-mean normalization, which normalizes the raw data set to have a mean of 0 and a variance of 1. In this study, we use a mini–maxi norm to apply the linear transformation to the raw data range. The formula is Xnorm=(X- Xmin)/(Xmax- Xmin), where Xnorm is the normalized data, X is the raw data, and Xmax and Xmin are the maximum and minimum values of the raw data set, respectively. This unique technique is suitable for the image at a predefined mask. The data normalization procedure is performed on the test data to obtain homogeneity, as necessary for the model to provide robust results.

The proposed model

The model is encouraged by the classical U-Net network (13) and considered the decoding-encoding idea. Meantime, the model introduces dense blocks, convolution block attention module (CBAM) and group norm-Atrous Spatial Pyramidal Pooling (GN-ASPP) (18) to capture more feature representation in segmentation. To effectively utilize shallow information, the proposed model fuses features from the contraction path into the expansion path at both symmetrical and asymmetrical levels. Figure 1 presents the details of the proposed model. DWI image is considered for model input. The convolution operations inside the model are all performed using 3×3 marked in red. The model consists of a contraction path (left side) and an expansion path (right side). The contraction path is designed to produce contextual information and the extension path is for precise positioning, and the two paths are mutually synchronous. The whole architecture utilizes short skip connections between various convolution layers at each step, which assists in steady parameter optimization. The union of long and short skips boosts the general efficiency of the network (19). The contraction path is responsible for downsampling and the number of channels increases from 64 to 1024. In the expansive path, each step involves an up-convolution of the prostate feature map, followed by a 2×2 convolution operation that reduces the number of feature channels by half. Another component is a concatenation with the tailoring prostate feature from the contracting path of the same layer. Apart from that, two 3×3 convolutions, each postulated with a ReLU and a CBAM, are included in the expansive path. The last layer employs three convolutions and a spatial pyramidal pooling with rates (6, 12, 18) to determine the number of classes. As boundary pixels are lost on each convolution, trimming is necessary.

A dense block (20) is a dense concatenation of numerous composite functions which makes up batch normalization, ReLU layer, convolutional layer, and dropout layer. It serves to mitigate the gradient disappearance and enhance the propagation of the prostate and its lesion features and reuse them in the subsequent network layer. The CBAM module (21) uses the attention mechanism to optionally optimize the multi-dimension image features and extract the interest features at each layer, inhibiting more non-relative noise. The network can generate the channel and

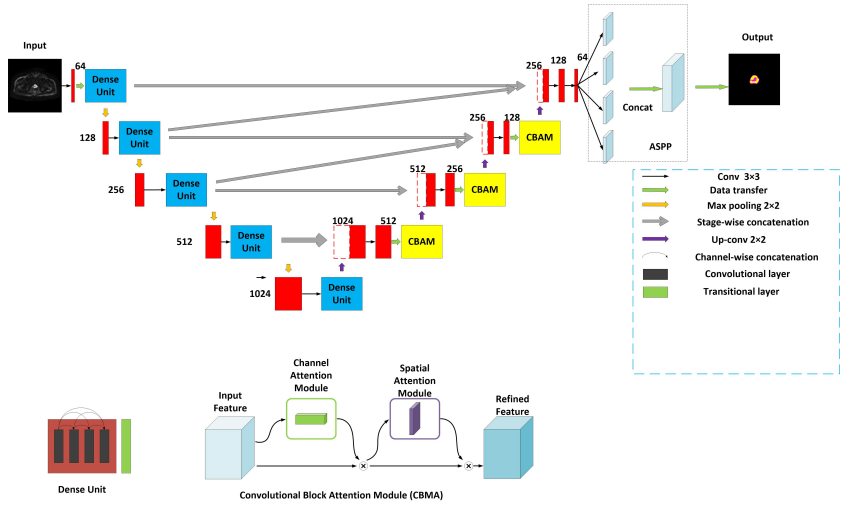


FIGURE 1
Structure of the proposed method. Yellow for CBMA module, dark blue for dense unit, with ASSP added to the end of the model.

spatial attention map by separately mining the inter-channel and inter-spatial relationship of features, which explains 'what' and 'where' issues. CBAM structure is made up of channel attention and spatial attention. The input DWI map is F (2-channel). It is also an intermediate feature map. CBAM defines a 1D channel attention map M_C and 1D spatial attention map M_S (18).

$$M_C(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F)))$$
$$= \sigma(W_1(W_0(F^C_{avg})) + W_1(W_0(F^C_{max})))$$

Where σ is the sigmoid function, $W_0 \in R^{(c/r\times c)}$, and $W_1 \in R^{(c/r\times c)}$. r denotes the reduction ratio. The hidden activation size is set to $R^{C/r\times 1\times 1}$. Note that multi-layer (MLP) weights, W_0 and W_1, are shared for both inputs and the ReLU is followed by W_0. F^C_{avg} and F^C_{max} denote two spatial context descriptors.

Where $f^{(7*7)}$ represents a convolution operation with a filter size of 7*7.

The attention principle can be explained as follows:

$F' = M_C(F) \otimes F$

$F'' = M_S(F') \otimes F'$

⊗ representatives element-wise multiplication. F^" denotes the final output.

The multi-scale feature maps obtained are then not directly used to predict the condition of object regions. To achieve more precise performance of the prostate and its lesion regions, we employ the spatial pyramidal pooling module to rescale attention features at various scales. As shown in Figure 2, features from up-sampling and dense block connections are fused to form a feature map of interest. The feature map is then processed by applying four parallel convolutions with different rates to collect various information. Our ASPP includes a 1×1 convolution and a triple 3×3 dilated convolution with rates of (6, 12, 18). Each convolution is followed by a normalization. We selected group norm (GN) over batch normalization because GN's accuracy is fairly stable over a wide range of batch sizes.

Evaluation of the proposed model

We evaluated the developed network in comparison with other state-of-the-art segmentation networks, including FCN (22), U-Net (13), U-Net++ (23), ResU-Net (24). To ensure a fair comparison, these models were retrained to produce the best separation results. During the training period, the training cross-entropy loss is exploited. The optimized method employed is Adam as it converges faster. The model was trained 150 times for almost all architectures. Checkpoint and stopping methods were utilized to reduce computation time.

To quantitatively assess the segmentation, we utilized several indicators, including Dice Similarity Coefficient (Dice), IoU, sensitivity, accuracy, and Hausdorff Distance (HD). Dice was utilized to assess the likelihood of similarity between the segmented volume and the ground truth. Dice are utilized to assess the likelihood of similarity between the segmented volume and the ground truth. Accuracy and Iou were appraised from the perspective of voxel classification for segmentation. Hausdorff

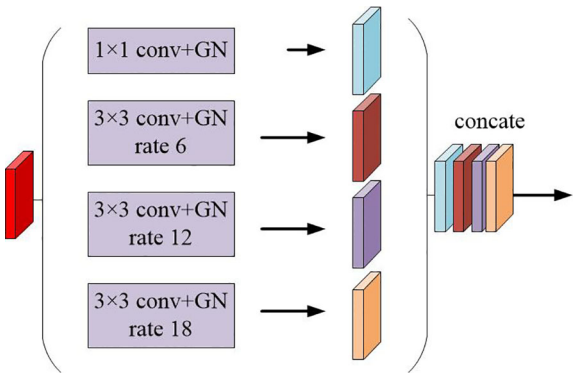


FIGURE 2
Schematic representation of the spatial pyramid set (ASPP) for dilated convolution and group norm (GN).

distance is a measure that describes the degree of similarity between two sets of points. The dataset was divided into training (70%), and testing set (30%) when each experiment was conducted in a randomized manner. Four-fold cross-validation method was used to obtain experimental results. All training data was randomly divided into 4 sets, 3 of which were used for training and the remaining one for validation. When this round was completed, 3 parts were randomly selected again to train the data. Finally, the optimal parameters was selected from loss evaluation.

Results

Comparison of the state-of-the-art algorithms

Loss vs Epoch

The training process was recorded, as shown in Figures 3 and 4. These two figures represent the effect of prostate area and lesion area vs epoch, respectively. Each epoch is one round of data re-iterations. The two figures show similar loss trends in the prostate and its lesion segmentation. The loss decreases from epoch 0 to 60. The model starts convergence from epoch 60. Although the ResUnet model exhibits a higher loss value compared to other models, all models exhibit similar convergence trends, and the model slowly converges as the number of training sessions increases.

Iou vs epoch

Iou is the intersection of the predicted and true results. It is often used as a metric to assess how well a model is learning. The curve chart of Iou vs epoch for the proposed algorithm is given in Figures 5 and 6, respectively. From 0 to 60 epochs, Iou of our method is unstable and fluctuates. But, the proposed approach is capable of producing better Iou than the state-of-the-art segmentation methods when the model converges to fit. In particular, it can be seen that the Iou of our algorithm is significantly greater than the other algorithms in Figure 5, which demonstrates that our model has better segmentation performance.

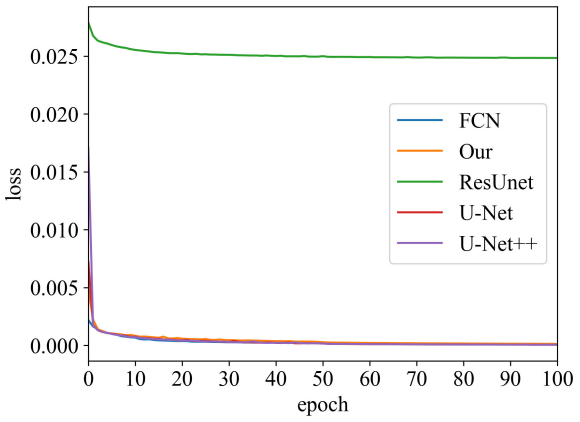


FIGURE 3
Loss vs epoch of prostate area on training data.

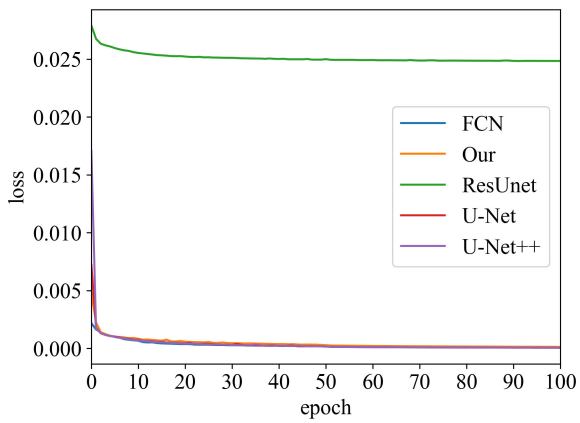


FIGURE 4
Loss vs epoch of prostate lesion area on training data.

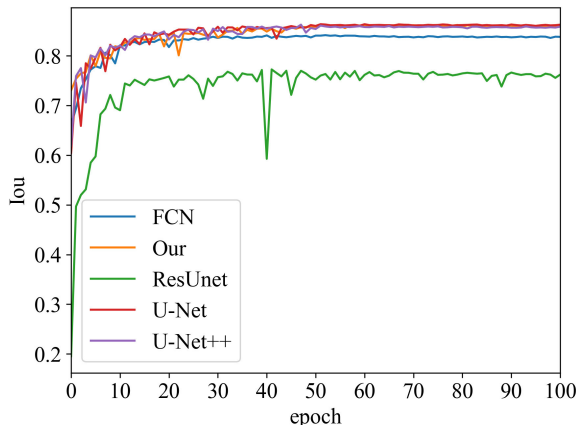


FIGURE 5
Iou vs epoch of the prostate area on testing data.

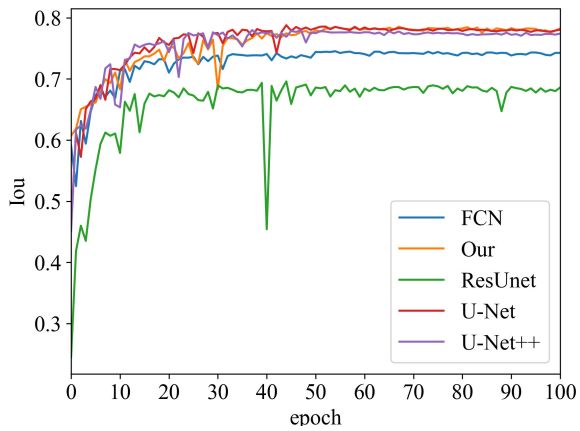


FIGURE 6
Iou vs epoch of prostate lesion area on testing data.

The segmentation performance was computed for the prostate and its lesion area with four cross-validations. In Table 1, the proposed network achieved an average dice sorce, Iou, accuracy, sensitivity for the prostate of 93.90%, 86.82%, 94.11%, and 93.80%. Our algorithm outperformed U-Net, U-Net++, FCN, and ResU-Net in Dice metrics by 1.7, 1.4, 2.6, and 7.19 percentage points and in terms of sensitivity metrics by 1.55, 2.68, 1.9, 10.5 percentage points for the prostate segmentation. In Table 2, for the segmentation performance of the prostate lesion, the proposed network yielded better results in terms of Dice score of 89.51%, Iou of 79.20%, accuracy of 88.43%, sensitivity of 89.31%, and 95%HD of 8.39 compared with the other 4 models. These findings suggest that this algorithm has superior performance compared to other models.

We did ablation experiments to verify the effect of each module on our model in Table 3. The dense block, CBMA, GN-ASPP were gradually increased on the backbone. In our dataset, the ablation experiment was implemented with identical model parameters, e.g. Adam, learning rate, model initialization, and loss function. This result shows that the guidance technology is reinforcing to each other. Our approach achieved the best performance and could learn more robust representation from dense block, CBMA, and GN-ASPP.

Visualization of segmentation effect

Figure 7 presents schematic images of the segmentation results obtained from our model. From the test results, we randomly selected 4 samples for the presentation of the results. Our model accurately distinguished between the prostate areas and lesion areas, with the segmentation of the prostate region being more accurate compared to the lesion area. These findings are consistent with the results presented in Table 1 and 2. To show the effect of CBAM in our model, the visualization attention results of the final layer of the model for the prostate lesion region are given in Figure 8. The rose mask denotes the area of interest which is the prostate lesion region.

Discussion

In this study, we propose a novel DL-based architecture that utilizes the dense block and CBAM, as well as the GN-ASPP module, to fully leverage the complementary information encoded in different layers of the model. Our proposed method is designed to improve the segmentation performance of the prostate and its lesion regions and aid clinical diagnosis. The segmentation output

TABLE 1 Segmentation performance of prostate area for five models.

Model	Dice	Iou	Accuracy	Sensitivity	95%HD(mm)
U-Net	92.20%	85.81%	93.0%	92.35%	8.94
U-Net++	92.50%	84.90%	93.6%	91.12%	8.89
FCN	91.30%	84.62%	92.4%	91.90%	8.71
ResU-Net	86.71%	76.01%	89.51%	83.30%	8.51
Proposed	93.90%	86.82%	94.11%	93.80%	7.84

TABLE 2 Segmentation performance of prostate lesion area for five models.

Model	Dice	Iou	Accuracy	Sensitivity	95%HD(mm)
U-Net	87.50%	77.91%	87.40%	88.53%	9.01
U-Net++	88.20%	77.45%	86.21%	87.56%	8.82
FCN	85.31%	75.06%	86.11%	85.03%	8.73
ResU-Net	81.21%	69.14%	86.30%	81.19%	8.66
Proposed	89.51%	79.20%	88.43%	89.31%	8.39

TABLE 3 Ablation experiments for the segmentation of the prostate and its lesion regions (√notes to introduce this technology in the model).

	Backbone (U-Net with feature fusion)					
Dense Block	√	√	√		√	
CBMA	√	√		√		√
GN-ASPP	√		√	√		
Dice of Prostate/Prostate lesion regions	93.90%/89.55%	93.00%/88.34%	91.24%/88.67%	89.71%/87.94%	88.86%/87.81%	88.67%/86.31%

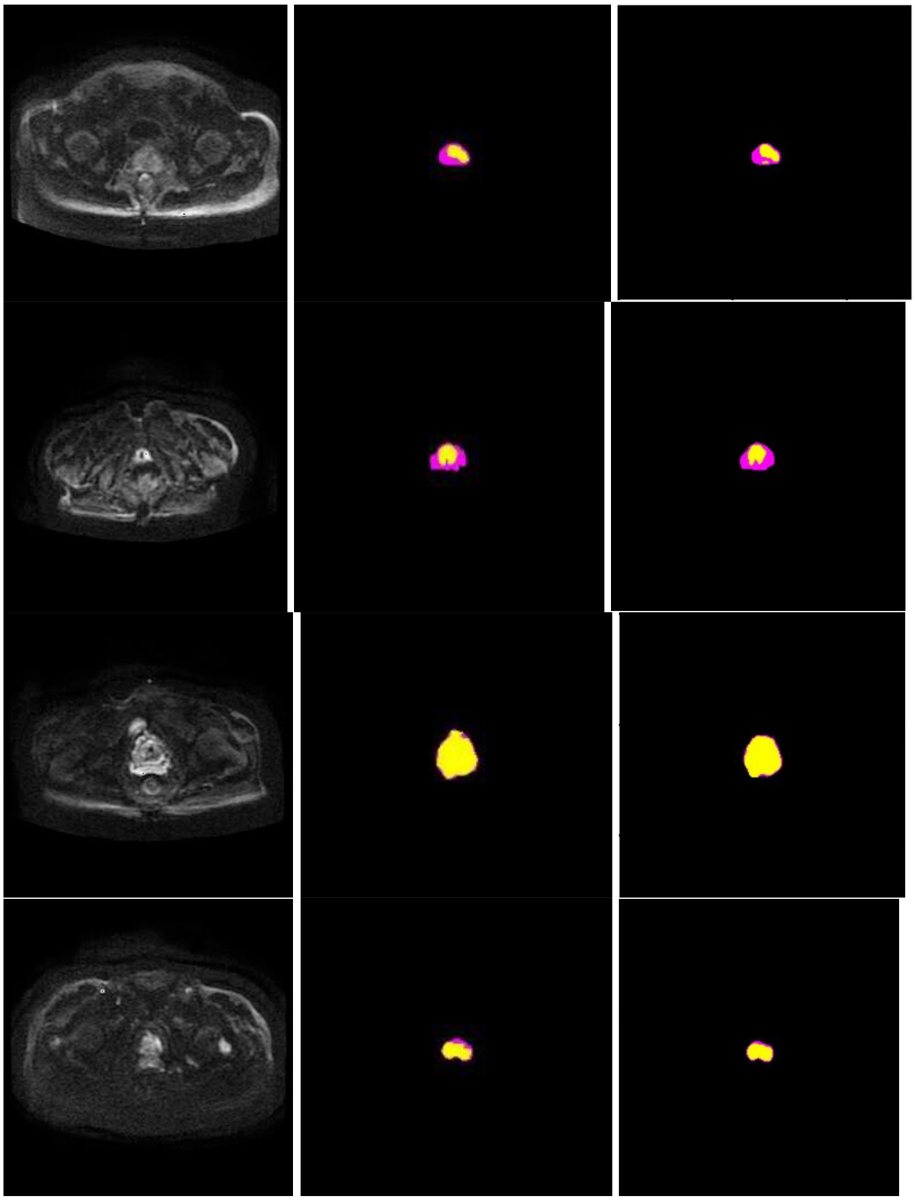


FIGURE 7
Segmentation performance of the proposed method in 4 different patients (row), and columns from left to right show input image, ground truth, and segmentation results of the proposed model. In the experiments, non-target regions were masked black to provide greater clarity. The lesion region is marked in yellow, while the prostate region in rose.

is obtained through an end-to-end approach. The model performance was evaluated on a real dataset. The experimental setup consisted of several aspects: observing the loss value and Iou change of each algorithm during training, quantitative comparison of the performance of each model, and visualization of model results. Finally, it was demonstrated that the proposed segmentation method outperformed the results of the state-of-the-art methods for segmentation of the prostate and lesion region. Specifically, the proposed method exhibited excellent results, especially for the lesion region, which is of great significance for clinical diagnosis and treatment.

Several studies demonstrates artifical intelligence is valid in urology works (25–27), especially using DCNN to segment the prostate or determine prostate cancer. Zhu et al. (28) designed a DCNN model to segment the prostate zone and outer contour. The model was derived from a cascade of two models. One model was responsible for segmenting the prostate region and one for segmenting the prostate zone. However, an end-to-end model, like the one proposed in our study, is more efficient in reducing training time and facilitating clinical diagnosis. Duran et al. (29) also developed a novel CNN model for PCa segmentation with an attention mechanism. This strategy is similar to our approach. Moreover, we used CBAM in the model which focused

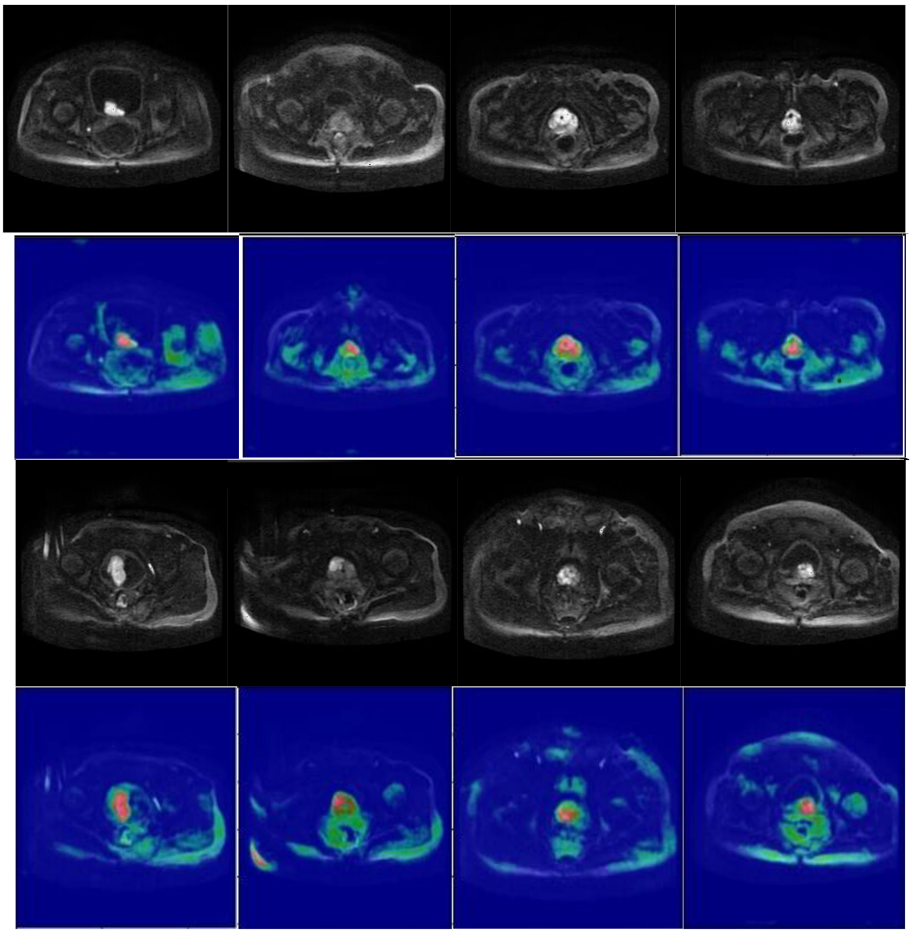


FIGURE 8
Visualization of the final layer of our model for prostate lesion region.

on both the spatial dimension and the channel dimension. It is essential to design a model that extracts as many effective features as possible to automatically and accurately segment the prostate and its lesion region.

Firstly, we adopt a symmetrical architecture inspired by the classical U-Net, with a contraction path and an expansion path, copy and concatenate connections in the same layer, as well as a fusion of features in different layers. There is also a fusion of features in different layers. The main idea of the architecture is to continuously perform deeper feature extraction of prostate image features using the systolic network and to supplement the images lost by the systolic network using the expansion network. At the same time, in order to make more efficient use of the underlying feature information, we fuse the lower-down sampling features into the up-sampling features of the upper layer.

Secondly, the dense block is added to the left side of the model to allow for effective retention and propagation of the prostate and its lesion features. Then, the introduction of the attention mechanism allows the model to focus on areas of the prostate and its lesion in both spatial and channel dimensions, as shown in Figure 8. The huge number of parameters of the DCNN model affects the prediction results. Dense block and CBAM are integrated into the model to enhance the performance of the segmentation model without increasing the burden on the backbone (20, 21). To expand the perceptual field of the convolution kernel without loss of resolution (no down-sampling), GN-ASPP is introduced in our model.

Thirdly, some deep models for prostate cancer discrimination take the combination of image sequences as input (14, 30, 31) or the models take multi-branch outputs (16, 28). The more complex the model, the more training examples it requires, leading to a higher risk of overfitting. In contrast, our model takes a single image input and produces one branch output while still achieving excellent results.

Moreover, our method has the same convergence effect as the classical U-Net because of the backbone of the model, as shown in Figures 3 and 4. Meantime, our model has also similar convergence rates and effects in both the prostate and lesion regions, which demonstrates that the model has generalization properties.

Some restrictions of our study are worth mentioning (1), In deep learning, the more the amount of data, the better the final result will be. But in this study, data for model training is scarce (2), The operation efficiency of this network could be improved. For each neighborhood, the network has to run once, and for the overlapping part of the neighborhood, the network performs repeated operations (3), Initialization of parameters has a great impact on model training. Compared to the random initialization

of the model parameters we studied, an effective initialization is more beneficial for the convergence of the model. To ensure the success of clinical application, it is essential to devep a robust and generalizable model. In the future, we will continue to collect more images of the prostate and increase the size of training set. Additionally, the combined use of data from different formats of MRI can compensate for the deficiencies of single data and improve the segmentation performance of the prostate and lesion regions.

In conclusion, we have proposed a DCNN model with dense block, convolution block attention module, and group norm-Atrous Spatial Pyramid Pooling for the segmentation of the prostate and its lesion regions. Experiments showed that this automatic segmentation model had excellent scores, which supports its potential to assist prostate disease diagnosis and treatment in clinical medicine.

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

MZ contributed to the conception and design of the article. HR and ZG analyzed and wrote the manuscript. ZG, CR, GZ, XL, ZR, HT, WL, HY, LH, and JW revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by projects Baoji Central Hospital BZXJJMDT-2020-2 and Shaanxi Provincial Science and Technology Department 2023-YBSF-061.

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.

References

1. Aldoj N, Lukas S, Dewey M, Penzkofer T. Semi-automatic classification of prostate cancer on multi-parametric MR imaging using a multi-channel 3D convolutional neural network. *Eur Radiol* (2020) 30(2):1243–53. doi: 10.1007/s00330-019-06417-z

2. Ferlay J, Colombet M, Soerjomataram I, Parkin DM, Piñeros M, Znaor A, et al. Cancer statistics for the year 2020: An overview. *Int J Cancer* (2021) 149(4):778–89. doi: 10.1002/ijc.33588

3. Zettinig O, Shah A, Hennersperger C, Eiber M, Kroll C, Kübler H, et al. Multimodal image-guided prostate fusion biopsy based on automatic deformable registration. *Int J Comput Assisted Radiol Surg* (2015) 10(12):1997–2007. doi: 10.1007/s11548-015-1233-y

4. Ozer S, Langer DL, Liu X, Haider MA, van der Kwast TH, Evans AJ, et al. Supervised and unsupervised methods for prostate cancer segmentation with multispectral MRI. *Med Phys* (2010) 37(4):1873–83. doi: 10.1118/1.3359459

5. Puech P, Betrouni N, Makni N, Dewalle AS, Villers A, Lemaitre L. Computer-assisted diagnosis of prostate cancer using DCE-MRI data: design, implementation and preliminary results. *Int J Comput Assisted Radiol Surg* (2009) 4(1):1–10. doi: 10.1007/s11548-008-0261-2

6. Liu X, Langer DL, Haider MA, Yang Y, Wernick MN, Yetik IS. Prostate cancer segmentation with simultaneous estimation of Markov random field parameters and class. *IEEE Trans Med Imaging* (2009) 28(6):906–15. doi: 10.1109/TMI.2009.2012888

7. Zwiggelaar R, Zhu Y, Williams S. Semi-automatic segmentation of the prostate, in: *Pattern Recognition and Image Analysis: First Iberian Conference, IbPRIA 2003, Puerto de Andratx, Mallorca, Spain,* June 4-6, 2003. Proceedings 1. Springer Berlin Heidelberg. (2003) pp. 1108–16.

8. Lindeberg T. Edge detection and ridge detection with automatic scale selection. *Int J Comput Vision* (1998) 30(2):117–56. doi: 10.1023/A:1008097225773

9. Flores-Tapia D, Thomas G, Venugopal N, McCurdy B, Pistorius S. Semi-automatic MRI prostate segmentation based on wavelet multiscale products, in: *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, IEEE. (2008) pp. 3020–3.

10. Klein S, Heide UAD, Lips IM, Vulpen MV, Staring M, Pluim JP. Automatic segmentation of the prostate in 3D MR images by atlas matching using localized mutual information. *Med Phys* (2008) 35(4):1407–17. doi: 10.1118/1.2842076

11. Gao Y, Sandhu R, Fichtinger G, Tannenbaum AR. A coupled global registration and segmentation framework with application to magnetic resonance prostate imagery. *IEEE Transactions on Medical Imaging* (2010) 29(10):1781–94. doi: 10.1109/TMI.2010.2052065

12. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. *Commun ACM* (2017) 60(6):84–90. doi: 10.1145/3065386

13. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation, in: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany,* October 5-9, 2015, Proceedings, Part III 18. Springer International Publishing. (2015) pp. 234–41.

14. Lai C, Wang HK, Wang FN, Peng YC, Lin TP, Peng HH, et al. Autosegmentation of prostate zones and cancer regions from biparametric magnetic resonance images by using deep-learning-based neural networks. *Sensors* (2021) 1 (8):2709. doi: 10.3390/s21082709

15. Wang Y, Dou H, Hu X, Zhu L, Yang X, Xu M, et al. Deep attentive features for prostate segmentation in 3D transrectal ultrasound. *IEEE Trans Med Imaging* (2019) 38 (12):2768–78. doi: 10.1109/TMI.2019.2913184

16. Chen Y, Xing L, Yu L, Bagshaw HP, Buyyounouski MK, Han B. Automatic intraprostatic lesion segmentation in multiparametric magnetic resonance images with proposed multiple branch UNet. *Med Phys* (2020) 47(12):6421–9. doi: 10.1002/mp.14517

17. Orlando N, Gillies DJ, Gyacskov I, Romagnoli C. Automatic prostate segmentation using deep learning on clinically diverse 3D transrectal ultrasound images. *Med Phys* (2020) 47(6):2413–26. doi: 10.1002/mp.14134

18. He K, Zhang X, Ren S, Sun J. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2015) 37(9):1904–16. doi: 10.1109/TPAMI.2015.2389824

19. Javadinasr M, Mohammadian A, Parsa A. A deep-learning based optimization approach to address stop-skipping strategy in urban rail transit lines, in: *International Conference on Transportation and Development*, ASCE. (2022) pp. 165–76.

20. Huang G, Liu Z, van der Maaten L, Weinberger KQ. Densely connected convolutional networks, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, IEEE. (2017) pp. 4700–8.

21. Woo S, Park J, Lee JY, Kweon IS. Cbam: Convolutional block attention module, in: *Proceedings of the European conference on computer vision (ECCV)*, Springer. (2018) pp. 3–19.

22. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, IEEE. (2015) pp. 3431–40.

23. Zhou Z, Rahman Siddiquee MM, Tajbakhsh N, Liang J. Unet++: A nested u-net architecture for medical image segmentation, in: *Deep learning in medical image analysis and multimodal learning for clinical decision support*. (2018) pp. 3–11.

24. Jha D, Smedsrud PH, Johansen D, de Lange T, Johansen HD, Halvorsen P, et al. A comprehensive study on colorectal polyp segmentation with ResUNet++, conditional random field and test-time augmentation. *IEEE J Biomed Health Inf* (2021) 25(6):2029–40. doi: 10.1109/JBHI.2021.3049304

25. Checcucci E, Autorino R, Cacciamani GE, Amparore D, Cillis S, Piana A, et al. Artificial intelligence and neural networks in urology: current clinical applications, Minerva urologica e nefrologica. *Ital J Urol Nephrol* (2019) 72(1):49–57. doi: 10.23736/s0393-2249.19.03613-0

26. Checcucci E, Rosati S, De Cillis S, Vagni M, Giordano N, Piana A, et al. Artificial intelligence for target prostate biopsy outcomes prediction the potential application of fuzzy logic. *Prostate Cancer Prostatic Dis* (2022) 25(2):359–62. doi: 10.1038/s41391-021-00441-1

27. Wessels F, Kuntz S, Krieghoff-Henning E, Schmitt M, Braun V, S.Worst T, et al. Artificial intelligence to predict oncological outcome directly from hematoxylin & eosin-stained slides in urology: a systematic review. *Minerva Urol Nephrol* (2022). doi: 10.23736/S2724-6051.22.04758-9

28. Zhu Y, Wei R, Gao G, Ding L, Zhang X, Wang X, et al. Fully automatic segmentation on prostate MR images based on cascaded fully convolution network. *J Magn Reson Imaging* (2019) 49(4):1149–56. doi: 10.1002/jmri.26337

29. Duran A, Dussert G, Rouvière O, Jaouen T, Jodoin PM, Lartizien C. ProstAttention-net: A deep attention model for prostate cancer segmentation by aggressiveness in MRI scans. *Med Image Anal* (2022) 77:102347. doi: 10.1016/j.media.2021.102347

30. Song Y, Zhang YD, Yan X, Liu H, Zhou M, Hu B, et al. Computer-aided diagnosis of prostate cancer using a deep convolutional neural network from multiparametric MRI. *J Magnetic Resonance Imaging* (2018) 48(6):1570–7. doi: 10.1002/jmri.26047

31. Cao R, Zhong X, Afshari S, Felker E, Suvannarerg V, Tubtawee T, et al. Performance of deep learning and genitourinary radiologists in detection of prostate cancer using 3-T multiparametric magnetic resonance imaging. *J Magnetic Resonance Imaging* (2021) 54(2):474–83. doi: 10.1002/jmri.27595

Frontiers in Oncology

Advances knowledge of carcinogenesis and tumor progression for better treatment and management

The third most-cited oncology journal, which highlights research in carcinogenesis and tumor progression, bridging the gap between basic research and applications to imrpove diagnosis, therapeutics and management strategies.

Discover the latest Research Topics

See more →

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact