# QUANTITATIVE IMAGING FOR CLINICAL DECISIONS

EDITED BY: Nandita Maria deSouza, Laure S. Fournier and
Daniela Elena Oprea-Lager
PUBLISHED IN: Frontiers in Oncology

**frontiers** Research Topics

## 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

# QUANTITATIVE IMAGING FOR CLINICAL DECISIONS

Topic Editors:
**Nandita Maria deSouza,** Institute of Cancer Research (ICR), United Kingdom
**Laure S. Fournier,** Assistance Publique Hopitaux De Paris, France
**Daniela Elena Oprea-Lager,** Academic Medical Center, Netherlands

# Table of Contents

# Editorial: Quantitative Imaging for Clinical Decisions

Nandita Maria deSouza[1]*, Daniela Elena Oprea-Lager[2] and Laure S. Fournier[3]

[1] Division of Radiotherapy and Imaging, The Institute of Cancer Research and Royal Marsden NHS Foundation Trust, London, United Kingdom, [2] Department of Radiology & Nuclear Medicine, Cancer Centre Amsterdam, Amsterdam University Medical Centers, VU University, Amsterdam, Netherlands, [3] Université de Paris, PARCC, INSERM, Radiology Department, AP-HP, Hopital Europeen Georges Pompidou, Paris, France

Editorial on the Research Topic

**Quantitative Imaging for Clinical Decisions**

Ever since the first captivating X-ray images of Mrs. Roentgen's left hand, medical imaging has been at the heart of clinical decision-making. Over a century later, the explosion in clinically available digital imaging techniques such as computerized tomography (CT), magnetic resonance imaging (MRI) and positron emission tomography (PET) has meant that more objective analysis of images has become desirable to facilitate clinical decisions. Therefore, the demand for quantitative imaging data is increasingly supplementing or sometimes replacing the subjective evaluation of disease visualised on scans. At the simplest level, image quantitation has involved linear measurements of visualised abnormalities and Response Evaluation Criteria In Solid Tumors (RECIST) have been the standard for assessing tumors, their regression, progression and control with treatment. This simple measurement remains at the heart of clinical decision-making in oncology and its role is showcased in a flagship expert statement by Fournier et al "RECIST 20 years on" which discusses the principles underlying RECIST measurements, their reproducibility, limitations and clinical relevance after two decades of use.

The nature of bone lesions has dictated that RECIST measurements are not applicable to the skeleton, so that bony lesions have traditionally been considered non measurable and relied on scoring indices (1). The review by Oprea-Lager et al. challenges this view and describes the newer imaging modalities such as whole-body magnetic resonance imaging (WB-MRI) with diffusion-weighted imaging and positron-emission tomography (PET) including the use of new targeted tracers that open the door to quantifying skeletal pathology. Exploitation of these techniques in order to introduce quantitative imaging for skeletal metastases has been endorsed and advocated by consortia and trial groups (2).

Although ultrasound (US), one of the earliest imaging modalities to be used for clinical decision-making, is not considered directly quantitative, its evaluation of tissue stiffness (a vital clinical sign exploited by clinical palpation) is quantifiable using shear-wave elastography. US shear-wave elastography is emerging as a viable technique (3), particularly in assessing and delineating liver fibrosis and prostate cancer. Hardware and software advances promise that it will be implemented more routinely in clinical practice. The research article by Wei et al investigates its utility as a biomarker for predicting change in biopsy-assigned Gleason score at radical prostatectomy, showing that tissue stiffness can predict upgrading of Gleason score. In future, if performed as part of lesion evaluation prior to US-guided biopsy, this technique has the potential to alter selection of surgical vs. non-surgical management options. Additionally, its use in guiding the biopsy

procedure itself or directing other therapeutic strategies such as US-guided high intensity focused ultrasound (HIFU) may become invaluable.

Tissue characterisation has largely been the domain of MRI with biologically-driven multiparameter evaluation dominating the landscape. In neuroimaging, the literature has been dominated by diffusion-weighted imaging. Li et al. examine the non-Gaussian diffusion in glioblastoma multiforme using diffusion kurtosis imaging to investigate whether any of the derived parameters are significant predictors of overall survival. They show that in the multivariate Cox model, the mean kurtosis in the gadolinium-enhanced gross tumor volume pre-radiotherapy was still a significant predictor of overall survival after adjusting effects of age, tumor tissue methylation status and extent of resection. Tissue characterization with quantifiable image data may also be achieved using dual-energy spectral CT (DESCT) (4). Cao et al. utilise spectral CT of the primary tumor in colorectal cancer to predict lymph node metastases, the detection of which remains a holy grail because size criteria are often poorly predictive of involvement. Their nomogram incorporating clinical and DESCT parameters shows clinical potential in this application, which also has been indicated in other cancer types (5).

More recently, a data-driven approach to biomarker extraction has been advocated using radiomics which probe the shape, first-order statistical and texture features within a region-of-interest. The bibliometric analysis by Ding et al. provides an overview of literature related to radiomics in oncology, highlighting artificial intelligence (AI), segmentation method, and use of radiomics for classification and diagnosis in oncology as hotspots. The reproducibility and statistical method of radiomics research, the relation between genomics and radiomics, and the applications of radiomics in sarcoma and intensity-modulated radiotherapy have been identified as research frontiers in the field. The link between radiomics features and histopathology is explored in *ex-vivo* ovarian cancer tissue using images acquired at 9.4T by Tardieu et al. and illustrates the correlation between radiomics features and stromal proportion. A relationship between tissue compartments has been shown in other studies (6), but the association between these features and stromal proportion on histology potentially offers avenues for understanding the biology of this disease by uncovering the histological changes that occur within individual lesions during tumour regression and progression.

Treatment response is an area where quantitative biomarkers are actively desired for clinical decision-making. Hellwig et al. address this in their study in head and neck cancer and develop a random forest based model with dynamic contrast-enhanced parameters to predict treatment response to induction chemotherapy. This is taken further using three-dimensional convolutional neural networks (CNN) in lung cancer by Hou et al. using deep transfer learning to stratify patients into subgroups with different response and progression risks. Their work illustrates the potential of CNN to stratify progression status in patients with epidermal growth factor receptor (EGFR) mutations treated with first-line tyrosine kinase inhibitors. Although such small single centre studies provide a handle on quantitative biomarker discovery, one of the weakest links in parameter generation is the reliability and reproducibility of the segmentation method. Conventionally, this is done manually by trained observers, but computer-aided segmentation is increasingly used (7). This potentially improves the reliability of segmentation methods. Li et al. examine the reproducibility of a computer-aided contouring tool in tumor measurements, and its impact on evaluation of tumor response in terms of RECIST 1.1 criteria. Their data highlight the improvements in interobserver variability that can be achieved with computer aided contouring, which is particularly evident when assigning patients to response categories, thus profoundly impacting individual patient management with regard to therapeutic decisions.

Quantitative imaging is now available with a variety of imaging techniques and there is an explosion in the wealth of parameters that can be derived, particularly with the advent of data-driven approaches of feature extraction. It is important that as imagers and clinicians we are not seduced by the ever-increasing amount of data available, but rather that we select appropriately the data that is truly meaningful and able to reliably influence our clinical decisions. This demands rigor in deriving, qualifying and validating quantitative biomarkers to advance patient management.

## AUTHOR CONTRIBUTIONS

All authors contributed to the article and approved the submitted version.

## REFERENCES

1. Mota JM, Armstrong AJ, Larson SM, Fox JJ, Morris MJ. Measuring the Unmeasurable: Automated Bone Scan Index as a Quantitative Endpoint in Prostate Cancer Clinical Trials. *Prostate Cancer Prostatic Dis* (2019) 22(4):522–30. doi: 10.1038/s41391-019-0151-4
2. Lecouvet FE, Talbot JN, Messiou C, Bourguet P, Liu Y, de Souza NM, et al. Monitoring the Response of Bone Metastases to Treatment With Magnetic Resonance Imaging and Nuclear Medicine Techniques: A Review and Position Statement by the European Organisation for Research and Treatment of Cancer Imaging Group. *Eur J Cancer* (2014) 50(15):2519–31. doi: 10.1016/j.ejca.2014.07.002
3. Sigrist RMS, Liau J, Kaffas AE, Chammas MC, Willmann JK. Ultrasound Elastography: Review of Techniques and Clinical Applications. *Theranostics* (2017) 7(5):1303–29. doi: 10.7150/thno.18650
4. Simons D, Kachelriess M, Schlemmer HP. Recent Developments of Dual-Energy CT in Oncology. *Eur Radiol* (2014) 24(4):930–9. doi: 10.1007/s00330-013-3087-4
5. Forghani R. An Update on Advanced Dual-Energy CT for Head and Neck Cancer Imaging. *Expert Rev Anticancer Ther* (2019) 19(7):633–44. doi: 10.1080/14737140.2019.1626234
6. Winfield JM, Wakefield JC, Brenton JD, AbdulJabbar K, Savio A, Freeman S, et al. Biomarkers for Site-Specific Response to Neoadjuvant Chemotherapy in Epithelial Ovarian Cancer: Relating MRI Changes to Tumour Cell Load and Necrosis. *Br J Cancer* (2021) 124(6):1130–7. doi: 10.1038/s41416-020-01217-5

7. Vivanti R, Joskowicz L, Karaaslan OA, Sosna J. Automatic Lung Tumor Segmentation With Leaks Removal in Follow-Up CT Studies. *Int J Comput Assist Radiol Surg* (2015) 10(9):1505–14. doi: 10.1007/s11548-015-1150-0

# Survival Prediction Analysis in Glioblastoma With Diffusion Kurtosis Imaging

Yuan Li[1,2], Michelle M. Kim[1], Daniel R. Wahl[1], Theodore S. Lawrence[1], Hemant Parmar[3] and Yue Cao[1,2,3*]

[1] Departments of Radiation Oncology, University of Michigan, Ann Arbor, MI, United States, [2] Department of Biomedical Engineering, University of Michigan, Ann Arbor, MI, United States, [3] Department of Radiology, University of Michigan, Ann Arbor, MI, United States

**Simple Summary:** Glioblastoma (GBM) is the most common and aggressive primary brain tumor. Diffusion kurtosis imaging (DKI) has characterized non-Gaussian diffusion behaviors in brain normal tissue and gliomas, but there are very limited efforts in investigating treatment responses of kurtosis in GBM. This study aimed to investigate whether any parameter derived from the DKI is a significant predictor of overall survival (OS). We found that the large mean, 80 and 90 percentile kurtosis values in the contrast enhanced gross tumor volume (Gd-GTV) on post-Gd T1-weighted images pre-RT were significantly associated with reduced OS. In the multivariate Cox model, the mean kurtosis Gd-GTV pre-RT after considering effects of age, extent of surgery, and methylation were significant predictors of OS. In addition, the 80 and 90 percentile kurtosis values in Gd-GTV post RT were significantly associated with progression free survival (PFS). The DKI model demonstrates the potential to predict outcomes in the patients with GBM.

**Purpose:** Non-Gaussian diffusion behaviors in gliomas have been characterized by diffusion kurtosis imaging (DKI). But there are very limited efforts in investigating the kurtosis in glioblastoma (GBM) and its prognostic and predictive values. This study aimed to investigate whether any of the diffusion kurtosis parameters derived from DKI is a significant predictor of overall survival.

**Methods and Materials:** Thirty-three patients with GBM had pre-radiation therapy (RT) and mid-RT diffusion weighted (DW) images. Kurtosis and diffusion coefficient (DC) values in the contrast enhanced gross tumor volume (Gd-GTV) on post-Gd T1 weighted images pre-RT and mid-RT were calculated. Univariate and multivariate Cox models were used to evaluate the DKI parameters and clinical factors for prediction of OS and PFS.

**Results:** The large mean kurtosis values in the Gd-GTV pre-RT were significantly associated with reduced OS ($p = 0.02$), but the values at mid-RT were not ($p > 0.8$). In the multivariate Cox model, the mean kurtosis in the Gd-GTV pre-RT ($p = 0.009$) was still a significant predictor of OS after adjusting effects of age, O6-Methylguanine-DNA Methyl transferase (MGMT) methylation and extent of resection. In Gd-GTV post-RT, 80 and 90 percentile kurtosis values were significant predictors ($p \leq 0.05$) for progression free survival (PFS).

**Conclusion:** The DKI model demonstrates the potential to predict OS and PFS in the patients with GBM. Further development and histopathological validation of the DKI model will warrant its role in clinical management of GBM.

Keywords: diffusion kurtosis imaging, diffusion MRI, glioblastoma, survival prediction, imaging analysis

# INTRODUCTION

Glioblastoma (GBM) is the most common and aggressive primary brain tumor in adults and has a poor prognosis with a median survival of approximately 14 months despite multimodality therapy with surgery, concurrent chemoradiation therapy, and adjuvant chemotherapy (1, 2). Standard clinical assessment of tumor progression or therapy response (3) is based primarily on post-contrast T1-weighted and fluid-attenuated inversion recovery (FLAIR) T2-weighted magnetic resonance images (MRI). There are some challenges to these conventional techniques. The contrast enhancement on the post-contrast T1-weighted MRI is affected by tumor growth, but also radiation, anti-angiogenesis drugs, and chemotherapy, all of which can be attributed to blood–brain barrier disruption. Abnormality on T2 FLAIR images is influenced by T2 changes of tumor cells as well as by edema that co-exists within GBM or is affected by radiation therapy. Limitations of conventional MRI in clinical management of GBM have motivated investigations of physiological and metabolic MRI.

Diffusion weighted (DW) imaging has been proposed to overcome these limitations. DW imaging is a technique to measure water molecule mobility in the microscopic tissue environment and is sensitive to cell density and size, cell membrane permeability, and extracellular space tortuosity. Apparent diffusion coefficient (ADC) quantified from conventional DW images fitted to a mono-exponential function is the commonly reported parameter in literature. The correlation between high cellularity and low ADC in tumor animal models and human cancers motivates investigations on roles of ADC in clinical GBM (4–7). However, heterogeneous tissue in GBM, especially edema, often results in elevated ADC compared to normal white matter (WM) (1) and gray matter (GM). To overcome this limitation, high b-value DW images and high-order diffusion models have been explored in clinical gliomas to differentiate tumor grade and assess therapy response (8–17). Diffusion tensor imaging (DTI) is an emerging technique to investigate brain tumor. Fractional anisotropy (FA) that derived from DTI has been suggested to provide information of cell density. A previous study of FA in GBM showed that FA was low in GBM and suggested that the directional diffusion has been corrupted in the tumor region (18). However, how to differentiate low FA caused by tumor from that affected by edema is a challenge.

The signal-to-noise ratio of diffusion weighted images acquired on clinical scanners is a limiting factor in the application of high-order diffusion models to GBM. Diffusion kurtosis imaging (DKI) is an emerging approach to estimate the non-Gaussian water diffusion behavior over high b values in tissue. DKI has shown the potential to characterize normal and pathologic tissue (17, 19). Previous research has suggested that DKI provides better separation of brain tumor grades (14, 17, 20), but there are very limited efforts in investigating treatment responses of kurtosis in GBM and its prognostic and predictive values for patient survival (21).

In this study, we hypothesized that high diffusion kurtosis in GBM correlated with decreased OS. We applied the diffusion kurtosis model to the DW images acquired in the patients with GBM before radiation therapy (pre-RT), during the course of RT (mid-RT) and after radiation therapy (post-RT). We analyzed the parameter differences between pre-RT and mid-RT to investigate the bio-physical meaning of the parameters and response to RT. Finally, we tested whether any parameter derived from the model is a significant predictor of overall survival (OS).

# MATERIALS AND METHODS

## Patients

Thirty-three patients with histologically confirmed, newly diagnosed GBM were enrolled on prospective, institute-review-board approved protocols. All patients signed written informed consent. The patients had research MRI scans, including anatomic scans and diffusion weighted (DW) images, pre-RT following maximal tumor surgical resection prior to chemo-radiation therapy (CRT) and during the 3rd–4th week of CRT (mid-RT). The twenty-one patients had the research MRI scans 3-month post-RT. The ten patients were treated based upon the institution protocol of concurrent CRT following chemotherapy with a median dose of 60 Gy (40.05–72 Gy), and the 23 patients were enrolled on a prospective radiation boosting clinical trial and treated to 75 Gy (NCT02805179) (22). All patients received concurrent temozolomide.

## *In Vivo* MR imaging

All MRI scans were performed on a 3.0-T scanner (Skyra, Siemens Healthineers) using a 20-channel head coil. Conventional MR images, 2D T2-FLAIR images, and 3D pre- and post-contrast T1-weighted images using a MPRAGE sequence, were acquired. DW images were acquired by a spin-echo echo-planar pulse sequence with diffusion weighting in three orthogonal directions and 11 b-values from 0 to 2,500 s/mm$^2$ with an incremental step of 250 s/mm$^2$. Other acquisition parameters included a parallel imaging factor of 4 (GRAPPA) (to reduce echo spacing and hence geometric distortion), TE/TR = 93/9,300 ms, bandwidth of 1,040 Hz/pixel, voxel size of approximately 1.3 × 1.3 × 5.2 mm, 30 slices to cover the whole

brain, one average and total scan time of 4.50 min. All DW images were acquired prior to contrast injection.

## Diffusion Model

The diffusion kurtosis model analyzes non-Gaussian water diffusivity with equation:

$$S = S_0 * e^{(-b*D+\frac{1}{6}(b*D)^2*K)} \qquad (1)$$

where $S_0$ is an amplitude of diffusion signals, $D$ is a diffusion coefficient (DC) that is corrected for the observed non-Gaussian diffusion behavior and $K$ represents an apparent diffusional kurtosis. Here, we did not consider an anisotropic diffusion kurtosis in GBM due to the low anisotropic diffusion behavior in the contrast-enhanced tumor volume.

## Computation of Kurtosis and DC Maps

Kurtosis and DC maps were generated from DW images with 11 b-values using in-house Functional Image Analysis Tools (imFIAT). We first took a logarithm of diffusion signals, and then used Simplex algorithm to fit the model. In the computation process, a 2D $3 \times 3$ Gaussian filter and brain mask were first applied to all phases of diffusion weighted images to reduce noise influence on the parameter maps.

The gadolinium enhancement gross tumor volumes (Gd-GTV) on post-Gd T1 weighted images were delineated by radiation oncologists who treated the patients. Surgical cavities were removed from the Gd-GTV. The median of the residual Gd-GTV is 20.97 cm$^3$ (ranges from 2.33 to 62.50 cm$^3$). In eighteen patients with gross total resection (**Table 1**), the median of the residual Gd-GTV (excluding the surgical cavity) was 14.00 cm$^3$ (ranges from 2.33 to 46.00 cm$^3$).

**TABLE 1 |** Patients characteristics.

| Count | N |
|---|---|
| Patients | 33 |
| **Age** | |
| Median (IQR) | 61 (50, 79) |
| **Gender** | |
| Female | 13 (39.4%) |
| Male | 20 (60.6%) |
| **ECOG** | |
| 0 | 7 (21.2%) |
| 1 | 23 (69.7%) |
| 2 | 3 (9.1%) |
| **Median physical dose** | |
| Institute protocol | 60 (40.05, 72) |
| Boosting protocol | 75 (75, 75) |
| **Extent of surgery** | |
| Biopsy | 6 (18.2%) |
| Subtotal resection | 9 (27.3%) |
| Gross total resection | 18 (54.5%) |
| **MGMT methylation** | |
| Positive | 9 (27.2%) |
| Negative | 22 (66.7%) |
| Unknown | 2 (6.1%) |
| **IDH status** | |
| Mutant | 1 (3%) |
| Wild type | 31 (94%) |
| Unknown | 1 (3%) |

Considering GBM is a heterogeneous tumor with edema (possible low cellular density) and high cellular components, a mean value of kurtosis or diffusion coefficient averaged over the whole volume of Gd-GTV-cavity could wash out the component that could be more aggressive and predict outcomes. Therefore, we attempted to analyze the part of the histogram of kurtosis or diffusion coefficient, which is associated with the aggressive tumor. Since high kurtosis values and low diffusion coefficients are associated with tumor aggressiveness, we choose high percentiles of kurtosis and low percentiles of DC to test whether they predicted OS. Therefore, mean, 80 and 90 percentile values of kurtosis, and mean, 10 and 20 percentile values of DC in the Gd-GTV pre-RT and mid-RT were calculated.

## Statistical Analysis

The primary endpoint of the study was to determine whether the DKI parameters provide additional predictive values over clinical variables for OS. OS was defined as the interval from the start of RT to death from any cause. Patients were censored at the time of last contact or clinical follow-up, whichever occurred last. Patients were generally followed every 8 weeks after chemoradiation with clinical exam and MRI. Progression-free survival (PFS) was defined as the interval from the start of RT to progression or death, whichever occurred first, and patients were censored at the time of last imaging follow-up. Progression was determined by a multidisciplinary tumor board, and worsened enhancement within 3 months of chemoradiation was generally managed by repeat imaging to rule out pseudoprogression. Progression was defined as worsened enhancement outside of the radiation field, or within the radiation field if progression was confirmed pathologically or with serial confirmatory imaging and clinical evaluation, or by change in therapy (*i.e.* initiation of next-line chemotherapy), whichever occurred first.

PFS and OS were calculated using Kaplan–Meier method. To test predictive values of the DKI parameters, univariate Cox proportional hazards model first was used to evaluate each of the DKI parameters as well as clinical factors for prediction of OS and PFS.

Clinical factors included age (continuous), sex, ECOG performance status (0 *vs.* 1 *vs.* 2), radiation dose (continuous), extent of resection (EOR, gross total resection = 2, subtotal resection = 1, or biopsy = 0), MGMT methylation status (methylated *vs.* unmethylated), and baseline contrast enhanced gross tumor volume (GTV-Gd). Multivariate Cox proportional hazards model was further performed to test whether the DKI parameters could provide additional values to clinical factors for prediction of OS and PFS, adjusting age, MGMT methylation status and EOR. The changes in the DKI parameters at mid-RT compared to pre-RT were also tested using a paired t test. A P-value <0.05 was considered significant.

## RESULTS

## Patient Characteristics and Outcomes

Thirty-three patients who had newly diagnosed GBM treated between October 2012 and December 2018 and had the diffusion

imaging scans pre-RT, mid-RT and post-RT as described in the section *In Vivo MR Imaging* were included in this analysis. The patient characteristics are provided in **Table 1**. The median age was 61 years old (50–79). Thirteen patients were female. ECOG performance status of thirty patients was 0–1. Eighteen patients had total surgical resection, nine had subtotal resection and six had biopsy only. Eight of the 31 patients who had MGMT methylation tests were methylated, and one of the 32 patients who had IDH tests had the mutated type.

Fourteen patients were still alive with a median follow-up of 17.4 months (9.07–49.4 months). The median survival was 13.7 months (0.6–37.5 months). Twenty-five patients progressed with a median progression of 8 months (0.6–25 months); one patient progressed (3 weeks) at mid-RT. **Figure 1** shows Kaplan–Meier curves of OS and PFS.

## Kurtosis and DC Values in the Gd-GTV Pre-RT, Mid-RT and Post-RT

Kurtosis and DC maps of the 33 patients pre-RT and mid-RT were calculated. An example of kurtosis maps and the diffusion curve in the Gd-GTV is shown in **Figure 2**. Note that the kurtosis values in the Gd-GTV were heterogeneous. We investigated the mean kurtosis values in the Gd-GTV as well as the 80 and 90 percentile values pre-RT and mid-RT. Similarly, we investigated the mean DC, 10 and 20 percentile values in the Gd-GTV pre-RT and mid-RT. All data are summarized in **Table 2**.

The kurtosis values and DC values in the Gd-GTV at mid-RT decreased and increased significantly (P-value <0.005) compared to pre-RT, respectively (**Figure 3**). The three outlier data points in the kurtosis plot that did not follow the decrease group trends from pre-RT to mid-RT were from one patient who had rapid
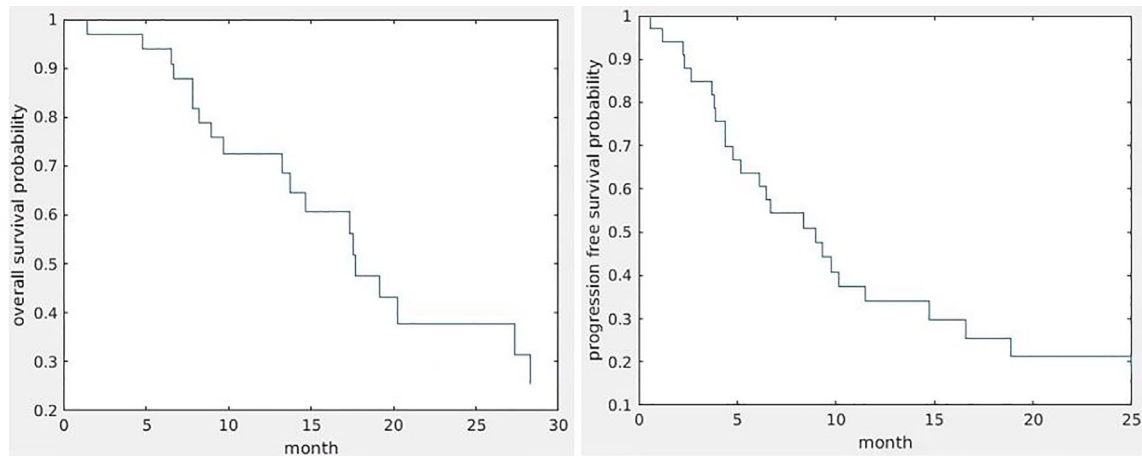


**FIGURE 1** | Kaplan–Meier curves of OS (left) and PFS (right).
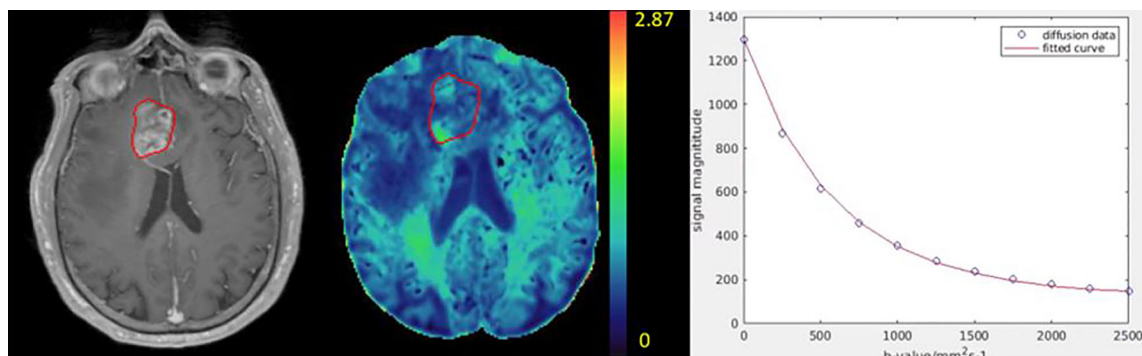


**FIGURE 2** | Illustration of a kurtosis map (color-coded, middle) of a patient with GBM. The color bar indicates kurtosis values. The post-Gd tumor volume (Gd-GTV, red contour) delineated on T1-weighted images (left) is overlaid on the kurtosis map. An example of diffusion weighted signals fitted by the diffusion kurtosis model is shown in the right panel. Blue dots represent original diffusion signal data in the Gd-GTV, and red solid line is the fitted curve. Note that the diffusion kurtosis model fits the diffusion signals well.

**TABLE 2 |** Kurtosis and DC values in the Gd-GTV pre-RT, mid-RT and post-RT.

| | Pre-RT | Mid-RT | Post-RT |
|---|---|---|---|
| Mean Kurtosis ± SD | 0.76 ± 0.10 | 0.73 ± 0.18 | 0.65 ± 0.14 |
| 80 percentile Kurtosis ± SD | 1.07 ± 0.18 | 0.98 ± 0.27 | 1.04 ± 0.67 |
| 90 percentile Kurtosis ± SD | 1.18 ± 0.24 | 1.07 ± 0.31 | 1.23 ± 1.04 |
| Mean DC (um$^2$/ms) ± SD | 1.54 ± 0.30 | 1.67 ± 0.34 | 1.71 ± 0.43 |
| 10 percentile DC (um$^2$/ms) ± SD | 0.89 ± 0.13 | 1.07 ± 0.17 | 1.81 ± 0.47 |
| 20 percentile DC(um$^2$/ms) ± SD | 1.02 ± 0.15 | 1.20 ± 0.20 | 1.89 ± 0.51 |
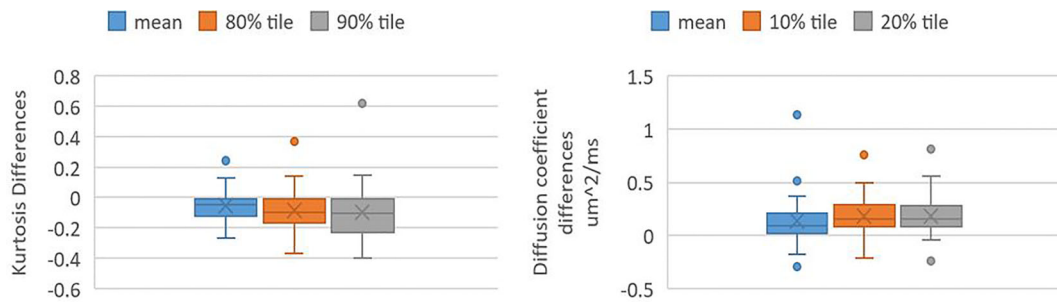


**FIGURE 3 |** Box and whisker plots shows values of kurtosis differences and DC differences in Gd-GTV pre-RT and mid-RT (mid-RT values–pre-RT values). Left panel shows kurtosis differences of mean, 80 and 90 percentile kurtosis values. Right panel shows DC differences of mean, 10 and 20 percentile DC values.

progression after treatment. In the DC plot, five outliers that deviated from the group trend came from three patients and were due to necrosis, tumor infiltration in the ventricle or adjacent to the surgical cavity.

The post-RT mean kurtosis, 80 and 90 percentile values, and mean DC, 10 and 20 percentile values in the Gd-GTV of the 21 patients are summarized in **Table 2**. Note that the post-RT mean kurtosis and DC values continued decreasing and increasing from the values from mid-RT, respectively. The large variances of kurtosis and DC over the group could be due to progression observed in two patients at 3 months post-RT.

## Correlation of Parameters With OS and PFS

Univariate Cox model analysis showed that large mean, 80 and 90 percentile kurtosis values in the Gd-GTV pre-RT were significantly associated with reduced OS (respective HR = 2.10, p = 0.03; HR = 2.29, p = 0.03; and HR = 2.30, p = 0.03; **Table 4**),

but not the values measured at mid-RT (p-value >0.83) and post-RT (p-value >0.47). The DC values including the mean, 10 and 20 percentile from the Gd-GTV pre-RT, mid-RT and post-RT did not show any significant association with OS (p-value >0.3). Univariate Cox model analysis of clinical prognostic factors and dose for prediction of OS are listed in **Table 3**.

We further addressed the question whether kurtosis values in the Gd-GTV added any values than clinical prognostic factors, such as EOR, age, MGMT and Gd-GTV, for prediction of OS. Due to the similarity that exists among mean, 80 and 90-percentile kurtosis values in the Gd-GTV, we only selected the mean kurtosis in the analysis. After adjusting these clinical factors, the mean kurtosis value pre-RT was a significant predictor of OS (HR = 3.06, p < 0.009), see **Table 4**.

The mean, 80 and 90 percentile kurtosis values and the DC values in the Gd-GTV pre-RT and mid-RT were not significant predictors for PFS (p >0.5) using univariate Cox model analysis. However, the post-RT values of kurtosis, specifically, the large

**TABLE 3 |** Univariate Cox model analysis of DKI parameters and clinical factors for prediction of OS.

| Parameters | Hazard ratio (HR) | p-value | 95% CI |
|---|---|---|---|
| Mean K pre-RT | 2.10 | 0.03* | [1.10, 4.02] |
| 80 percentile K pre-RT | 2.29 | 0.03* | [1.10, 4.71] |
| 90 percentile K pre-RT | 2.30 | 0.03* | [1.07, 4.96] |
| Gd-GTV pre-RT | 0.74 | 0.25 | [0.44, 1.23] |
| Age | 1.72 | 0.14 | [0.84, 3.52] |
| MGMT | 0.45 | 0.2 | [0.14, 1.47] |
| Dose | 1.20 | 0.07 | [0.98, 1.46] |
| EOR | 0.34 | 0.52 | [0.63,2.52] |

*Significant with p < 0.05. The continuous data were normalized.*

**TABLE 4 |** Multivariate cox model analysis of clinical factors and MK for prediction of OS.

| parameters | Hazard ratio (HR) | p-value | 95% CI |
|---|---|---|---|
| Age | 2.92 | 0.03* | [1.08, 7.94] |
| MGMT | 0.25 | 0.09 | [0.05, 1.24] |
| EOR | 0.55 | 0.21 | [0.21, 1.42] |
| mean K pre-RT | 3.06 | 0.009* | [1.32, 7.13] |

*Significant with p < 0.05. The continuous variables were normalized to their means and standard deviations.

values of kurtosis at the 80 and 90 percentile in the Gd-GTV were associated with reduced PFS (p = 0.05) in the univariate Cox model analysis (p = 0.03 and p = 0.05, respectively), see **Table 5**, which could be a useful indicator for time of progression.

## DISCUSSION

In this study, we investigated the diffusion kurtosis model and characterized non-Gaussian diffusion properties in the Gd-GTVs pre-RT, mid-RT and post-RT in the patients with GBM. We found that the mean kurtosis value in the Gd-GTV pre-RT was significantly prognostic of OS as a high mean kurtosis was associated with inferior of survival. Also, the diffusion kurtosis added a predictive value to the extent of surgery, age and methylation status for survival. The post-RT kurtosis values in the Gd-GTV predicted time to progression. In addition to glioma grading (14, 17, 20), the kurtosis model has potential to aid in conventional MRI for outcome prognosis. Further validation with another cohort of patients will warrant the role of the diffusion kurtosis model in the clinical management of GBM.

Many diffusion models have been investigated in gliomas. An apparent diffusion coefficient quantified from conventional DW images with b-values between 0 and 1,000 s/mm$^2$ using a mono-exponential decay is the commonly reported parameter in literature. Previous studies have suggested that a low ADC was associated with a decrease in survival for patients with gliomas (23–27). One limitation of the mono-exponential model is that there are large deviations of fitted curves from the diffusion weighted signals with b-values greater than 1,500 s/mm$^2$. Another problem is that with a single diffusion parameter is hard to describe the complex microstructure effects on water diffusion. To deal with the deviation of diffusion weighted signals from the mono-exponential function, a bi-exponential model with fast and slow diffusion components has been proposed (11). In the initial interpretation of the bi-exponential model, fast and slow diffusion coefficients are considered from respective extra- and intra-cellular water compartments, but the estimated fraction of the intra-cellular water in the tissue from the

bi-exponential model cannot be matched with that measured by other methods (9). The bi-exponential model fits the diffusion curves better than the mono-exponential model. A study suggests that the fast diffusion coefficient is close to the reported human brain diffusion coefficient (28). To fit the bi-exponential model, it is necessary to take diffusion weighted images with more b-values, which increases the acquisition time. In addition, the bi-exponential model that fits four parameters is unstable to noise, which makes it difficult to generate high quality voxel-by-voxel brain maps.

In addition to the mono and bi-exponential models, other high order diffusion models that have been investigated in clinical gliomas, such as the fractional order calculus model (FROC) and restricted diffusion model (RDM) (12, 18). Those high-order diffusion models require diffusion weighted images with more b-values and high SNR. The FROC requires b-values up to 4,000 s/mm$^2$, and the diffusion coefficient in the model is pre-determined by fitting a mono-exponential model before fitting the entire model (12), which may lead to some errors in parameters. The RDM is insensitive to intracellular diffusion coefficient and is instable to voxel fitting (18), which leads to difficulty in generating parameter maps in the patients with GBM. The diffusion kurtosis model improves the goodness of fit and is more stable than those high-order diffusion models (12, 18). In addition, the kurtosis model has been investigated in clinical gliomas (20). Research suggests that mean kurtosis shows better separation of glioma grades than fractional anisotropy and mean diffusivity. Overall, the kurtosis model is convenient to generate voxel maps and provides the potential measurement of non-Gaussian diffusion in GBM.

When considering underlying of tissue morphology and physiology of diffusion parameters, low ADC is considered to correlate with high cellularity. However, co-existence of edema and high vascularity in a single pixel of the tumor can elevate ADC compared to normal white matter and gray matter. To mitigate the influence of perfusion on measured diffusion coefficients, a bi-exponential model that quantifies fast and slow DCs has been investigated. The fast DC derived from the model is found to be significantly higher in high-grade gliomas

**TABLE 5 |** Univariate Cox model analysis of DKI parameters post-RT for prediction of PFS.

| Parameters | Hazard ratio (HR) | p-value | 95% CI |
|---|---|---|---|
| Mean K post-RT | 1.85 | 0.10 | [0.88, 3.88] |
| 80-percentile K post-RT | 2.18 | 0.03* | [1.10, 4.30] |
| 90-percentile K post-RT | 1.82 | 0.05* | [1.00, 3.33] |

*Significant with p < 0.05. The continuous data were normalized.

than in low-grade gliomas (20), which could be due to hyper-vascularization in the high-grade gliomas. One limitation of the bi-exponential model is that the fraction of slow DC component has discordance with microstructure parameters, *e.g.*, the fraction of intra-cellular water. Some investigations suggest that the discordance may result from the restricted cell membrane and cell size (29, 30). The RDM considers restricted intracellular diffusion and modulations of diffusion gradients into the model (16, 18). To obtain accurate estimations of the apparent cell radius and the extracellular diffusion coefficient derived from the RDM in the GBM and brain normal tissue requires short diffusion times that may be beyond the clinical scanner hardware. The heterogeneous tissue could present even more challenges for the model (18). The FROC model shows that DC, fractional order and spatial parameter all differentiate high-grade pediatric brain tumors from low-grade ones (12). In addition, the fractional order has high predictive values for tumor outcomes (12). There are also some limitations of the FROC model. First, the parameters derived from the model may not differentiate tumor from normal tissue (12). Another challenge is that parameters are not sensitive enough to generate high contrast maps (12).

Previous research has suggested that the mean kurtosis could serve as the optimal parameter for grading glioma in practice (20). Zhang et al. investigated the correlation between OS and kurtosis in high grade gliomas, including grade III and grade VI, and found that mean kurtosis of glioma was a significant predictor of OS (21). Hempel et al. also assessed whether mean kurtosis was a prognostic factor in grade II, grade III, and grade IV gliomas, and found PFS and OS were significantly better in patients with lower mean kurtosis (31). However, different grades of gliomas could have specific features, which may contribute to prediction power. In our analysis, we only included grade IV glioma.

In this study, we found that high mean kurtosis values pre-RT were significantly correlated with reduced OS. To illustrate the unique contribution of the mean kurtosis, we also tested the Gd-GTV for prediction of survival using the cox model. We found that the Gd-GTV volume itself did not predict OS, which suggests that the mean kurtosis provides information beyond the enhanced tumor volume. In the Gd-GTV that consists of heterogeneous tumor with mixture of high cellular tumor and edema, the kurtosis values vary from high to low. The region with high kurtosis values may imply an aggressive component in the tumor, which is supported by the observations: 1) a higher grade of gliomas associated with higher mean kurtosis values (17, 20) and 2) high mean kurtosis values in GBM associated with inferior survival. The decreased mean kurtosis and the increased DC in the Gd-GTV of GBM after receiving radiation treatment are expected to represent a tumor response to therapy, but not specific enough to predict outcomes. Radiation likely causes cell degeneration and necrosis (32), which may decrease mean kurtosis and increase DC to an extent for some GBMs. In contrast, we observed substantially increased mean kurtosis and decreased DC at the mid-RT in two patients who had rapid tumor growth. Further research could be carried out to investigate pathology associated with mean kurtosis changes.

## LIMITATIONS

The DKI model quantifies non-Gaussian water diffusion in heterogeneous tissue and demonstrates the potential to predict OS in GBM patients. However, there are some limitations in the current study. First, we used 11 b-values up to 2,500 s/mm$^2$, which increase acquisition time. Also, the model is sensitive to noise. To overcome the noise influence on model fitting, we applied a 2D Gaussian filter that blurs images. Second, the mean kurtosis decreased while the DC increased in mid treatment, but these changes are not significantly associated with survival. This may be affected by radiation treatment or the small patient sample size. Third, this is a retrospective analysis with a small sample size. Fourth, pathology correlated to the imaging finding is lacking in our research. Understanding of the mean kurtosis and DC changes after radiation and relationship to tumor changes is limited. The DKI model needs to be further validated in an independent large cohort of patients in future.

## CONCLUSIONS

The DKI model demonstrates the potential to predict OS and PFS in the patients with GBM. The model needs to be further investigated with pathologic correlation and validated in an independent large cohort of patients in the future. Further development and histopathological validation of the DKI model will warrant its role in clinical management of GBM.

## 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 IRB. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

YC contributed to conception and design of the study. MK, DW, and HP helped collect data. YL performed the imaging analysis and statistical analysis. YL also wrote the first draft of the manuscript. YC, MK, and TL revised the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

## FUNDING

# 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. Prados MD, Chang SM, Butowski N, DeBoer R, Parvataneni R, Carliner H, et al. Phase II Study of Erlotinib Plus Temozolomide During and After Radiation Therapy in Patients With Newly Diagnosed Glioblastoma Multiforme or Gliosarcoma. *J Clin Oncol* (2009) 27(4):579–84.

3. Wen PY, Macdonald DR, Reardon DA, Cloughesy TF, Sorensen AG, Galanis E, et al. Updated Response Assessment Criteria for High-Grade Gliomas: Response Assessment in Neuro-Oncology Working Group. *J Clin Oncol* (2010) 28(11):1963–72. doi: 10.1200/JCO.2009.26.3541

4. Chenevert TL, Stegman LD, Taylor JMG, Robertson PL, Greenberg HS, Rehemtulla A, et al. Diffusion Magnetic Resonance Imaging: An Early Surrogate Marker of Therapeutic Efficacy in Brain Tumors. *J Natl Cancer I* (2000) 92(24):2029–36. doi: 10.1093/jnci/92.24.2029

5. Guo AC, Cummings TJ, Dash RC, Provenzale JM. Lymphomas and High-Grade Astrocytomas: Comparison of Water Diffusibility and Histologic Characteristics. *Radiology* (2002) 224(1):177–83. doi: 10.1148/radiol.2241010637

6. Lyng H, Haraldseth O, Rofstad EK. Measurement of Cell Density and Necrotic Fraction in Human Melanoma Xenografts by Diffusion Weighted Magnetic Resonance Imaging. *Magnet Reson Med* (2000) 43(6):828–36. doi: 10.1002/1522-2594(200006)43:6<828::AID-MRM8>3.0.CO;2-P

7. Sugahara T, Korogi Y, Kochi M, Ikushima I, Shigematu Y, Hirai T, et al. Usefulness of Diffusion-Weighted MRI With Echo-Planar Technique in the Evaluation of Cellularity in Gliomas. *JMRI-J Magn Reson Im* (1999) 9(1):53–60. doi: 10.1002/(SICI)1522-2586(199901)9:1<53::AID-JMRI7>3.0.CO;2-2

8. Kwee TC, Galban CJ, Tsien C, Junck L, Sundgren PC, Ivancevic MK, et al. Intravoxel Water Diffusion Heterogeneity Imaging of Human High-Grade Gliomas. *NMR BioMed* (2010) 23(2):179–87. doi: 10.1002/nbm.1441

9. Maier SE, Bogner P, Bajzik G, Mamata H, Mamata Y, Repa I, et al. Normal Brain and Brain Tumor: Multicomponent Apparent Diffusion Coefficient Line Scan Imaging. *Radiology* (2001) 219(3):842–9. doi: 10.1148/radiology.219.3.r01jn02842

10. Mardor Y, Pfeffer R, Spiegelmann R, Roth Y, Maier SE, Nissim O, et al. Early Detection of Response to Radiation Therapy in Patients With Brain Malignancies Using Conventional and High B-Value Diffusion-Weighted Magnetic Resonance Imaging. *J Clin Oncol* (2003) 21(6):1094–100. doi: 10.1200/JCO.2003.05.069

11. Niendorf T, Dijkhuizen RM, Norris DG, Campagne MV, Nicolay K. Biexponential Diffusion Attenuation in Various States of Brain Tissue: Implications for Diffusion-Weighted Imaging. *Magnet Reson Med* (1996) 36(6):847–57. doi: 10.1002/mrm.1910360607

12. Sui Y, Wang H, Liu G, Damen FW, Wanamaker C, Li Y, et al. Differentiation of Low- and High-Grade Pediatric Brain Tumors With High B-Value Diffusion-Weighted MR Imaging and a Fractional Order Calculus Model. *Radiology* (2015) 277(2):489–96. doi: 10.1148/radiol.2015142156

13. Szczepankiewicz F, van Westen D, Englund E, Westin CF, Stahlberg F, Latt J, et al. The Link Between Diffusion MRI and Tumor Heterogeneity: Mapping Cell Eccentricity and Density by Diffusional Variance Decomposition (DIVIDE). *Neuroimage* (2016) 142:522–32. doi: 10.1016/j.neuroimage.2016.07.038

14. Van Cauter S, Veraart J, Sijbers J, Peeters RR, Himmelreich U, De Keyzer F, et al. Gliomas: Diffusion Kurtosis MR Imaging in Grading. *Radiology* (2012) 263(2):492–501. doi: 10.1148/radiol.12110927

15. White NS, McDonald CR, Farid N, Kuperman JM, Kesari S, Dale AM. Improved Conspicuity and Delineation of High-Grade Primary and Metastatic Brain Tumors Using "Restriction Spectrum Imaging": Quantitative Comparison With High B-Value DWI and ADC. *AJNR Am J Neuroradiol* (2013) 34(5):958–64, S951. doi: 10.3174/ajnr.A3327

16. Jiang X, Li H, Devan SP, Gore JC, Xu J. MR Cell Size Imaging With Temporal Diffusion Spectroscopy. *Magn Reson Imaging* (2020) 77:109–123. doi: 10.1016/j.mri.2020.12.010

17. Raab P, Hattingen E, Franz K, Zanella FE, Lanfermann H. Cerebral Gliomas: Diffusional Kurtosis Imaging Analysis of Microstructural Differences. *Radiology* (2010) 254(3):876–81. doi: 10.1148/radiol.09090819

18. Li Y, Kim M, Lawrence TS, Parmar H, Cao Y. Microstructure Modeling of High B-Value Diffusion-Weighted Images in Glioblastoma. *Tomography* (2020) 6(1):34–43. doi: 10.18383/j.tom.2020.00018

19. Wu EX, Cheung MM. MR Diffusion Kurtosis Imaging for Neural Tissue Characterization. *NMR BioMed* (2010) 23(7):836–48. doi: 10.1002/nbm.1506

20. Bai Y, Lin Y, Tian J, Shi D, Cheng J, Haacke EM, et al. Grading of Gliomas by Using Monoexponential, Biexponential, and Stretched Exponential Diffusion-Weighted MR Imaging and Diffusion Kurtosis MR Imaging. *Radiology* (2016) 278(2):496–504. doi: 10.1148/radiol.2015142173

21. Zhang J, Jiang J, Zhao L, Zhang J, Shen N, Li S, et al. Survival Prediction of High-Grade Glioma Patients With Diffusion Kurtosis Imaging. *Am J Transl Res* (2019) 11(6):3680–8.

22. Kim MM, Sun Y, Aryal MP, Parmar H, Piert M, Rosen BS, et al. A Phase II Study of Dose-Intensified Chemoradiation Using Biologically-Based Target Volume Definition in Patients With Newly Diagnosed Glioblastoma. *Int J Radiat Oncol* (2020) 108(3):S20–S20. doi: 10.1016/j.ijrobp.2020.07.2105

23. Zulfiqar M, Yousem DM, Lai H. ADC Values and Prognosis of Malignant Astrocytomas: Does Lower ADC Predict a Worse Prognosis Independent of Grade of Tumor?–A Meta-Analysis. *AJR Am J Roentgenol* (2013) 200(3):624–9. doi: 10.2214/AJR.12.8679

24. Yamasaki F, Sugiyama K, Ohtaki M, Takeshima Y, Abe N, Akiyama Y, et al. Glioblastoma Treated With Postoperative Radio-Chemotherapy: Prognostic Value of Apparent Diffusion Coefficient at MR Imaging. *Eur J Radiol* (2010) 73(3):532–7. doi: 10.1016/j.ejrad.2009.01.013

25. Murakami R, Sugahara T, Nakamura H, Hirai T, Kitajima M, Hayashida Y, et al. Malignant Supratentorial Astrocytoma Treated With Postoperative Radiation Therapy: Prognostic Value of Pretreatment Quantitative Diffusion-Weighted MR Imaging. *Radiology* (2007) 243(2):493–9. doi: 10.1148/radiol.2432060450

26. Higano S, Yun X, Kumabe T, Watanabe M, Mugikura S, Umetsu A, et al. Malignant Astrocytic Tumors: Clinical Importance of Apparent Diffusion Coefficient in Prediction of Grade and Prognosis. *Radiology* (2006) 241(3):839–46. doi: 10.1148/radiol.2413051276

27. Cui Y, Ma L, Chen X, Zhang Z, Jiang H, Lin S. Lower Apparent Diffusion Coefficients Indicate Distinct Prognosis in Low-Grade and High-Grade Glioma. *J Neurooncol* (2014) 119(2):377–85. doi: 10.1007/s11060-014-1490-6

28. Mulkern RV, Zengingonul HP, Robertson RL, Bogner P, Zou KH, Gudbjartsson H, et al. Multi-Component Apparent Diffusion Coefficients in Human Brain: Relationship to Spin-Lattice Relaxation. *Magn Reson Med* (2000) 44(2):292–300. doi: 10.1002/1522-2594(200008)44:2<292::AID-MRM17>3.0.CO;2-Q

29. Sehy JV, Ackerman JJ, Neil JJ. Evidence That Both Fast and Slow Water ADC Components Arise From Intracellular Space. *Magn Reson Med* (2002) 48(5):765–70. doi: 10.1002/mrm.10301

30. Sehy JV, Zhao L, Xu J, Rayala HJ, Ackerman JJ, Neil JJ. Effects of Physiologic Challenge on the ADC of Intracellular Water in the Xenopus Oocyte. *Magn Reson Med* (2004) 52(2):239–47. doi: 10.1002/mrm.20132

31. Hempel JM, Brendle C, Bender B, Bier G, Kraus MS, Skardelly M, et al. Diffusion Kurtosis Imaging Histogram Parameter Metrics Predicting Survival in Integrated Molecular Subtypes of Diffuse Glioma: An Observational Cohort Study. *Eur J Radiol* (2019) 112:144–52. doi: 10.1016/j.ejrad.2019.01.014

32. Matsusue E, Fink JR, Rockhill JK, Ogawa T, Maravilla KR. Distinction Between Glioma Progression and Post-Radiation Change by Combined Physiologic MR Imaging. *Neuroradiology* (2010) 52(4):297–306. doi: 10.1007/s00234-009-0613-9

# Predicting Tyrosine Kinase Inhibitor Treatment Response in Stage IV Lung Adenocarcinoma Patients With EGFR Mutation Using Model-Based Deep Transfer Learning

*Runping Hou[1,2†], Xiaoyang Li[2,3†], Junfeng Xiong[1,4], Tianle Shen[2], Wen Yu[2], Lawrence H. Schwartz[5], Binsheng Zhao[5], Jun Zhao[1*‡] and Xiaolong Fu[2*‡]*

[1] School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai, China, [2] Department of Radiation Oncology, Shanghai Chest Hospital, Shanghai Jiao Tong University, Shanghai, China, [3] The First Affiliated Hospital of USTC, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, China, [4] Division of Health Care, Tencent, Shenzhen, China, [5] Department of Radiology, Columbia University Irving Medical Center, New York, NY, United States

**Background:** For stage IV patients harboring EGFR mutations, there is a differential response to the first-line TKI treatment. We constructed three-dimensional convolutional neural networks (CNN) with deep transfer learning to stratify patients into subgroups with different response and progression risks.

**Materials and Methods:** From 2013 to 2017, 339 patients with EGFR mutation receiving first-line TKI treatment were included. Progression-free survival (PFS) time and progression patterns were confirmed by routine follow-up and restaging examinations. Patients were divided into two subgroups according to the median PFS (<=9 months, > 9 months). We developed a PFS prediction model and a progression pattern classification model using transfer learning from a pre-trained EGFR mutation classification 3D CNN. Clinical features were fused with the 3D CNN to build the final hybrid prediction model. The performance was quantified using area under receiver operating characteristic curve (AUC), and model performance was compared by AUCs with Delong test.

**Results:** The PFS prediction CNN showed an AUC of 0.744 (95% CI, 0.645–0.843) in the independent validation set and the hybrid model of CNNs and clinical features showed an AUC of 0.771 (95% CI, 0.676–0.866), which are significantly better than clinical features-based model (AUC, 0.624, P<0.01). The progression pattern prediction model showed an AUC of 0.762(95% CI, 0.643–0.882) and the hybrid model with clinical features showed an AUC of 0.794 (95% CI, 0.681–0.908), which can provide compensate information for clinical features-based model (AUC, 0.710; 95% CI, 0.582–0.839).

**Conclusion:** The CNN exhibits potential ability to stratify progression status in patients with EGFR mutation treated with first-line TKI, which might help make clinical decisions.

Keywords: deep learning—convolutional neural networks, computed tomography, lung cancer, transfer learning, epidermal growth factor receptor mutation

# INTRODUCTION

Non-small cell lung cancer (NSCLC) has the highest mortality both in United States and China (1, 2), of which lung adenocarcinoma accounts for about 50%. For stage IV lung adenocarcinoma patients harboring EGFR mutations, tyrosine kinase inhibitor (TKI) is recommended to be the first-line treatment modality especially for Asian patients with a relatively higher possibility of EGFR mutations (3). First-line TKI treatment could achieve the median progression-free survival (PFS) of approximately 10 months and a response rate of about 70% (4, 5). However, the disease inevitably progresses owing to acquired resistance to TKI treatment after a period of response. Because of inter-patient and inter-lesion heterogeneity, PFS and progression pattern of first-line TKI treatment are heterogeneous between patients. Different PFS and progression pattern determines different subsequent treatment strategy. For example, it is helpful to increase the PFS and even overall survival (OS) of TKI treatment by the addition of local ablative therapy for patients with favorable PFS and oligoprogression (6) and the enhancement of systematic therapy for patients with poor PFS and systematic progression (7). Therefore, accurate prediction of PFS and progression pattern of first-line TKI treatment is of great significance to the subsequent clinical decision making.

Nowadays, the prediction of PFS and progression pattern of TKI treatment in clinical practice is mainly based on the conventional information such as patient demographics, pathology, and genetics. Nevertheless, these features are low-dimensional with limited representational ability, which may lead to unsatisfactory accuracy. Recently, medical imaging has been widely used to help clinicians for decision making according to some morphological features about the tumor. However, these subjective and qualitative morphological features often result in low inter-observer agreement and limited accuracy. Thus, a more objective and quantitative method to accurately predict PFS and progression patterns of first-line TKI treatment is urgently needed.

Convolutional neural network (CNN) is an artificial intelligence algorithm with the capability to excavate the underlying biological information from medical imaging. Compared with the traditional feature engineering, CNN has great advantages in automatically extracting the latent deep representative features and developing robust end-to-end prediction models. It has been recently utilized in various medical domains with satisfactory results (8). In thoracic oncology, CNN could distinguish malignant pulmonary nodules (9), identify pathological types of lung cancer (10, 11), detect driven oncogene status, and other tasks (12–14) using chest CT images. Therefore, we decided to develop a CNN model to predict PFS and progression patterns of first-line TKI treatment of lung adenocarcinoma patients based on chest CT images.

For the training of CNN, the weights of network are often randomly initialized and then updated under the supervision of image labels, which is called "training from scratch." This method requires large amounts of data to learn the huge number of CNN parameters. However, in this study, as the number of patients harboring EGFR mutations and treated with first-line TKI is limited, this training strategy may cause overfitting of the CNN model and lead to poor generalization performance. Thus, how to train the CNN network with limited data is the major concern of this study.

Transfer learning is a technique that can help overcome the problem of insufficient training data. Researches have shown that the pre-trained weights and features from one domain are transferable to another domain with similar characteristics (15, 16). In this study, considering the available data are limited, we decided to train the CNN network with deep transfer learning using pretrained CNN models. The basic pretrained CNN model has been developed to distinguish benign and malignant pulmonary nodules in a large data set (with 8472 samples). Then, in light of the large dissimilarity between distinguishing pulmonary nodules (in early-stage patients) and predicting PFS and progression patterns (in stage IV patients), we added the domain of detecting EGFR mutations of stage IV patients for fine-tuning of the basic model. Afterward, this fine-tuned model for EGFR mutation prediction was further utilized to help train the progression prediction models.

Overall, this study aims to develop and validate a CNN model with model-based deep transfer learning to predict the PFS and progression patterns of first-line TKI treatment of lung adenocarcinoma patients based on pretherapy CT images. The pretrained model in source domain based on large data set would help the CNN model in target domain be better trained with limited data.

# MATERIALS AND METHODS

## Study Design

This study was approved by Shanghai Chest Hospital, Shanghai Jiaotong University. Ethical approval (ID: KS 1716) was obtained for the use of the CT images and clinical information of patients. Informed consent was waived for the respective nature of the study. The study design was illustrated in **Figure 1**. The basic CNN model was previously constructed by the domain of distinguishing malignant and benign pulmonary nodules. Then, this basic model was fine-tuned by the domain of detecting EGFR mutation of lung adenocarcinoma, and then transferred to predict the PFS and progression patterns of TKI treatment.

## Patients

We retrospectively analyzed patients receiving first-line TKI treatment from 2013 to 2017 in Shanghai Chest Hospital. The inclusion criteria: a. Patients were diagnosed with stage IIIB and IV lung adenocarcinoma harboring EGFR mutations; b. Patients received first-line TKI treatment. c. The smoking history should be clear. d. Patients should undertake completed staging

---

**Abbreviations:** 3D CNN, three-dimensional convolutional neural network; CI, confidence interval; PFS, progression free survival; TKI, tyrosine kinase inhibitor; ROI, region of interest.
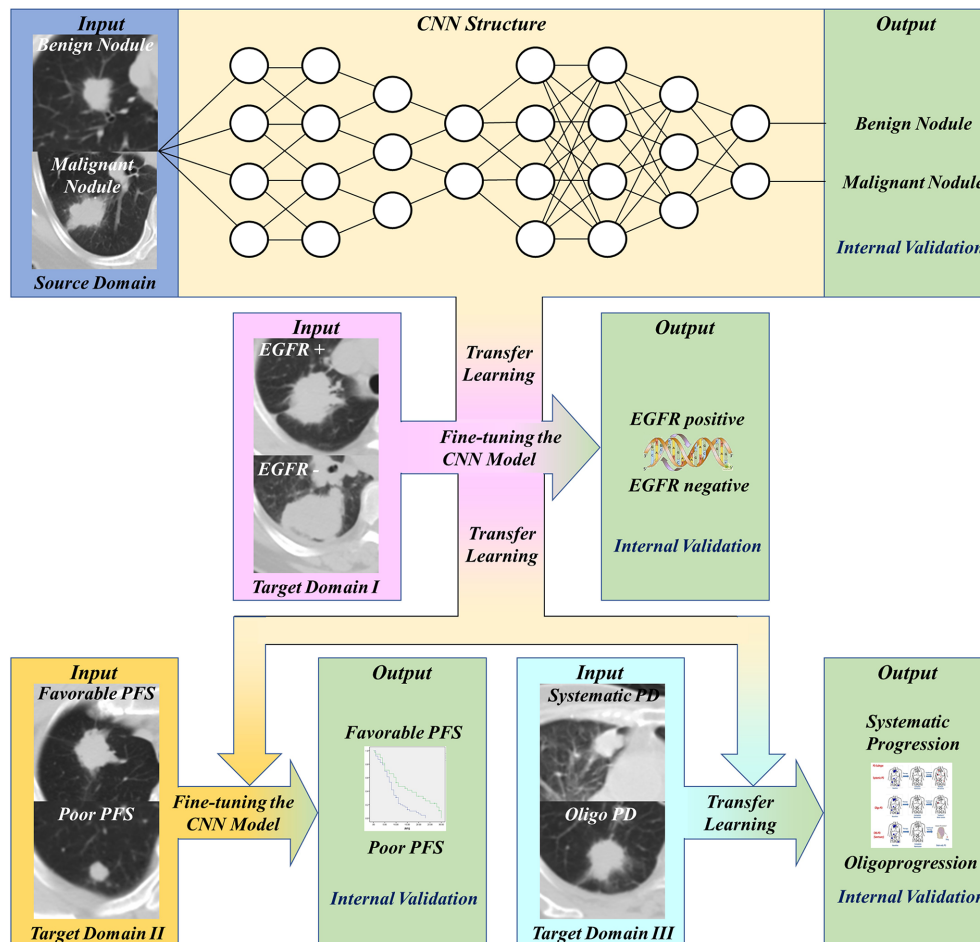
**FIGURE 1** | Workflow of our work. First, transfer and finetune the basic pulmonary nodule recognition model to EGFR prediction. Then, transfer the EGFR recognition model to PFS prediction and progression patterns prediction.

examination to confirm the clinical stage. e. The pulmonary nodules should be solid with the max diameter over 0.8 cm. f. Patients should receive completed follow-up every 3 months to confirm the accurate PFS and restaging examination for judging progression patterns. The exclusion criteria: a. Patients with non-adenocarcinoma; b. Patients without completed staging examinations to confirm the clinical stages of IIIB and IV. c. Patients did not undertake routine follow-up and restaging examinations to confirm the accurate PFS and progression patterns. PFS was defined from the start of TKI treatment to first progression or last follow-up date. For metastatic pattern at initial diagnosis, systemic metastasis was defined as over five metastatic sites or over three organs, and oligometastasis was less than five metastatic sites within three or fewer organs. The progression pattern was classified into oligoprogression and systematic progression. The systematic progression was defined as multi-sites progression, which may include both new metastatic sites, as well as regrowth in previously responsive sites of disease. The oligoprogression was defined to CNS

progression without leptomeningeal progression and extra-cerebral progression four or less sites. All the enrolled patients were assigned into training set and validation set randomly.

## CT Image Acquisition and Preprocessing

Chest CT scans were taken with voltage from 120 kV to 140 kV, current 170 mA, scan layer thickness 5 mm, and spatial resolution about 1 mm using Brilliance 64 CT from PHILIPS. Tumors were manually segmented by an experienced radiologist (window level -400 and window width 1600) on the platform Pinnalce2 for Varian®. The radiologist was asked to only delineate a rough region of interest covering nodules. Linear interpolation was applied to the original CT images to get isotropic images (1 mm × 1 mm × 1 mm). Image patches were cropped from the interpolated images centered as the tumor.

## Model Development and Evaluation

We established binary classifiers using three dimensional CNN (3D CNN) to distinguish patients with different PFS and progression

patterns. Model-based deep transfer learning was utilized to train the CNN more effectively. The 3D CNN was pre-trained on a source task and then the weights of some layers or features were transferred to the target domain task. Moreover, parameter fine-tuning was used to retrain the network on the task of PFS prediction. The workflow of our work was shown in **Figure 1**. Based on the basic model for benign and malignant pulmonary nodules recognition, we use transfer learning and fine-tuning to develop the EGFR classification model, then further transferred this EGFR classification model to PFS and progression patterns prediction.

## Establishment and Fine-Tuning of the Basic Models

The structure of the basic model for nodule recognition ($CNN_{BM}$) was based on a 3D residual network with prior attention. The residual network can effectively tackle the vanishing gradient problem in deep neural networks. The inputs of the network include both the CT image patch and corresponding mask, which cover the region of interest (ROI), to make the network focusing on image pixels within the ROI. As the sample size was large for training (8,472 samples), the basic model was trained from scratch. The details of the basic CNN architecture can be found in **Supplementary Table S1**.

An EGFR classification model ($CNN_{EGFR}$) was constructed through fine-tuning of the pre-trained basic model. The fine-tuning process makes the network more applicable for the IV stage patients, and some latent EGFR mutation-related features can be learned, which may be helpful for the PFS prediction after TKI treatment.

## Establishment of PFS Prediction by Transfer Learning

For PFS prediction, patients with favorable PFS (>9 months) were regarded as positive samples with label 1, and those with poorer PFS (<=9 months) were negative samples with label 0. To develop the PFS prediction model, we transfer and fine-tune the pre-trained 3D CNN model in two steps. First, we freeze the top- layers' parameters and only train the fully connected layers with a larger initial learning rate of 1e-2. After 10 epochs training, we unfreeze the frozen layers and fine-tune the whole network with a smaller initial learning rate of 1e-4. The $CNN_{EGFR}$ was also trained in this way based on the $CNN_{BM}$. To evaluate the effect of transfer learning and the influence of domain difference, we compare different 3D CNN PFS prediction models respectively fine-tuned from the $CNN_{EGFR}$, $CNN_{BM}$, and trained from scratch. Furthermore, clinical features, such as age, sex, smoking, clinical stages, and molecular pathology status, were fused with 3D CNN model by logistic regression for better prediction. Tensorflow (tensorflow.org) was used for network training.

## Establishment of Progression Pattern Prediction by Transfer Learning

For progression patterns prediction, patients with systematic progression were regarded as positive samples, and patients with oligoprogression were negative samples. To develop the progression patterns prediction model, we use the pre-trained $CNN_{EGFR}$ as

feature extractor and then construct classifiers. The deep features (dimensional feature vectors, 128) were extracted from the last layer before the outputs. After feature extraction, univariate feature selection and recursive feature elimination were used to select features, then decision tree, random forest, and K-Nearest Neighbor classifiers were constructed to realize the final prediction. Furthermore, because T stage and metastasis status are significant factors related to patients' progression patterns, we used the two factors to build a logistic regression model as the baseline. Finally, this basic model and the image-based model were fused to develop the hybrid prediction model. The algorithms were implemented with scikit-learn (scikit-learn.org) in python.

## Statistical Analysis

Statistical analysis was conducted in R software (Rproject.org). Fisher's exact test, Wilcoxon test, and chi-square test were used to compare the differences of clinical features between training and validation groups. For model evaluation, the receiver operating characteristic (ROC) curve and AUC were used to describe model performance, and DeLong (17) test was used to pairwise compare the difference of two ROCs. Kaplan-Meier survival curves of the subgroups stratified by our model (favorable/poor PFS) were plotted, and log-rank test was used to compare difference of two KM curves. P value less than 0.05 was considered as significant.

# RESULTS

## Patient Characteristics

We retrospectively analyzed 339 patients for the creation of PFS prediction model. Patients were randomly divided into training group (70.5%) and validation group (29.5%). No significant difference was found between the two groups in terms of all clinical characteristic (**Table 1**). The median PFS of total patients was 9 months. There were 169, 160, and 10 patients harboring EGFR exon 19, exon 21, and double site mutation, respectively. At Cox proportional hazard regression, all the clinical characteristics, including age, gender, smoking status, clinical stage, and EGFR mutation site, were not prognostic for PFS. After excluding patients without confirmed progression pattern, totally 255 patients were enrolled for the establishment of progression pattern prediction model. The detailed characteristics of patients were shown in **Table 2**. For the metastatic pattern in the initial diagnosis, there were 186 (72.9%) and 55 (21.6%) patients demonstrating systematic metastasis and oligometastasis, respectively. While at acquired resistance to TKI, 153 (60%) and 102 (40%) displayed systematic progression and oligoprogression, respectively. At multivariate logistic regression, T stage (OR=1.70, p<0.001) and metastatic pattern (OR=3.29, p=0.006) were recognized to be related with progression pattern.

## Structure and Performance of Basic Models

The basic model we developed to distinguish malignant pulmonary nodules achieved good performance with a high

**TABLE 1 |** Comparison of clinical features in patients with PFS information.

| Clinical Features | Training group (n = 239) | Validation group (n = 100) | p value |
|---|---|---|---|
| **Age** | | | |
| Median (Range) | 61 (33-84) | 61 (26-82) | t-test p=0.217 |
| **Gender (n%)** | | | |
| Male | 97 (40.6) | 32 (32.0) | Pearson χ² Test |
| Female | 142 (59.4) | 68 (68.0) | p=0.143 |
| **Smoking History** | | | |
| Yes | 55 (23.0) | 17 (17.0) | Pearson χ² Test |
| No | 184 (77.0) | 83 (83.0) | p=0.265 |
| **PFS (months)** | | | |
| Median | 9 | 9 | Log-rank Test p=0.265 |
| ≤9 months | 138 (57.7) | 56 (56.0) | Pearson χ² Test |
| >9 months | 101 (42.3) | 44 (44.0) | p=0.810 |
| **Clinical Staging** | | | |
| IIIA | 8 (3.3) | 3 (3.0) | Mann-Whitney Test |
| IIIB | 22 (9.2) | 7 (7.0) | p=0.989 |
| IV | 209 (87.5) | 90 (90.0) | |
| **EGFR mutation site** | | | |
| 19del | 121 (50.6%) | 48 (48%) | Pearson χ² Test |
| 21L858R | 113 (47.3%) | 47 (47%) | p=0.346 |
| Double Site | 5 (2.1%) | 5 (5%) | |

**TABLE 2 |** Comparison of clinical features in patients with progression patterns information.

| Clinical Features | Training group (n = 195) | Validation group (n = 60) | p value |
|---|---|---|---|
| **Age** | | | |
| Median (range) | 61 (26–81) | 59 (35–84) | t-test p=0.777 |
| **Gender (n%)** | | | |
| Male | 77 (39.5) | 20 (33.3) | Pearson χ² test p=0.737 |
| Female | 118 (60.5) | 40 (66.7) | |
| **Smoking history** | | | |
| Yes | 156 (80.0) | 49 (81.7) | Pearson χ² test p=0.776 |
| No | 39 (20.0) | 11 (18.3) | |
| **PFS (months)** | | | |
| Median | 8 | 11 | Log-rank test p=0.131 |
| **T Stage** | | | |
| T1 | 38 (19.5%) | 11 (18.3%) | Mann-Whitney test |
| T2 | 41 (21.0%) | 12 (20%) | p=0.865 |
| T3 | 16 (8.2%) | 6 (10.0%) | |
| T4 | 100 (51.3%) | 31 (51.7%) | |
| **Metastasis pattern at initial diagnosis** | | | |
| Oligometastasis | 26 (13.3%) | 13 (21.7%) | Pearson χ² test p=0.117 |
| Systematic metastasis | 169 (86.7) | 47 (78.3) | |
| **Progression pattern** | | | |
| Oligoprogression | 77 (39.5) | 25 (41.7) | Pearson χ² test p=0.763 |
| Systematic progression | 118 (60.5) | 35 (58.3) | |
| **EGFR mutation site** | | | |
| 19del | 105 (53.8%) | 33 (55%) | Pearson χ² test p=0.756 |
| 21L858R | 83 (42.6%) | 26 (43.3%) | |
| Double site | 7 (3.6%) | 1 (1.7%) | |

AUC value of 0.932 (95% CI, 0.924–0.947). When utilizing the domain of detecting EGFR mutation for fine-tuning, the AUC value of the model was 0.863 (95% CI, 0.763–0.897) (**Supplementary Figures S2, S3**).

## PFS Prediction

The 3D-CNN model trained from scratch ($CNN_{Scratch}$), transferred directly from the basic model ($CNN_{TL-BM}$) and from the EGFR mutation fine-tuned model ($CNN_{TL-EGFR}$) showed AUCs of 0.668 (95% CI, 0.559–0.776), 0.701 (95% CI, 0.598–0.805), and 0.744 (95% CI, 0.645–0.843) in the validation group, respectively (**Figures 2**, **3A**). The 3D CNN model with EGFR classification fine-tuning achieved better performance than the model directly transferred from the basic model, mainly because the domain difference between EGFR classification and PFS stratification is smaller than the difference between nodule classification and PFS stratification. Moreover, after adding the clinical features, the corresponding fusion model's performance improved to 0.715 (95% CI, 0.614–0.816), 0.756 (95% CI, 0.659–0.854), and 0.771 (95% CI, 0.676–0.866), respectively (**Figure 2** and **Table 3**). As shown in **Figure 3B**, the best 3D CNN model ($CNN_{TL-EGFR}$) performed better than clinical features-based model (AUC, 0.744 vs 0.624). Furthermore, the fusion model ($CNN_{TL-EGFR}$ and Clinical) achieved significantly better performance than the clinical model alone (P=0.008).

Then, according to the prediction results of different 3D CNN models, we divided the validation group into high-risk and low-risk subgroups. The optimal cutoff threshold was confirmed by X-tile (18). Based on this, Kaplan-Meier survival curves were plotted respectively in the two subgroups. As shown in **Figure 4**, the $CNN_{TL-EGFR}$ and clinical fusion model achieved the best performance and can significantly distinguish the difference in PFS between the stratified progression subgroups (log-rank test, P<0.001).

## Progression Pattern Prediction

As above-mentioned, the addition of EGFR recognition fine-tuning achieved the highest prediction efficacy among all the
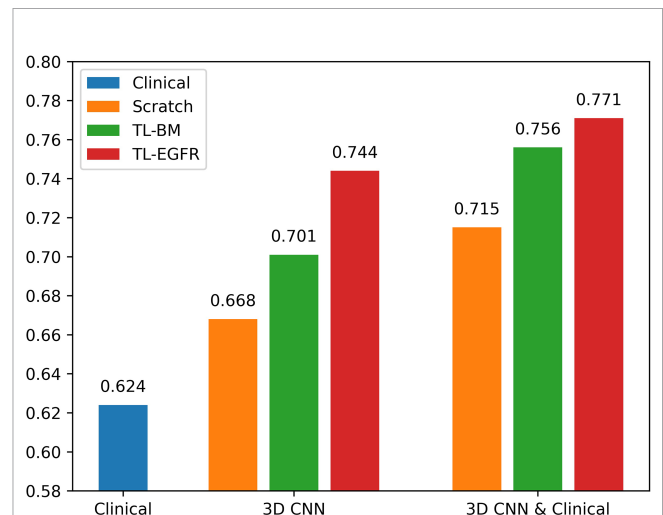


**FIGURE 2 |** AUCs of each PFS prediction models in the validation group. The blue ones correspond to the clinical alone model. The orange, green, and red ones correspond to CNN model trained from scratch, transferred from nodule, and transferred from EGFR classification models, receptively.
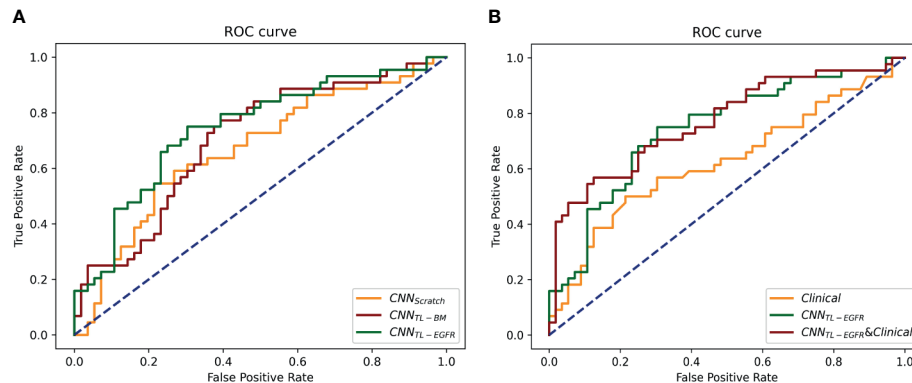
**FIGURE 3 |** ROCs of 3D CNN models for the prediction of PFS in the validation group. **(A)** The ROCs of 3D CNN model trained from scratch, using transfer learning based on nodule or EGFR classification models. The corresponding AUCs were 0.668, 0.701, and 0.744, receptively. **(B)** The ROCs of only using clinical features, 3D CNN (using transfer learning based on EGFR classification model), and the combination of 3D CNN and clinical features, and the corresponding AUCs were 0.624, 0.744, and 0.771.

**TABLE 3 |** Performance of different PFS prediction models in the validation group.

| Models | $CNN_{TL\text{-}EGFR}$ | $CNN_{TL\text{-}BM}$ | $CNN_{Scratch}$ |
|---|---|---|---|
| AUC | 0.744 | 0.701 | 0.668 |
| 95% CI | 0.645 to 0.843 | 0.598 to 0.805 | 0.559 to 0.776 |
| Threshold | 0.449 | 0.379 | 0.490 |
| Accuracy | 72.0% | 68.0% | 68.0% |
| Sensitivity | 75.0% | 77.3% | 54.5% |
| Specificity | 69.6% | 60.7% | 78.6% |
| **Models** | **$CNN_{TL\text{-}EGFR}$ and Clinical** | **$CNN_{TL\text{-}BM}$ and Clinical** | **$CNN_{Scratch}$ and Clinical** |
| AUC | 0.771 | 0.756 | 0.715 |
| 95% CI | 0.676 to 0.866 | 0.659 to 0.854 | 0.614 to 0.816 |
| Threshold | 0.575 | 0.615 | 0.496 |
| Accuracy | 74.0% | 75.0% | 70.0% |
| Sensitivity | 56.8% | 52.3% | 56.8% |
| Specificity | 87.5% | 92.9% | 80.4% |

*CNN, convolutional neural network; AUC, area under receiver operating characteristic curve; threshold, threshold at the optimal decision point; CI, confidence interval.*

3D-CNN models. Therefore, in the prediction of progression pattern, we utilized transfer learning from the EGFR classification to develop the progression pattern model.

The progression patterns prediction model transferred from EGFR classification achieved an AUC of 0.762 (95% CI, 0.643–0.882; sensi, 0.92; speci, 0.571). Clinical features-based model achieved an AUC of 0.710 (95% CI, 0.582–0.839; sensi, 0.686; speci, 0.760), and the hybrid model achieved an AUC of 0.794 (95% CI, 0.681–0.908; sensi, 0.92; speci, 0.66). The ROCs of the models in the validation group were shown in **Figure 5**.

## DISCUSSION

In this study, we developed a PFS prediction and a progression pattern prediction model using model-based deep transfer learning based on a pre-trained EGFR classification CNN model. Results show that the hybrid model combining transfer learning-based and clinical features-based model finally achieved satisfactory performance for PFS prediction (AUC = 0.771) and progression pattern prediction (AUC = 0.794). Also, the PFS prediction model can significantly stratify patients with different progression risk after first-line TKI treatment (P < 0.001). Utilization of the established CNN model could instruct clinical practice to individually modify TKI treatment for a better prognosis.

The above results indicate that image-based deep learning can mine more informative features for the prediction of tumor's biological behavior. Also, the results indicate that 3D CNN trained with model-based deep transfer learning performs better than model training from scratch (AUC, 0.668–0.744), and the smaller the difference between source domain and target domain, the better performance can transfer learning achieve (AUC, 0.701–0.744). Compared with the most widely used transfer learning pretrained on 2D natural image data set (ImageNet) (14, 19), our 3D transfer learning is based on 3D medical image data set, which can not only mine more spatial information but also effectively reduce the domain difference.

A recent study about predicting EGFR-TKI treatment response using CT images (20) used a self-supervised learning-based model called BigBiGAN as a feature extractor, then utilize the extracted features to construct a Cox regression model for distinguishing patients with different progression risk. In comparison, we used a supervised learning based pre-trained model for transfer learning, then utilize the progression label to finetune the model and update the extracted deep features. Compared with the BigBiGAN model trained in self-supervised ways, our pre-trained model trained with EGFR status can learn not only the inherent grayscale-based features but also some implicit biologically related image features. Moreover, because of the small difference between the source
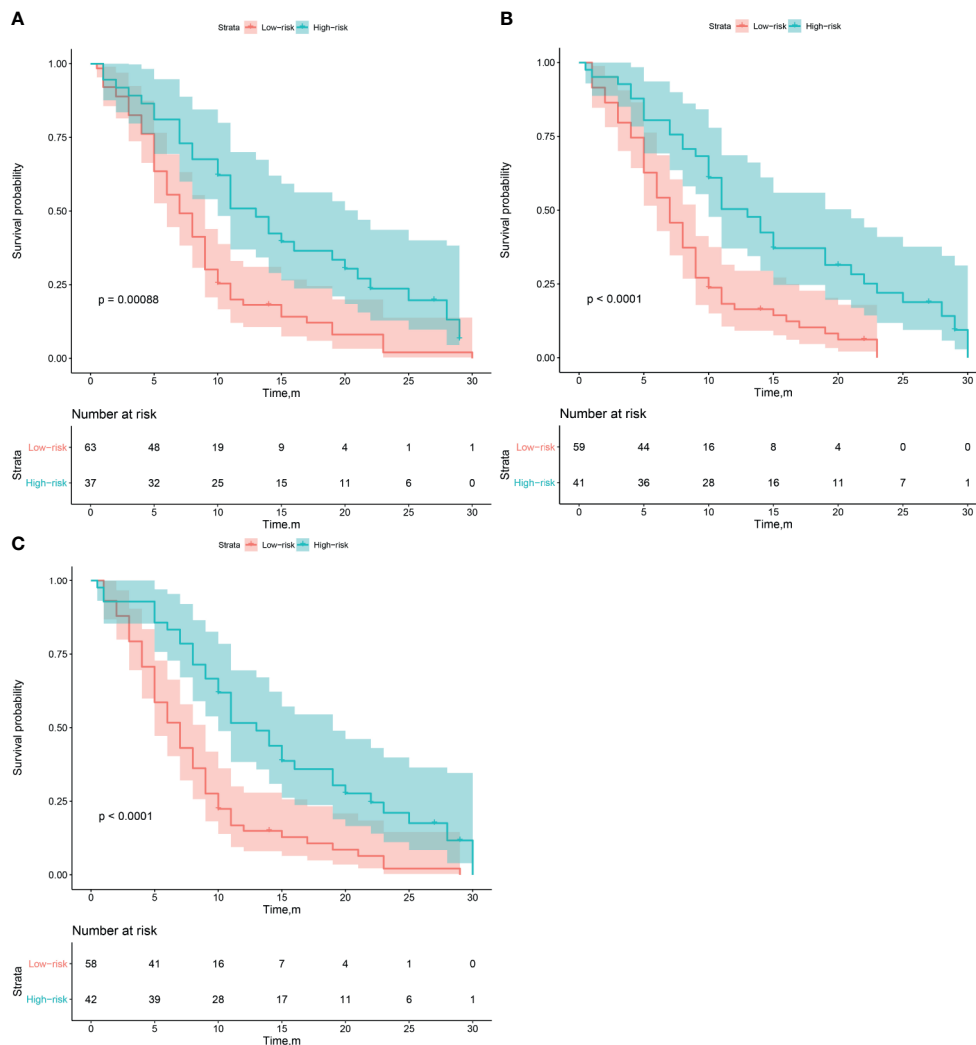
**FIGURE 4** | Survival analysis of PFS in low and high risk patients in the validation group. **(A–C)** The CNN<sub>Scratch</sub> and clinical, CNN<sub>TL-BM</sub> and clinical, and CNN<sub>TL-EGFR</sub> and clinical model's KM curves, respectively.

domain (classification of EGFR mutation) and target domain (prediction of EGFR-TKI therapy response) of our proposed method, the network can make fully use of the pre-learned features effectively, which can better help the prediction of disease progression.

This research also has several limitations. First, because of the limitation of sample size, this study only realized the simple binary classification of patients' PFS with the median survival as the cutoff threshold. In the future, we will collect more samples, and further attempt deep Cox regression to realize the end-to-end survival prediction. Second, our hypothesis of the relationship between EGFR mutation and patient's PFS is that the mutation abundance is thought to be related with the patient's survival (21). Therefore, we thought the CNN classifying EGFR mutation status will also be able to learn

information about the mutation abundance, which may be useful for PFS prediction. However, this mutation abundance information learning was clearly insufficient. In the future, if the mutation abundance information of the EGFR mutation patients can be collected, a more efficient network can be built and further correlate with the patient's PFS. Finally, the model should be validated in a prospective cohort to confirm its efficacy.

## CONCLUSION

We developed a deep transfer learning-based PFS prediction and progression pattern prediction model in EGFR mutation
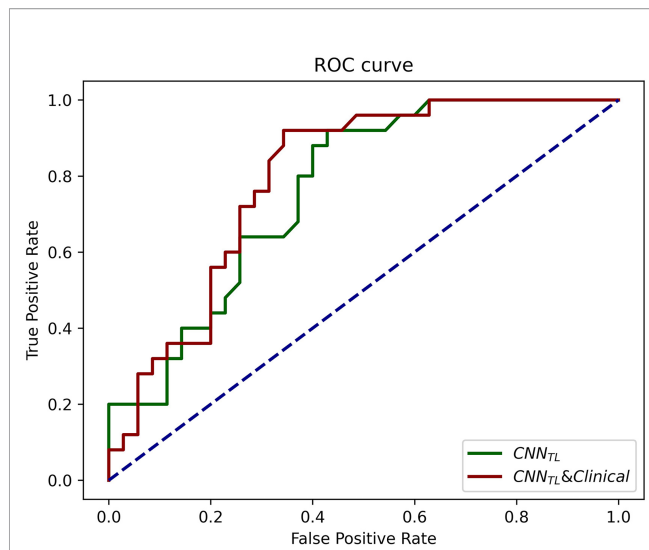
**FIGURE 5** | The ROCs of CNN$_{TL}$ (transferred from the EGFR classification) and the CNN$_{TL}$ and clinical models for progression patterns prediction in the validation group. The corresponding AUCs were 0.76 and 0.79.

patients treated with TKIs. The results showed that the prediction model transferred from EGFR classification can significantly stratify patients with different progression risk after TKI treatment, which may be able to further help the clinical decision making.

## DATA AVAILABILITY STATEMENT

The data sets presented in this article are not readily available because the data sets are privately owned by Shanghai Chest Hospital and are not made public. Requests to access the data sets should be directed to XF, xlfu1964@hotmail.com.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Shanghai Chest Hospital. The ethics committee waived the requirement of written informed consent for participation.

## AUTHOR CONTRIBUTIONS

XF, JZ, RH, and XL contributed to the study concept and design. XL, RH, and TS contributed to acquisition of data. RH, XL, JX, WY, LS, and BZ contributed to analysis and interpretation of data. RH and XL contributed to drafting of the manuscript. The corresponding author had full access to all of the data and took full responsibility for the veracity of the data and the statistical analyses. All authors contributed to the article and approved the submitted version.

## FUNDING

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2021.679764/full#supplementary-material

## REFERENCES

1. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer Statistics, 2021. *CA: Cancer J Clin* (2021) 71(1):7–33. doi: 10.3322/caac.21654
2. Gao S, Li N, Wang S, Zhang F, Wei W, Li N, et al. Lung Cancer in People's Republic of China. *J Thoracic Oncol Off Publ Int Assoc Study Lung Cancer* (2020) 15(10):1567–76. doi: 10.1016/j.jtho.2020.04.028
3. Han B, Tjulandin S, Hagiwara K, Normanno N, Wulandari L, Laktionov K, et al. EGFR Mutation Prevalence in Asia-Pacific and Russian Patients With Advanced NSCLC of Adenocarcinoma and Non-Adenocarcinoma Histology: The IGNITE Study. *Lung Cancer (Amsterdam Netherlands)* (2017) 113:37–44. doi: 10.1016/j.lungcan.2017.08.021
4. Mok TS, Wu YL, Thongprasert S, Yang CH, Chu DT, Saijo N, et al. Gefitinib or Carboplatin-Paclitaxel in Pulmonary Adenocarcinoma. *N Engl J Med* (2009) 361(10):947–57. doi: 10.1056/NEJMoa0810699
5. Rosell R, Carcereny E, Gervais R, Vergnenegre A, Massuti B, Felip E, et al. Erlotinib Versus Standard Chemotherapy as First-Line Treatment for European Patients With Advanced EGFR Mutation-Positive Non-Small-Cell Lung Cancer (EURTAC): A Multicentre, Open-Label, Randomised Phase 3 Trial. *Lancet Oncol* (2012) 13(3):239–46. doi: 10.1016/s1470-2045(11)70393-x
6. Gomez DR, Tang C, Zhang J, Blumenschein GRJr., Hernandez M, Lee JJ, et al. Local Consolidative Therapy Vs. Maintenance Therapy or Observation for Patients With Oligometastatic Non-Small-Cell Lung Cancer: Long-Term Results of a Multi-Institutional, Phase II, Randomized Study. *J Clin Oncol Off J Am Soc Clin Oncol* (2019) 37(18):1558–65. doi: 10.1200/jco.19.00201
7. Hosomi Y, Morita S, Sugawara S, Kato T, Fukuhara T, Gemma A, et al. Gefitinib Alone Versus Gefitinib Plus Chemotherapy for Non–Small-Cell Lung Cancer With Mutated Epidermal Growth Factor Receptor: NEJ009 Study. *J Clin Oncol* (2020) 38(2):115–23. doi: 10.1200/JCO.19.01488
8. Bi WL, Hosny A, Schabath MB, Giger ML, Birkbak NJ, Mehrtash A, et al. Artificial Intelligence in Cancer Imaging: Clinical Challenges and Applications. *CA: Cancer J Clin* (2019) 69(2):127–57. doi: 10.3322/caac.21552
9. Ather S, Kadir T, Gleeson F. Artificial Intelligence and Radiomics in Pulmonary Nodule Management: Current Status and Future Applications. *Clin Radiol* (2019) 75(1):13–9. doi: 10.1016/j.crad.2019.04.017
10. Zhao W, Yang J, Sun Y, Li C, Wu W, Jin L. 3d Deep Learning From CT Scans Predicts Tumor Invasiveness of Subcentimeter Pulmonary Adenocarcinomas. *Cancer Res* (2018) 78(24):6881–9. doi: 10.1158/0008-5472.can-18-0696
11. Shen W, Zhou M, Yang F, Yang C, Tian J. Multi-Scale Convolutional Neural Networks for Lung Nodule Classification. *Inf Process Med Imaging* (2015) 24:588–99. doi: 10.1007/978-3-319-19992-4_46
12. Xiong JF, Jia TY, Li XY, Yu W, Xu ZY, Cai XW, et al. Identifying Epidermal Growth Factor Receptor Mutation Status in Patients With Lung

Adenocarcinoma by Three-Dimensional Convolutional Neural Networks. *British J Radiol* (2018) 91(1092):20180334. doi: 10.1259/bjr.20180334

13. Kim H, Goo JM, Lee KH, Kim YT, Park CM. Preoperative CT-Based Deep Learning Model for Predicting Disease-Free Survival in Patients With Lung Adenocarcinomas. *Radiology* (2020) 296(1):216–24. doi: 10.1148/radiol.2020192764

14. Wang S, Shi J, Ye Z, Dong D, Yu D, Zhou M, et al. Predicting EGFR Mutation Status in Lung Adenocarcinoma on Computed Tomography Image Using Deep Learning. *Eur Respir J* (2019) 53(3):1800986. doi: 10.1183/13993003.00986-2018

15. C Tan, F Sun, T Kong, W Zhang, C Yang, C Liu eds. A Survey on Deep Transfer Learning. In: . *International Conference on Artificial Neural Networks*. Springer (2018) 270–9.

16. Yosinski J, Clune J, Bengio Y, Lipson H eds. How Transferable Are Features in Deep Neural Networks? In: *Proceedings of the 27th International Conference on Neural Information Processing Systems* (2014). 27:3320–8.

17. 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

18. Camp RL, Dolled-Filhart M, Rimm DL. X-Tile: A New Bio-Informatics Tool for Biomarker Assessment and Outcome-Based Cut-Point Optimization. *Clin Cancer Res* (2004) 10(21):7252–9. doi: 10.1158/1078-0432

19. Shin HC, Roth HR, Gao M, Lu L, Xu Z, Nogues I, et al. Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset

Characteristics and Transfer Learning. *IEEE Trans Med Imaging* (2016) 35 (5):1285–98. doi: 10.1109/TMI.2016.2528162

20. Song J, Wang L, Ng N, Zhao M, Shi J, Wu N, et al. Development and Validation of a Machine Learning Model to Explore Tyrosine Kinase Inhibitor Response in Patients With Stage IV EGFR Variant-Positive Non-Small Cell Lung Cancer. *JAMA Network Open* (2020) 3(12):e2030442. doi: 10.1001/jamanetworkopen.2020.30442

21. Li X, Cai W, Yang G, Su C, Ren S, Zhao C, et al. Comprehensive Analysis of EGFR-Mutant Abundance and Its Effect on Efficacy of EGFR TKIs in Advanced NSCLC With EGFR Mutations. *J Thorac Oncol* (2017) 12 (9):1388–97. doi: 10.1016/j.jtho.2017.06.006

# Radiomics in Oncology: A 10-Year Bibliometric Analysis

*Haoran Ding, Chenzhou Wu, Nailin Liao, Qi Zhan, Weize Sun, Yingzhao Huang, Zhou Jiang and Yi Li\**

*State Key Laboratory of Oral Diseases & National Clinical Research Center for Oral Diseases & Department of Head and Neck Oncology Surgery, West China Hospital of Stomatology, Sichuan University, Chengdu, China*

**Objectives:** To date, radiomics has been applied in oncology for over a decade and has shown great progress. We used a bibliometric analysis to analyze the publications of radiomics in oncology to clearly illustrate the current situation and future trends and encourage more researchers to participate in radiomics research in oncology.

**Methods:** Publications for radiomics in oncology were downloaded from the Web of Science Core Collection (WoSCC). WoSCC data were collected, and CiteSpace was used for a bibliometric analysis of countries, institutions, journals, authors, keywords, and references pertaining to this field. The state of research and areas of focus were analyzed through burst detection.

**Results:** A total of 7,199 pieces of literature concerning radiomics in oncology were analyzed on CiteSpace. The number of publications has undergone rapid growth and continues to increase. The USA and Chinese Academy of Sciences are found to be the most prolific country and institution, respectively. In terms of journals and co-cited journals, *Scientific Reports* is ranked highest with respect to the number of publications, and *Radiology* is ranked highest among co-cited journals. Moreover, Jie Tian has published the most publications, and Phillipe Lambin is the most cited author. A paper published by Gillies et al. presents the highest citation counts. Artificial intelligence (AI), segmentation methods, and the use of radiomics for classification and diagnosis in oncology are major areas of focus in this field. Test-retest statistics, including reproducibility and statistical methods of radiomics research, the relation between genomics and radiomics, and applications of radiomics to sarcoma and intensity-modulated radiotherapy, are frontier areas of this field.

**Conclusion:** To our knowledge, this is the first study to provide an overview of the literature related to radiomics in oncology and may inspire researchers from multiple disciplines to engage in radiomics-related research.

**Keywords: radiomics, oncology, bibliometric analysis, hotspots, trends**

# INTRODUCTION

Unlike the natural intelligence displayed by humans and animals, artificial intelligence (AI) is intelligence demonstrated by machines. AI can be applied to develop systems possessing characteristics of human beings: the ability to learn, reasoning, sensing, and actioning. Initially, the development of automated interpretation of medical images was based on human decision models to perform high-level interpretations of images. At the time, logical rules were applied to AI machines. AI machines sought specific structures such as lines or circles for identification. Such systems have succeeded in other fields, such as in business and manufacturing. Next came the second generation of AI algorithms. Instead of focusing on certain symbols of images, such algorithms were designed to be more statistical. This kind of model of medical images may develop from healthy individuals, and its parameters are inferred from data. Such an algorithm can assist in helping radiologists identify lesions. The segmentation method serves as a classic example of this algorithm. Currently, the explosion of big data has ushered AI into a new era, and algorithms are called data-driven/model-free approaches, which involve automating knowledge discovery. This approach is now widely applied in medical research, and a popular application of this method is called radiomics [1].

Radiomics, first pioneered by Philippe Lambin [2], uses high-throughput data to extract certain features from medical images for personalized precision medicine development. With the development of AI, the field of radiomics has grown rapidly and been widely used in every phase of tumor treatment. Relying on quantitative data generated by medical imaging and the support of technology, radiomics offers a risk-free and efficient method for diagnosis [3, 4], classification [5], and prognosis prediction [6, 7] in oncology.

Data selection, medical imaging, feature extraction, exploratory analysis, and modelling are the five steps of radiomics [2, 8, 9]. Applying standard imaging protocols to generate high-quality images from computed tomography (CT), positron-emission CT (PET-CT), magnetic resonance imaging (MRI), radiography or using high-quality photographs of lesions can enable radiomics reproducibility. Then, the volume of interest (VOI), which identifies lesions from images, is delineated by experienced radiologists or semiautomated or automated segmentation methods. The features extracted from VOIs are inputted to generate quantitative descriptions, which contain semantic and agnostic features. The value of extracted features is then analyzed, and only the features most contributing to the classifiers are retained for future modelling. In practice, this step can be supported by statistical approaches and AI, including a univariate analysis of variance, the least absolute shrinkage and selection operator, decision trees, neural networks, and support vector machines [9]. The relationship among the algorithms is presented in **Supplementary Figure S1**.

Although radiomics can be applied in a large number of conditions, it is most well developed and widely used in oncology due to initial support received from the National Cancer Institute [9, 10] and Quantitative Imaging Biomarker Alliance [11]. In 1973, some researchers intended to use texture features to classify images [12]. In 1995, researchers started to use a convolutional neural network (CNN) to identify lung nodes, suggesting that it is possible to train computer algorithms to identify medical images [13]. In late 2000, researchers attempted to identify the relationship between the imaging of tumors and their genomic types [14, 15]. At the time, most studies were performed on relatively small datasets and lacked external examination, meaning that the established radiomic models were only based on small datasets from individual organizations and could not be validated by data from external organizations. With innovations made in the field of medical imaging, radiomics in oncology has rapidly progressed [16–19]. In around 2012, radiomics was first proposed by Philippe Lambin; ever since, thousands of researchers have been encouraged to conduct radiomics-based research. In 2014, radiomics in oncology was used to examine CT imaging features for diagnosis and prediction [20]. In 2016, researchers found for the first time that the radiomics signature could predict lymph node metastasis in patients with colorectal cancer [21]. In 2018, researchers found that radiomics features could predict the treatment responses and prognosis of patients receiving immunotherapy [18]. Some researchers have developed models for the automated identification of lesions from videos and images [3, 4]. In addition, some researchers have found a correlation between radiomic features and tumor histology [22]. At present, by extracting various features from medical images and translating these image features into high-throughput and quantitative data for analysis, radiomics can be used for the classification and differentiation of different lesions and subtypes of tumors [3–5] and for survival prediction [23] and prognosis prediction for patients undergoing radiation therapy [6, 24, 25]. Even with common limitations, such as a lack of outside validations, the use of small datasets or the variabilities caused by medical imaging protocols, radiomics research has offered a significant opportunity for researchers to make clinical decisions from an entirely new perspective.

Since radiomics studies mostly rely on medical data, which are subject to approaches to data acquisition and analysis used, creating a gold standard for medical models remains a great challenge. When establishing a clinical model, the input and processing of radiomic features can drastically influence the model, as these features depend on the radiologists who record the clinical characteristics, on AI and on statistical methods. For example, inexperienced radiologists may fail to delineate a lesion entirely or miss significant clinical features, and different AI and statistical methods may create clinical diagnosis or prediction models with inconsistent accuracy. For these reasons, standardizing the procedures of radiomics studies and finding robust features with which to establish models are essential in achieving clinical goals of radiomics research.

Bibliometric analysis evaluates scientific activities in a certain field [26]. A simple quantitative technique provided by citation analysis provides a means to estimate the impact of an article [27], such as the influence of bridging articles between themes or the influence of articles laying the foundation in certain fields.

CiteSpace, developed by Chaomei Chen, is a Java-based application for detecting and visualizing possible trends and radical changes in scientific disciplines over time (28) and is a valuable tool for bibliometric analysis. The program can assist researchers in identifying influential and effective areas of research, trends, and prospects in certain fields. CiteSpace has been widely applied in many subjects for bibliometric analysis, such as neuroscience, oncology, and cardiovascular science (29–31).

In this article, we use data collected from the Web of Science Core Collection (WoSCC) and CiteSpace to analyze 7,199 publications related to radiomics in oncology and generated knowledge maps for the first time, to our knowledge. Since radiomics provides a new means for clinicians to examine entire tumors with rather minimally invasive methods, we sought to provide a more comprehensive understanding of the ever-changing field of radiomics. Furthermore, to encourage researchers from various disciplines to actively and creatively participate in practicing radiomics, we conducted our bibliometric analysis based on relevant literature in this field to outline the countries, authors, institutions, and journals that have made significant contributions to this field. In applying this method, we also identify areas of focus and future trends.

## METHODS

### Data Acquisition and Search Strategy

Relevant literature was collected from the WoSCC. The following search terms were employed, the searching formula is also presented in **Supplementary Figure S2**.

$$TS = (image^* \text{ OR } picture^* \text{ OR } photograph^* \text{ OR } X-ray^* \text{ OR } CT \text{ } OR \text{ } MRI \text{ OR } panorama^* \text{ OR }$$
$$Computer \text{ } Tomography \text{ OR } Magnetic \text{ } Resonance \text{ } Imaging \text{ OR } tomography \text{ OR } PET \text{ } CT$$
$$AND$$
$$(TS = (AI \text{ OR } Artificial \text{ } Intelligence \text{ OR } deep \text{ } learning \text{ OR } machine \text{ } learning \text{ OR } computational \qquad (1)$$
$$intellegen^* \text{ OR } Convolutional \text{ } Neural \text{ } Network \text{ OR } CNN) \text{ OR}(TS = radiomic^*)$$
$$AND$$
$$TS = (tumor^* \text{ OR } cancer^* \text{ OR } carcino^* \text{ OR } onco^*)$$

$$(TS = radiomic^*)$$
$$AND \qquad\qquad (2)$$
$$TS = (tumor^* \text{ OR } cancer^* \text{ OR } carcino^* \text{ OR } onco^*)$$

The time interval was set to 2011 to 2020. Only articles and reviews were included, and no language restrictions were applied. The search and download process was carried out on March 14, 2021 to eliminate substantial errors caused by daily database updates. Given that data were directly downloaded from the database, ethical approval was not required.

### Data Analysis

CiteSpace V was used to remove duplicates and analyze the 7,199 unique records exported. Then, to visualize emerging trends and areas of focus in tumor radiomics research, CiteSpace was applied to generate knowledge mappings of countries, institutions, co-occurrences of keywords, references, authors and co-cited authors, and co-cited journals (28). With each year covered by a dataset assigned a different color, CiteSpace uses colorful node edges or

crosses to discriminate between different research objects, including countries, institutions, words, references, etc. The size of rings on nodes indicates the number of publications for each node. Purple rings surrounding circles indicate the centrality of nodes. Moreover, "Burst detection" is a function provided by the software to detect current and prospective areas of focus. The detection objects of this function can include noun phrases, keywords included in abstracts, papers, and so on, revealing words, or papers undergoing citation bursts in a given period. When the time period is set to the present, this means that some keywords or papers are undergoing a citation burst, which may indicate further prospects for a given field. The impact factors (IFs) for all publications were documented based on the Journal Citation Report (2019).

Generally, the productivity of individuals, journals, and countries can be measured by the total number of papers, whereas the total counts of citations of authors, journals, or references measure the total impact. Defined as the maximum value of $h$ such that the given author/journal has published at least $h$ papers that have each been cited at least $h$ times, the H index is used to characterize a researcher's output in scientific research. The impact factor (IF), defined as a scientometric index, is also a measurement of journals and articles (32). The value is calculated as the average number of citations a publication receives in 2 or 5 years as indexed by the Web of Science. Furthermore, co-citation is defined as the number of times two documents are cited together (33); that is, when two publications or authors are cited at the same time, they may focus on the same theme of research, which may indicate their cooperation. A burst of an event refers to a surge in the frequency of a certain event, such as the emergence of a keyword or the citation of a specific article (34). These parameters allow us to identify productive institutions and their countries and outstanding individuals in the studied field. **Figure 1** illustrates the research steps of this study.

## RESULTS

### State of Publication Output

A total of 7,199 publications were examined in the present study and include 6,417 (89.1%) original articles and 782 (10.9%) reviews. **Figure 2** shows the chronological distribution of the publications for 2011 to 2020. With technological breakthroughs in AI, an increasing number of researchers have been attracted to radiomics in oncology. As depicted in the diagram, the number of articles and reviews grew steadily in the first 4 years. From 2015 to 2020, the annual number of publications grew exponentially and peaked in 2020.

### Active Countries and Institutions

The included publications were published in 82 countries and regions over the last decade. The top 10 contributors are presented in **Supplementary Table S1**, and the cooperative relationships among them are shown in **Supplementary Figure S3**. When counting the number of publications, the USA (2,280) ranks first, followed by China (2136), India (456), Germany (454), and England (401).
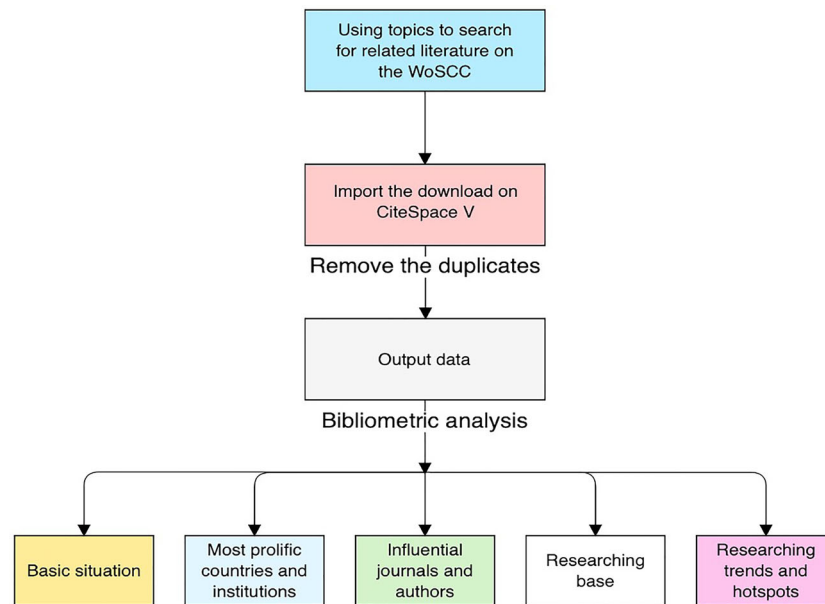
**FIGURE 1** | Workflows of this study.
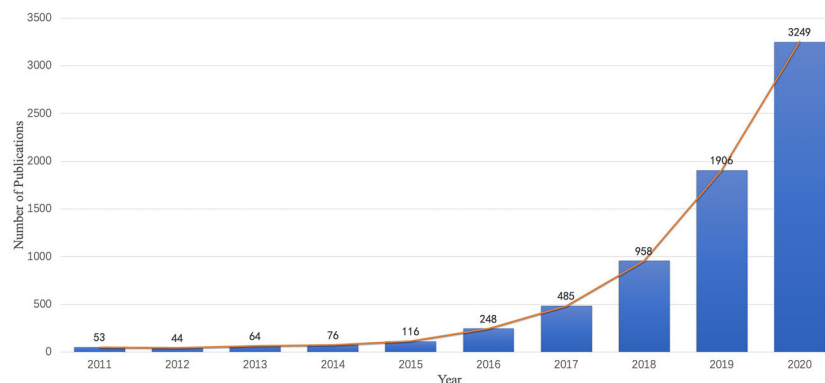


**FIGURE 2** | Chronological distribution of publications in radiomics for oncology from 2011 to 2020.

Overall, 715 institutions contributed to this field. **Supplementary Table S1** shows the 10 most productive institutions, and **Supplementary Figure S4** shows the cluster of institutions engaged in radiomics research in oncology by keyword. The most productive institution is the Chinese Academy of Sciences with 224 publications, followed by Sun Yat-Sen University (168), Fudan University (155), Harvard Medical School (145), and Stanford University (136). The three most prolific institutions are Chinese universities.

## Productive Journals

A total of 1,247 journals published articles or reviews in this field. We list the 10 most productive journals with their IFs in this field in **Supplementary Table S2** and provide an overdual map of citing and cited journals in **Figure 3**. According to statistics from

the WoSCC, *Scientific Reports* published 253 publications over the last decade and thus ranks first. *Medical Physics* ranks second (244 publications), followed by *IEEE Access* (241 publications), *European Radiology* (213 publications), and *Frontiers in Oncology* (180 publications).

The 10 most co-cited journals are given in **Supplementary Table S2**. Among co-cited journals, *Radiology* was cited the most (3258 times), followed by *IEEE Transactions on Medical Imaging* (2800 times), *Scientific Reports* (2410 times), *PLoS One* (2407 times), and *Medical Physics* (2320 times).

## Productive Authors

More than 2,000 authors have contributed to this field of research. **Supplementary Table S3** shows the 10 most prolific authors, and
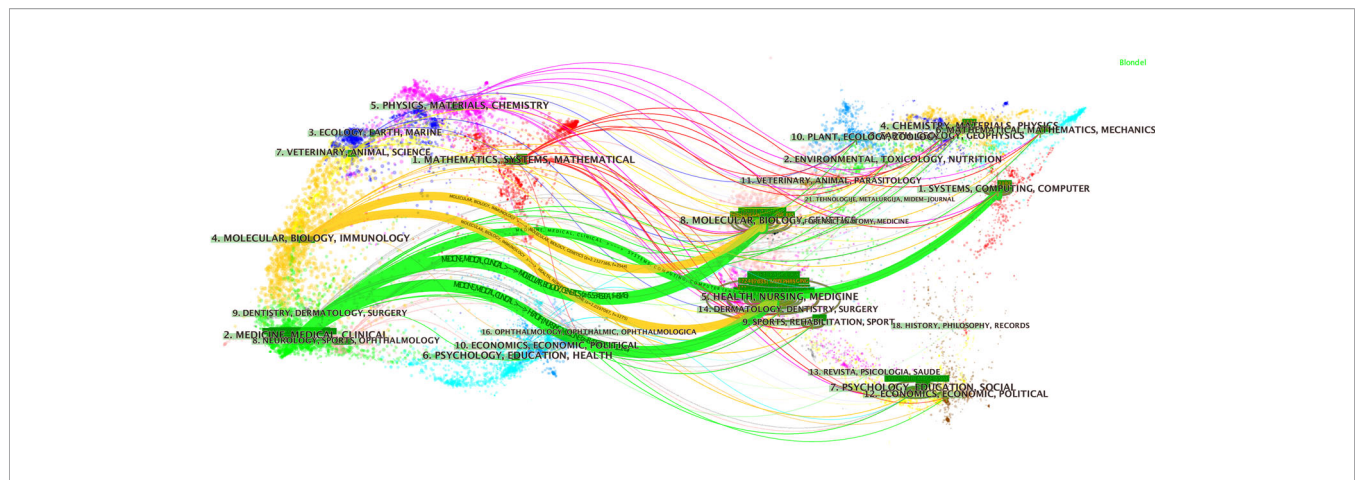
**FIGURE 3** | The dual map overlay of journals. This figure can be divided into two sections. Each dot represents one journal, and this knowledge map uses different colors to symbolize journals from different subjects. On the left, there are the citing journals of this field, and on the right, lays the cited journals in this field. The waves link to two sides means the publications on the journals of the left side may cite publications from the journals on the other side. For example, publications on journals in the field of medicine medical and clinical (labeled 2 on the left), may refer to the publications on the journal of systems, computing, and computer.

**Figure 4** presents a timeline of authors' contributions to this field. Jie Tian is identified as the most productive author with 121 publications, and Anant Madabushi (56 publications) ranks second, followed by Dong Di (52 publications), Philippe Lambin (49 publications), and Zhenyu Liu (45 publications).

**Supplementary Table S3** shows the leading 10 authors in terms of numbers of citations. The author with the most citations is Philippe Lambin (1,171 times), followed by Alex Krizhevsky (1,090 times), Robert J. Gillies (1,066 times), Hugo J. W. L Aerts (1,052 times), and Yann Lecun (1,020 times). **Supplementary Figure S5** presents the authors' potential cooperative relationships, as links between nodes indicate instances where authors are cited together. Since the authors on the left mainly focus on applications of radiomics while authors on the right have most laid the foundations of this field, Robert M. Haralick, shown in the middle of the network, has contributed in connecting applications and algorithms of radiomics in oncology (12).

With more than 2,000 author contributions, the knowledge map of cited authors provides information regarding the most influential authors and the collaborative relationships among them.

## Popular References

CiteSpace provides a mixed map of terms and co-cited references, as shown in **Figure 5**. We present the 10 most cited references in **Table 1**. Of the 6,218 publications shown, an article published by Robert J. Gillies et al. in 2016 ranks first with 1,036 citations (9). This report describes the processes, challenges, and opportunities of radiomics in detail, particularly in reference to the field of oncology.

The burst detection results show articles that have attracted the attention of peer scientists. Citation bursts note the duration and strength of each burst, or the duration and intensity of burst status, respectively (34). **Figure 6** shows the 100 references with the strongest citation bursts. The citation burst analysis shows that a
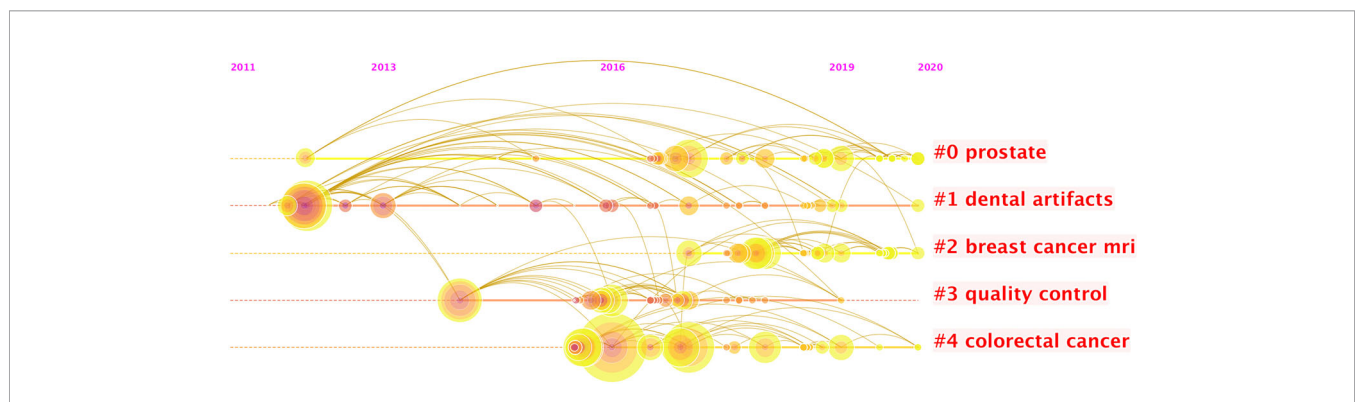


**FIGURE 4** | The timeline view listed authors by clustering through keywords. Each node represents one author. The position of the node here represents the time of an author's first publication. There were 15 clusters of keywords. In each cluster, the size of each node shows the contribution of the author. It seems that the keywords "prostate" and "breast cancer MRI" occur most recently, which suggest the active participations of researchers in practicing radiomics for oncology related to them, and it also shows that researchers have been practicing radiomics research related to dental artifacts since it lasts the longest duration.
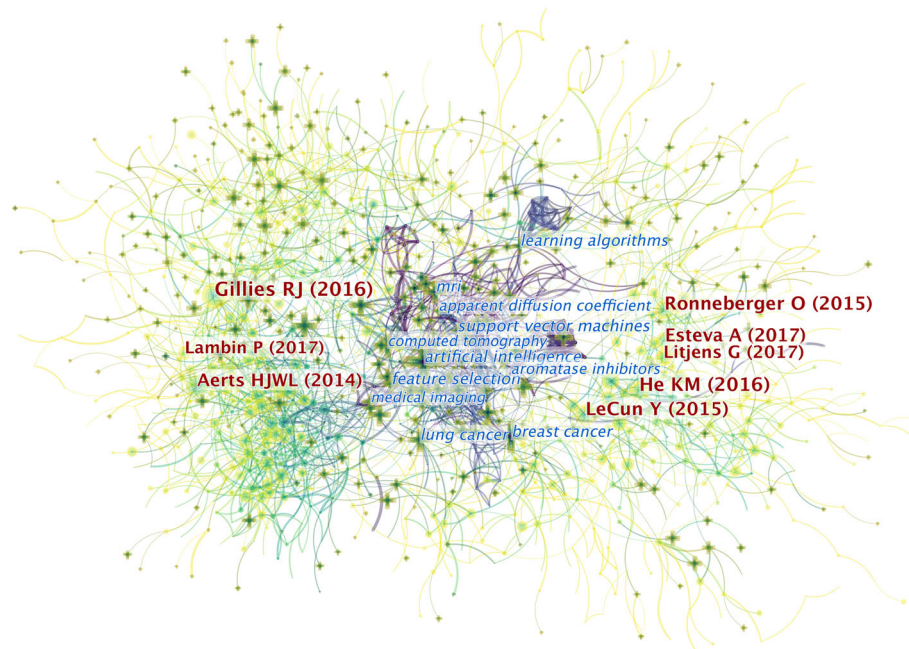
**FIGURE 5** | The mixed science map consists of the most cited noun phrases in publications and co-cited references in this field. By doing so, we illustrate the most co-cited references and the noun phrases in this field and uncover the relationship between them. There are two types of shapes in this picture. Each cross symbolizes a noun phrase, and each node represents a piece of co-cited references. There are links between the crosses and the circles. The links between two circles or noun phrases indicate there are some relationships between two pieces of articles or two phrases since they can be cited together. Also, the links between circles and crosses indicate that a piece of paper can be cited with certain noun phrases. Generally, there are three domains of this map. On the right and the left sides lay the most co-cited articles in the field. The ones on the left are mainly related to the definition and application of radiomics while the references on the right are mostly related to the AI algorithms involved in this field. In the middle are the most-cited noun phrases in the articles in this field. As indicated in the picture, the publications that contain these phrases are the bridges to relate publications from both sides. It illustrated that the most-cited articles from both sides may focus on conducting researches related to the noun phrases in the middle.

**TABLE 1** | The top 10 co-cited references.

| Rank | Title | Author | NOC |
|---|---|---|---|
| 1 | Radiomics: Images Are More than Pictures, They Are Data | Robert J. Gillies et al. | 1,036 |
| 2 | Deep Residual Learning for Image Recognition | Kaiming He et al. | 831 |
| 3 | Deep Learning | Yann LeCun et al. | 741 |
| 4 | U-Net: Convolutional Networks for Biomedical Image Segmentation | Olaf Ronneberger et al. | 707 |
| 5 | Decoding Tumour Phenotype by Noninvasive Imaging Using a Quantitative Radiomics Approach | Hugo J.W.L.Aerts et al. | 638 |
| 6 | Dermatologist-Level Classification of Skin Cancer With Deep Neural Networks | Andre Esteva et al. | 583 |
| 7 | A Survey on Deep Learning in Medical Image Analysis | Geert Litjens et al. | 470 |
| 8 | Radiomics: The Bridge between Medical Imaging and Personalized Medicine | Phillipe Lambin et al. | 429 |
| 9 | ImageNet Classification With Deep Convolutional Neural Networks | Alex Krizhevsky et al. | 386 |
| 10 | Fully Convolutional Networks for Semantic Segmentation | Jonathan Long et al. | 377 |

*NOC, number of citations.*

publication by Hugo J. W. L Aerts et al. earned the highest burst value (95.69) from 2015 to 2020 (35). This result indicates that this publication underwent a citation burst of the highest intensity from 2015 to 2020. The article reports that the radiomics data of cancer patients contain prognostic information and are associated with underlying gene expression patterns.

## Keyword Research

Over 800 keywords were extracted from publications. **Figure 7** shows the keywords mentioned most frequently in

publications. In terms of frequency, the term "artificial intelligence" ranks first (2,931 times), followed by "oncology" (1,972 times), "radiomics" (1,437 times), "classification" (1,436 times), "diagnosis" (1,020 times), and "segmentation" (995 times).

We identify the top 60 keywords with citation bursts, as shown in **Figure 8**, and we provide five keywords with the strongest recent citation bursts in **Table 2**. Of the 60 keywords with the strongest citation bursts, "radiation therapy" shows the highest burst strength level of 16.81. "Test retest", "sarcoma",

## Top 100 References with the Strongest Citation Bursts



**FIGURE 6** | Detection of top 100 references with the strongest citation bursts.
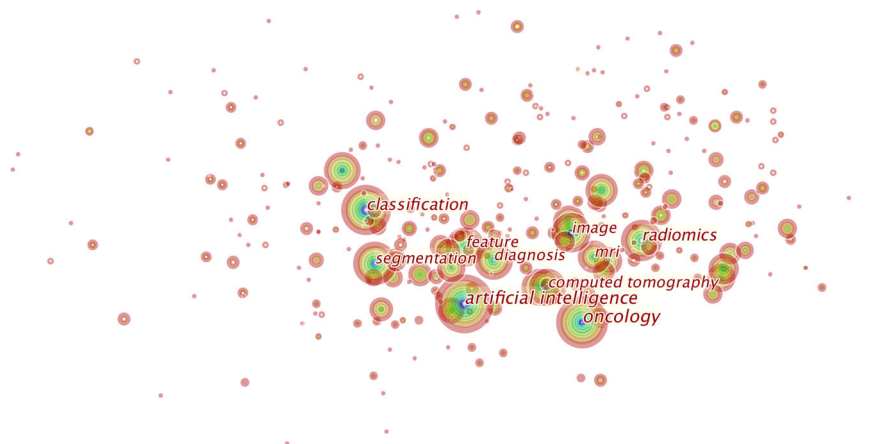


**FIGURE 7** | This is the knowledge map of the most cited keywords in this field. Each node represents a keyword, and the sizes of rings on the node denote the number of publications related to the keyword in a certain year. This map suggests the hotspots in this research field.

"statistics", "intensity-modulated radiotherapy," and "genomics" are keywords with recent citation bursts, and the term "test retest" achieves the highest burst strength and the longest duration for 2016 to 2020. These keywords indicate that the reproducibility and statistical methods of radiomics, the relationship between radiomics and genomic types, and applications of radiomics to sarcomas and intensity-modulated radiotherapy are major focuses of research in this field.

### Top 60 Keywords with the Strongest Citation Bursts

| Keywords | Year | Strength | Begin | End |
|---|---|---|---|---|
| feature selection | 2011 | 12.77 | 2011 | 2017 |
| image analysis | 2011 | 10.52 | 2011 | 2016 |
| positron emission tomography | 2011 | 10.25 | 2011 | 2017 |
| aromatase inhibitor | 2011 | 10.16 | 2011 | 2017 |
| artificial neural network | 2011 | 6.34 | 2011 | 2016 |
| postmenopausal women | 2011 | 6.26 | 2011 | 2016 |
| cancer detection | 2011 | 5.8 | 2011 | 2017 |
| mammography | 2011 | 4.96 | 2011 | 2014 |
| classification | 2011 | 4.43 | 2011 | 2014 |
| randomized trial | 2011 | 4.07 | 2011 | 2013 |
| tissue | 2011 | 5.86 | 2012 | 2016 |
| lesion | 2011 | 5.71 | 2012 | 2015 |
| segmentation | 2011 | 5.66 | 2012 | 2013 |
| diffusion | 2011 | 5.18 | 2012 | 2017 |
| masse | 2011 | 5.18 | 2012 | 2017 |
| tamoxifen | 2011 | 4.09 | 2012 | 2016 |
| radiation therapy | 2011 | 16.81 | 2013 | 2017 |
| cell lung cancer | 2011 | 11.76 | 2013 | 2018 |
| quantification | 2011 | 6.05 | 2013 | 2017 |
| volume | 2011 | 5.8 | 2013 | 2017 |
| textural feature | 2011 | 5.74 | 2013 | 2018 |
| glioblastoma multiforme | 2011 | 5.69 | 2013 | 2018 |
| magnetic resonance spectroscopy | 2011 | 5.06 | 2013 | 2015 |
| registration | 2011 | 4.65 | 2013 | 2018 |
| pattern recognition | 2011 | 4.28 | 2013 | 2017 |
| prostate cancer | 2011 | 4.15 | 2013 | 2016 |
| deformable image registration | 2011 | 5.15 | 2014 | 2017 |
| evolution | 2011 | 4.9 | 2014 | 2017 |
| parameter | 2011 | 4.55 | 2014 | 2017 |
| computer-aided detection | 2011 | 4.11 | 2014 | 2016 |
| reproducibility | 2011 | 10.36 | 2015 | 2017 |
| image feature | 2011 | 5.78 | 2015 | 2018 |
| tumor response | 2011 | 5.7 | 2015 | 2016 |
| survival | 2011 | 5.27 | 2015 | 2015 |
| mammogram | 2011 | 4.71 | 2015 | 2016 |
| reduction | 2011 | 3.77 | 2015 | 2016 |
| f 18 fdg pet | 2011 | 13.59 | 2016 | 2018 |
| tumor heterogeneity | 2011 | 7.23 | 2016 | 2018 |
| big data | 2011 | 6.75 | 2016 | 2018 |
| fdg pet radiomics | 2011 | 6.68 | 2016 | 2017 |
| test retest | 2011 | 5.97 | 2016 | 2020 |
| tumorheterogeneity | 2011 | 5.18 | 2016 | 2018 |
| scan | 2011 | 4.97 | 2016 | 2018 |
| pet image | 2011 | 4.77 | 2016 | 2017 |
| texture | 2011 | 4.54 | 2016 | 2017 |
| spectroscopy | 2011 | 4.49 | 2016 | 2018 |
| random forest | 2011 | 4.29 | 2016 | 2017 |
| architecture | 2011 | 4.04 | 2016 | 2018 |
| intratumor heterogeneity | 2011 | 3.99 | 2016 | 2018 |
| variability | 2011 | 3.95 | 2016 | 2017 |
| nuclei | 2011 | 5.53 | 2017 | 2017 |
| quantitative assessment | 2011 | 5.39 | 2017 | 2018 |
| intratumoral heterogeneity | 2011 | 3.92 | 2017 | 2018 |
| irradiation | 2011 | 3.92 | 2017 | 2018 |
| lung nodule | 2011 | 4.6 | 2018 | 2018 |
| growth factor receptor | 2011 | 4.44 | 2018 | 2018 |
| genomics | 2011 | 3.82 | 2018 | 2020 |
| sarcoma | 2011 | 4.7 | 2019 | 2020 |
| statistics | 2011 | 4.1 | 2019 | 2020 |
| intensity modulated radiotherapy | 2011 | 3.91 | 2019 | 2020 |

**FIGURE 8** | Detection of top 60 keywords with the strongest citation bursts.

**TABLE 2** | Top 5 keywords in the network burst recently.

| Rank | Keywords | Strength | Begin | End |
|---|---|---|---|---|
| 1 | Test retest | 5.97 | 2016 | 2020 |
| 2 | Genomics | 3.82 | 2018 | 2020 |
| 3 | Sarcoma | 4.7 | 2019 | 2020 |
| 4 | Statistics | 4.1 | 2019 | 2020 |
| 5 | Intensity-modulated radiotherapy | 3.91 | 2019 | 2020 |

## DISCUSSION

To our knowledge, this is the first bibliometric analysis of radiomics in oncology. This article provides an in-depth and visualized analysis of publications of this field, which may help researchers gain a basic understanding, develop areas of focus and trends and pursue further practice in this field.

The 10 leading countries include four Asian countries, two American countries, and four European countries. The USA has contributed a great volume of publications (2,280) and has collaborated frequently with other countries. Publications from the USA and China comprise >40% of all publications. Among the 10 leading institutional contributors, the Chinese Academy of Sciences ranks highest in terms of volume. Furthermore, six of the 10 leading institutional contributors are Chinese universities. However, even with the country's large volume of publications and despite including three of the five most prolific institutions, collaboration in China has been rare and limited, suggesting that although China has been carrying out radiomics research in oncology for the last decade, extensive collaborative work is needed.

*Scientific Reports* has been the leading contributions in this field with an IF of 3.998. *Radiology* has published fewer publications but is the most cited and hence may be viewed as an influential journal in this field with a high IF of 7.931; several ground-breaking articles have been published in *Radiology* (8, 9, 35).

We list the 10 leading contributing authors and the 10 most frequently cited authors. These authors have devoted themselves to conducting research in oncology radiomics and laid the foundation of basic knowledge in this field. **Supplementary Figure S5** shows the corelationships of authors, which may indicate collaboration. With more than 100 publications, Jie Tian is the most prolific author with an H-index of 65. In 2016, to assist researchers in evaluating malignancy uncertainty, Jie Tian and his team developed a multicrop convolutional neural network to effectively characterize nodules instead of carefully segmenting using imaging and time-consuming feature extraction procedures (36). With the most citations of the 10 leading co-cited authors, Philippe Lambin is considered a pioneering and influential researcher in this field. He was the first to define radiomics and has made substantial efforts to standardize radiomics research (2, 8).

Among the 10 most cited references, an article by Robert J. Gilles et al. has been cited most, as this work provides basic information for researchers seeking to participate in radiomics research on work procedures, applications, challenges, and potential uses (9). In addition, three of the five most influential references focus on algorithms or basic knowledge of this field (37–39).

Keywords can represent areas of focus in a given field as shown in **Figure 7**, and we identified "artificial intelligence," "tumors," "classification," "segmentation," and "diagnosis" as areas of focus in this field. We summarize these areas as follows:

1. **AI**: Much of the development of radiomics relies on AI algorithms, as they can mimic human performance. To date, AI algorithms have been most widely used in radiomics in oncology. In fact, the development of AI can directly influence this field. In the past, radiomics studies most rely on traditional machine learning methods: the random forest, decision tree, and regression algorithms. These algorithms mostly rely on manual segmentation, require a huge amount of matrix manipulations, and can only perform well in small given datasets. These time-consuming processes hindered the applications of radiomics studies. Unlike the traditional machine learning methods, deep learning is a subset of AI that can acquire discriminative features from data. Instead of requiring experienced radiologists to evaluate medical images, deep learning algorithms excel at delineating and monitoring cancerous lesions and can translate medical images into quantitative data to be automatically analyzed. This approach has been most frequently applied in radiomics in oncology (40, 41). Moreover, radiomic models integrated by AI algorithms and clinical features can increase the capacity to judge individual treatment. With efforts made to develop AI technology, groundbreaking AI algorithms may enable computers to act more similar to human beings in the future.

2. **Oncology**, **radiomics**, **diagnosis**, **and classification**: When analyzing a medical image, clinicians usually depend on their personal experience, which is subjective, and results therefore vary among different radiologists (41). By applying quantitative data extracted from medical images for analysis, radiomics provides a new objective means using AI algorithms to detect lesions. Radiomics research today mainly focuses on diagnosis and classification (42, 43). For example, a computer-assisted diagnosis system can automatedly identify cancerous lesions with images and videos (44), and researchers have made considerable efforts to build radiomics models to classify identical cancerous lesions and lymph node status (5, 45).

3. **Segmentation**: To achieve reliable radiomic models in oncology, robustly and precisely delineating lesions is essential. However, traditional manual segmentation usually takes a very long time to perform, generates interobserver variability, and requires the involvement of experienced radiologists for analysis. To address these problems, automated segmentation methods have been established in radiomic research in oncology (46, 47).

Our burst detection results of references reveal articles that have attracted the attention of peer scientists (34). It seems that the ground-breaking articles and reviews with the highest citation burst strength and longest durations were published in 2015 (35). Since citation bursts may help researchers obtain a quick review of research focuses and perspectives, below we list some recent radiomic studies identified by citation burst detection that may be defined as ground-breaking works in this field in leading prospective research:

1. Some radiomic signatures have prognostic power, and there is a prognostic radiomic signature associated with underlying gene expression patterns (35);

2. Researchers have built a prostate cancer MRI computer-aided detection system evaluated on a per-patient basis and compared with the prospective performance of radiologists (48);

3. Multiparametric MRI has been used to accurately locate and segment rectal cancer (49);

4. Researchers have found that the preprocessing of CT images may influence feature volume dependence and its significance in univariate analysis models (50);

5. Using simple linear regression, a subset of radiomic features extracted from CT and cone-beam CT images can be interchangeable, and cone-beam CT radiomics can be used as a prognostic imaging biomarker (51).

We also found that radiomics research in oncology may focus on the following five keywords: test-retest, sarcoma, statistics, intensity-modulated radiotherapy, and genomics. These terms may reflect prospective areas of focus in radiomics for oncology and are summarized as follows:

1. **Test-retest and statistics**: Test-retest methods involve repeating the process of acquiring medical images to test the stability of radiomic features that they may generate. As there are many radiomics models based on various imaging parameter settings and algorithms, test-retest studies of these radiomics features have become essential for future applications. Researchers in this field have engaged in examining the reproducibility of radiomic features (52–54). However, according to a systematic review by Alberto Traverso et al., under different settings, there is no consensus on the most reproducible features (55). Moreover, the statistical method used in radiomics is of great significance. To conduct radiomic research of high quality, standardized statistical methods are essential (56, 57). Likewise, different facilities or radiologists involved during the acquisition of medical images and various analysis algorithms used may lead to bias in radiomics models. Sometimes, owing to the nonstandardized procedure of medical imaging, pictures are too distorted or of low resolution to be read by an AI algorithm. Inexperienced doctors may also fail to identify lesions. Such bias may lead to the conversion of the output result when establishing a radiomic model. Thus, future research may focus on the reproducibility of radiomics models.

2. **Sarcoma**: To date, radiomics has been used in the classification (58) and prediction of metastasis for sarcomas (59); additionally, nomograms have been used in survival prediction (60) for sarcomas. There have been relatively few radiomic analyses of sarcomas, since the prevalence of this disease is relatively low. However, this result implies that more radiomic studies should focus on sarcomas.

3. **Intensity-modulated radiotherapy**: Researchers have used radiomic models to predict patients' responses to intensity-modulated radiotherapy (25, 61). As datasets expand, further applications of radiomics in predicting prognosis for patients undergoing radiotherapy may be identified.

4. **Genomics**: Radiogenomics involves the use of data generated by radiomic analysis to correlate with genomic patterns. Evidence has shown that radiomics features are correlated with gene patterns (35, 62, 63). Radiomic features that do not

relate to gene types may supply independent information, which may enable precision medicine (9).

Our study has some limitations. First, we only focus on literature included in the WoSCC; thus, not all publications are considered and citation counts may be underestimated. Second, CiteSpace only analyses the main conclusions of publications instead of reviewing full texts; thus, some information may have been overlooked. Finally, our results only reflect the current state of radiomics research in oncology, as data are typically prone to frequent changes.

To our knowledge, this is the first analysis of radiomics in oncology conducted from the perspective of bibliometrics. The presented results may help researchers gain a basic understanding and detect areas of focus and trends and may encourage further practice in this field.

## CONCLUSION

Due to advances in technology, radiomics in oncology has significantly evolved over the past decade. According to results found through CiteSpace, we can conclude that current studies in this field focus on AI algorithms and on using radiomics to realize automated segmentation and classify and diagnose lesions.

Such a trend may be attributed to the rapid growth of AI algorithms, which can identify medical images. Researchers have used this new tool to train machines to identify lesions. Many articles on such issues have emerged in the last decade. This may explain why in recent decades the number of papers in this field has sharply increased. However, even with AI algorithms surpassing the performance of physicians, we must carefully validate them.

Since CNN is the most widely used AI algorithm in this field, the terms "test-retest" and "statistics" were identified during our burst detection of keywords. This finding is attributed to the characteristics of CNN itself. Unlike other algorithms, CNN requires magnitude data to train models. When such data are available, CNN can establish a robust model with high precision for clinical decision-making. However, the accessibility of medical data is always limited, and many studies mainly establish their models based on relatively small datasets. Such models may perform well only for the studied datasets. Meanwhile, a given CNN network can only perform a single defined task based on the given labels and dataset. In regard to combining several radiomic models in identifying the same oncolesions, relabeling images and retraining models may be needed due to the limitations of CNN. Therefore, model test-retest and statistical methods are likely to be widely used in future work in this field. The identification of robust features and use of standardized statistical methods may offer opportunities for the combination of various CNN models. If this can be eventually achieved, robust CNN models may be developed and may surpass the capacities of human beings. Thus, with the application of CNN models, researchers can realize their limitations. The field will thus develop with the use of a new kind of algorithm that may overcome the limitations of the CNN network (1, 64).

From our findings, we must note that not all kinds of lesions have received equal attention. According to the provided evidence, lung, breast, and prostate cancer have been the most frequently studied malignancies. According to 2020 cancer statistics (65), some of the most prevalent cancers have not been widely studied, including colon and rectum tumors, bladder cancer, kidney cancer, etc. Additionally, some region-related cancers, including liver, gastric, and oral cancer (66), which are major forms of malignancy in Asia, have been less widely reported on. This may relate to barriers of radiomics technology and poor cross-country collaboration. "Sarcoma" was highlighted in the burst detection analysis, indicating that research has focused on this field; however, only a few studies have specifically investigated this type of malignancy. This trend may be due to the relatively low disease incidence and fast progression of the disease, rendering the availability of imaging datasets more constrained. An increased use of radiomics to diagnose the above less-reported malignancies may accompany the development of AI algorithms and the sharing of databases across regions and countries.

Intensity-modulated radiotherapy and genomics may be more heavily integrated with radiomics in the future, potentially because it usually takes a long time to establish a radiomic model that can predict the survival rate of intensity-modulated radiotherapy. With the establishment of a large dataset, the use of radiomics to predict the prognoses of patients undergoing such therapy may increase in the future. Additionally, studies have found relationships between cancerous lesions and their gene patterns. Since radiomics is a risk-free means to examine the gene patterns of oncology, it may become the next area of focus of this field.

In conclusion, with active participation and regulated practices, radiomics may be applied in every phase of oncology treatment, which could further advance the development of oncology and will likely change the state of oncology imaging.

## 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

YL conceived and designed the structure of this manuscript. HD, CW, NL, WS, QZ, YH, and ZJ wrote the paper. YL revised the paper. All authors contributed to the article and approved the submitted version.

## FUNDING

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2021.689802/full#supplementary-material

# REFERENCES

1. Savadjiev P, Chong J, Dohan A, Vakalopoulou M, Reinhold C, Paragios N, et al. Demystification of AI-Driven Medical Image Interpretation: Past, Present and Future. *Eur Radiol* (2019) 29(3):1616–24. doi: 10.1007/s00330-018-5674-x

2. 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

3. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-Level Classification of Skin Cancer With Deep Neural Networks. *Nature* (2017) 542(7639):115–8. doi: 10.1038/nature21056

4. Guo L, Xiao X, Wu C, Zeng X, Zhang Y, Du J, et al. Real-Time Automated Diagnosis of Precancerous Lesions and Early Esophageal Squamous Cell Carcinoma Using a Deep Learning Model (With Videos). *Gastrointest Endosc* (2020) 91(1):41–51. doi: 10.1016/j.gie.2019.08.018

5. Yin P, Mao N, Zhao C, Wu J, Chen L, Hong N. A Triple-Classification Radiomics Model for the Differentiation of Primary Chordoma, Giant Cell Tumor, and Metastatic Tumor of Sacrum Based on T2-Weighted and Contrast-Enhanced T1-Weighted MRI. *J Magn Reson Imaging* (2019) 49(3):752–9. doi: 10.1002/jmri.26238

6. Nie K, Shi L, Chen Q, Hu X, Jabbour SK, Yue N, et al. Rectal Cancer: Assessment of Neoadjuvant Chemoradiation Outcome Based on Radiomics of Multiparametric MRI. *Clin Cancer Res* (2016) 22(21):5256–64. doi: 10.1158/1078-0432.CCR-15-2997

7. Bologna M, Calareso G, Resteghini C, Sdao S, Montin E, Corino V, et al. Relevance of Apparent Diffusion Coefficient Features for a Radiomics-Based Prediction of Response to Induction Chemotherapy in Sinonasal Cancer. *NMR Biomed* (2020), e4265. doi: 10.1002/nbm.4265

8. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, et al. Radiomics: The Bridge Between Medical Imaging and Personalized Medicine. *Nat Rev Clin Oncol* (2017) 14(12):749–62. doi: 10.1038/nrclinonc.2017.141

9. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More Than Pictures, They Are Data. *Radiology* (2016) 278(2):563–77. doi: 10.1148/radiol.2015151169

10. Clarke LP, Nordstrom RJ, Zhang H, Tandon P, Zhang Y, Redmond G, et al. The Quantitative Imaging Network: NCI's Historical Perspective and Planned Goals. *Transl Oncol* (2014) 7(1):1–4. doi: 10.1593/tlo.13832

11. Buckler AJ, Bresolin L, Dunnick NR, Sullivan DC. A Collaborative Enterprise for Multi-Stakeholder Participation in the Advancement of Quantitative Imaging. *Radiology* (2011) 258(3):906–14. doi: 10.1148/radiol.10100799

12. Haralick RM, Shanmugam K, Dinstein I. TEXTURAL FEATURES FOR IMAGE CLASSIFICATION. *IEEE Trans Syst Man Cybernetics* (1973) SMC3(6):610–21. doi: 10.1109/TSMC.1973.4309314

13. Lo SCB, Lou SLA, Lin JS, Freedman MT, Chien MV, Mun SK. Artificial Convolution Neural Network Techniques and Applications for Lung Nodule Detection. *IEEE Trans Med Imaging* (1995) 14(4):711–8. doi: 10.1109/42.476112

14. Segal E, Sirlin CB, Ooi C, Adler AS, Gollub J, Chen X, et al. Decoding Global Gene Expression Programs in Liver Cancer by Noninvasive Imaging. *Nat Biotechnol* (2007) 25(6):675–80. doi: 10.1038/nbt1306

15. Diehn M, Nardini C, Wang DS, McGovern S, Jayaraman M, Liang Y, et al. Identification of Noninvasive Imaging Surrogates for Brain Tumor Gene-Expression Modules. *Proc Natl Acad Sci U.S.A.* (2008) 105(13):5213–8. doi: 10.1073/pnas.0801279105

16. 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

17. Parmar C, Leijenaar RT, Grossmann P, Rios Velazquez E, Bussink J, Rietveld D, et al. Radiomic Feature Clusters and Prognostic Signatures Specific for Lung and Head & Neck Cancer. *Sci Rep* (2015) 5:11044. doi: 10.1038/srep11044

18. Sun R, Limkin EJ, Vakalopoulou M, Dercle L, Champiat S, Han SR, et al. A Radiomics Approach to Assess Tumour-Infiltrating CD8 Cells and Response to Anti-PD-1 or Anti-PD-L1 Immunotherapy: An Imaging Biomarker, Retrospective Multicohort Study. *Lancet Oncol* (2018) 19(9):1180–91. doi: 10.1016/S1470-2045(18)30413-3

19. Huang C, Cintra M, Brennan K, Zhou M, Colevas AD, Fischbein N, et al. Development and Validation of Radiomic Signatures of Head and Neck Squamous Cell Carcinoma Molecular Features and Subtypes. *EBioMedicine* (2019) 45:70–80. doi: 10.1016/j.ebiom.2019.06.034

20. Hawkins SH, Korecki JN, Balagurunathan Y, Yuhua G, Kumar V, Basu S, et al. Predicting Outcomes of Nonsmall Cell Lung Cancer Using CT Image Features. *IEEE Access* (2014) 2:1418–26. doi: 10.1109/ACCESS.2014.2373335

21. Huang YQ, Liang CH, He L, Tian J, Liang CS, Chen X, et al. Development and Validation of a Radiomics Nomogram for Preoperative Prediction of Lymph Node Metastasis in Colorectal Cancer. *J Clin Oncol* (2016) 34(18):2157–64. doi: 10.1200/JCO.2015.65.9128

22. Wu W, Parmar C, Grossmann P, Quackenbush J, Lambin P, Bussink J, et al. Exploratory Study to Identify Radiomics Classifiers for Lung Cancer Histology. *Front Oncol* (2016) 6:71. doi: 10.3389/fonc.2016.00071

23. van Timmeren JE, Leijenaar RTH, van Elmpt W, Reymen B, Oberije C, Monshouwer R, et al. Survival Prediction of non-Small Cell Lung Cancer Patients Using Radiomics Analyses of Cone-Beam CT Images. *Radiother Oncol* (2017) 123(3):363–9. doi: 10.1016/j.radonc.2017.04.016

24. Liu Z, Zhang XY, Shi YJ, Wang L, Zhu HT, Tang Z, et al. Radiomics Analysis for Evaluation of Pathological Complete Response to Neoadjuvant Chemoradiotherapy in Locally Advanced Rectal Cancer. *Clin Cancer Res* (2017) 23(23):7253–62. doi: 10.1158/1078-0432.CCR-17-1038

25. Abdollahi H, Mofid B, Shiri I, Razzaghdoust A, Saadipoor A, Mahdavi A, et al. Machine Learning-Based Radiomic Models to Predict Intensity-Modulated Radiation Therapy Response, Gleason Score and Stage in Prostate Cancer. *Radiol Med* (2019) 124(6):555–67. doi: 10.1007/s11547-018-0966-4

26. Oelrich B, Peters R, Jung K. A Bibliometric Evaluation of Publications in Urological Journals Among European Union Countries Between 2000-2005. *Eur Urol* (2007) 52(4):1238–48. doi: 10.1016/j.eururo.2007.06.050

27. Khan MS, Ullah W, Riaz IB, Bhulani N, Manning WJ, Tridandapani S, et al. Top 100 Cited Articles in Cardiovascular Magnetic Resonance: A Bibliometric Analysis. *J Cardiovasc Magn Reson* (2016) 18(1):87. doi: 10.1186/s12968-016-0303-9

28. Chen C. CiteSpace II: Detecting and Visualizing Emerging Trends and Transient Patterns in Scientific Literature. *J Am Soc Inf Sci Technol* (2006) 57(3):359–77. doi: 10.1002/asi.20317

29. Da W, Tao Z, Meng Y, Wen K, Zhou S, Yang K, et al. A 10-Year Bibliometric Analysis of Osteosarcoma and Cure From 2010 to 2019. *BMC Cancer* (2021) 21(1):115. doi: 10.1186/s12885-021-07818-4

30. Yan W, Zheng K, Weng L, Chen C, Kiartivich S, Jiang X, et al. Bibliometric Evaluation of 2000-2019 Publications on Functional Near-Infrared Spectroscopy. *Neuroimage* (2020) 220:117121. doi: 10.1016/j.neuroimage.2020.117121

31. Chen S, Zhang Y, Dai W, Qi S, Tian W, Gu X, et al. Publication Trends and Hot Spots in Postoperative Cognitive Dysfunction Research: A 20-Year Bibliometric Analysis. *J Clin Anesth* (2020) 67:110012. doi: 10.1016/j.jclinane.2020.110012

32. Garfield E. The History and Meaning of the Journal Impact Factor. *Jama-Journal Am Med Assoc* (2006) 295(1):90–3. doi: 10.1001/jama.295.1.90

33. Small H. COCITATION IN SCIENTIFIC LITERATURE - NEW MEASURE OF RELATIONSHIP BETWEEN 2 DOCUMENTS. *J Am Soc Inf Sci* (1973) 24(4):265–9. doi: 10.1002/asi.4630240406

34. Chen C, Dubin R, Kim MC. Emerging Trends and New Developments in Regenerative Medicine: A Scientometric Update (2000 - 2014). *Expert Opin Biol Ther* (2014) 14(9):1295–317. doi: 10.1517/14712598.2014.920813

35. 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

36. Shen W, Zhou M, Yang F, Yu D, Dong D, Yang C, et al. Multi-Crop Convolutional Neural Networks for Lung Nodule Malignancy Suspiciousness Classification. *Pattern Recognition* (2017) 61:663–73. doi: 10.1016/j.patcog.2016.05.029

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

38. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. *Lecture Notes Comput Sci* (2015) 9351:234–41. doi: 10.1007/978-3-319-24574-4_28

39. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE. (2016). p. 770–8.

40. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts H. Artificial Intelligence in Radiology. *Nat Rev Cancer* (2018) 18(8):500–10. doi: 10.1038/s41568-018-0016-5

41. Bi WL, Hosny A, Schabath MB, Giger ML, Birkbak NJ, Mehrtash A, et al. Artificial Intelligence in Cancer Imaging: Clinical Challenges and Applications. *CA Cancer J Clin* (2019) 69(2):127–57. doi: 10.3322/caac.21552

42. El-Dahshan E-SA, Mohsen HM, Revett K, Salem A-BM. Computer-Aided Diagnosis of Human Brain Tumor Through MRI: A Survey and a New Algorithm. *Expert Syst Appl* (2014) 41(11):5526–45. doi: 10.1016/j.eswa.2014.01.021

43. Liu C, Ding J, Spuhler K, Gao Y, Serrano Sosa M, Moriarty M, et al. Preoperative Prediction of Sentinel Lymph Node Metastasis in Breast Cancer by Radiomic Signatures From Dynamic Contrast-Enhanced MRI. *J Magn Reson Imaging* (2019) 49(1):131–40. doi: 10.1002/jmri.26224

44. Guo R, Guo J, Zhang L, Qu X, Dai S, Peng R, et al. CT-Based Radiomics Features in the Prediction of Thyroid Cartilage Invasion From Laryngeal and Hypopharyngeal Squamous Cell Carcinoma. *Cancer Imaging* (2020) 20(1):81. doi: 10.1186/s40644-020-00359-2

45. Ho TY, Chao CH, Chin SC, Ng SH, Kang CJ, Tsang NM. Classifying Neck Lymph Nodes of Head and Neck Squamous Cell Carcinoma in MRI Images With Radiomic Features. *J Digit Imaging* (2020) 33(3):613–8. doi: 10.1007/s10278-019-00309-w

46. Parmar C, Rios Velazquez E, Leijenaar R, Jermoumi M, Carvalho S, Mak RH, et al. Robust Radiomics Feature Quantification Using Semiautomatic Volumetric Segmentation. *PloS One* (2014) 9(7):e102107. doi: 10.1371/journal.pone.0102107

47. Polan DF, Brady SL, Kaufman RA. Tissue Segmentation of Computed Tomography Images Using a Random Forest Algorithm: A Feasibility Study. *Phys Med Biol* (2016) 61(17):6553–69. doi: 10.1088/0031-9155/61/17/6553

48. Litjens G, Debats O, Barentsz J, Karssemeijer N, Huisman H. Computer-Aided Detection of Prostate Cancer in MRI. *IEEE Trans Med Imaging* (2014) 33(5):1083–92. doi: 10.1109/TMI.2014.2303821

49. Trebeschi S, van Griethuysen JJM, Lambregts DMJ, Lahaye MJ, Parmar C, Bakers FCH, et al. Deep Learning for Fully-Automated Localization and Segmentation of Rectal Cancer on Multiparametric Mr. *Sci Rep* (2017) 7(1):5301. doi: 10.1038/s41598-017-05728-9

50. Fave X, Zhang L, Yang J, Mackin D, Balter P, Gomez D, et al. Impact of Image Preprocessing on the Volume Dependence and Prognostic Potential of Radiomics Features in non-Small Cell Lung Cancer. *Trans Cancer Res* (2016) 5(4):349–63. doi: 10.21037/tcr.2016.07.11

51. Shiri I, Rahmim A, Ghaffarian P, Geramifar P, Abdollahi H, Bitarafan-Rajabi A. The Impact of Image Reconstruction Settings on 18F-FDG PET Radiomic Features: Multi-Scanner Phantom and Patient Studies. *Eur Radiol* (2017) 27(11):4498–509. doi: 10.1007/s00330-017-4859-z

52. Leijenaar RTH, Carvalho S, Velazquez ER, Van Elmpt WJC, 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

53. van Velden FH, Kramer GM, Frings V, Nissen IA, Mulder ER, de Langen AJ, et al. Repeatability of Radiomic Features in Non-Small-Cell Lung Cancer [(18)F]FDG-PET/CT Studies: Impact of Reconstruction and Delineation. *Mol Imaging Biol* (2016) 18(5):788–95. doi: 10.1007/s11307-016-0940-2

54. Zhao B, Tan Y, Tsai WY, Qi J, Xie C, Lu L, et al. Reproducibility of Radiomics for Deciphering Tumor Phenotype With Imaging. *Sci Rep* (2016) 6:23428. doi: 10.1038/srep23428

55. Traverso A, Wee L, Dekker A, Gillies R. Repeatability and Reproducibility of Radiomic Features: A Systematic Review. *Int J Radiat Oncol Biol Phys* (2018) 102(4):1143–58. doi: 10.1016/j.ijrobp.2018.05.053

56. Leijenaar RT, Nalbantov G, Carvalho S, van Elmpt WJ, Troost EG, Boellaard R, et al. The Effect of SUV Discretization in Quantitative FDG-PET Radiomics: The Need for Standardized Methodology in Tumor Texture Analysis. *Sci Rep* (2015) 5:11075. doi: 10.1038/srep11075

57. Ha S, Choi H, Paeng JC, Cheon GJ. Radiomics in Oncological PET/CT: A Methodological Overview. *Nucl Med Mol Imaging* (2019) 53(1):14–29. doi: 10.1007/s13139-019-00571-4

58. Hermessi H, Mourali O, Zagrouba E. Deep Feature Learning for Soft Tissue Sarcoma Classification in MR Images *via* Transfer Learning. *Expert Syst Appl* (2019) 120:116–27. doi: 10.1016/j.eswa.2018.11.025

59. Vallieres M, Freeman CR, Skamene SR, El Naqa I. A Radiomics Model From Joint FDG-PET and MRI Texture Features for the Prediction of Lung Metastases in Soft-Tissue Sarcomas of the Extremities. *Phys Med Biol* (2015) 60(14):5471–96. doi: 10.1088/0031-9155/60/14/5471

60. Wu Y, Xu L, Yang P, Lin N, Huang X, Pan W, et al. Survival Prediction in High-Grade Osteosarcoma Using Radiomics of Diagnostic Computed Tomography. *EBioMedicine* (2018) 34:27–34. doi: 10.1016/j.ebiom.2018.07.006

61. Li S, Wang K, Hou Z, Yang J, Ren W, Gao S, et al. Use of Radiomics Combined With Machine Learning Method in the Recurrence Patterns After Intensity-Modulated Radiotherapy for Nasopharyngeal Carcinoma: A Preliminary Study. *Front Oncol* (2018) 8:648. doi: 10.3389/fonc.2018.00648

62. Zhu Y, Mohamed ASR, Lai SY, Yang S, Kanwar A, Wei L, et al. Imaging-Genomic Study of Head and Neck Squamous Cell Carcinoma: Associations Between Radiomic Phenotypes and Genomic Mechanisms *via* Integration of The Cancer Genome Atlas and The Cancer Imaging Archive. *JCO Clin Cancer Inform* (2019) 3:1–9. doi: 10.1200/CCI.18.00073

63. Zwirner K, Hilke FJ, Demidov G, Socarras Fernandez J, Ossowski S, Gani C, et al. Radiogenomics in Head and Neck Cancer: Correlation of Radiomic Heterogeneity and Somatic Mutations in TP53, FAT1 and KMT2D. *Strahlenther Onkol* (2019) 195(9):771–9. doi: 10.1007/s00066-019-01478-x

64. Fujita H. AI-Based Computer-Aided Diagnosis (AI-CAD): The Latest Review to Read First. Igaku Butsuri : Nihon Igaku Butsuri Gakkai Kikanshi. *Japanese J Med Phys an Off J Japan Soc Med Phys* (2020) 40(4):140. doi: 10.1007/s12194-019-00552-4

65. Siegel RL, Miller KD, Jemal A. Cancer Statistics, 2020. *CA: A Cancer J Clin* (2020) 70(1):7–30. doi: 10.3322/caac.21590

66. Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, et al. Cancer Statistics in China, 2015. *CA: A Cancer J Clin* (2016) 66(2):115–32. doi: 10.3322/caac.21338

# Development of a Nomogram Combining Clinical Risk Factors and Dual-Energy Spectral CT Parameters for the Preoperative Prediction of Lymph Node Metastasis in Patients With Colorectal Cancer

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

[1] Department of Radiology, Affiliated Hospital of Qinghai University, Xining, China, [2] Second Clinical School, Lanzhou University, Lanzhou, China, [3] Department of Radiology, Lanzhou University Second Hospital, Lanzhou, China, [4] Key Laboratory of Medical Imaging of Gansu Province, Lanzhou, China, [5] Gansu International Scientific and Technological Cooperation Base of Medical Imaging Artificial Intelligence, Lanzhou, China, [6] Department of Radiology, The Fifth Affiliated Hospital of Zunyi Medical University, Zhuhai, China, [7] Department of Radiology, Sichuan Provincial People's Hospital, Chengdu, China, [8] Department of Critical Medicine, Affiliated Hospital of Qinghai University, Xining, China, [9] Department of Hepatopancreatobiliary Surgery, Affiliated Hospital of Qinghai University, Xining, China, [10] Department of Pharmaceuticals Diagnosis, General Electrics (GE) Healthcare, Beijing, China

**Objective:** This study aimed to develop a dual-energy spectral computed tomography (DESCT) nomogram that incorporated both clinical factors and DESCT parameters for individual preoperative prediction of lymph node metastasis (LNM) in patients with colorectal cancer (CRC).

**Material and Methods:** We retrospectively reviewed 167 pathologically confirmed patients with CRC who underwent enhanced DESCT preoperatively, and these patients were categorized into training ($n = 117$) and validation cohorts ($n = 50$). The monochromatic CT value, iodine concentration value (IC), and effective atomic number (Eff-Z) of the primary tumors were measured independently in the arterial phase (AP) and venous phase (VP) by two radiologists. DESCT parameters together with clinical factors were input into the prediction model for predicting LNM in patients with CRC. Logistic regression analyses were performed to screen for significant predictors of LNM, and these predictors were presented as an easy-to-use nomogram. The receiver operating characteristic curve and decision curve analysis (DCA) were used to evaluate the clinical usefulness of the nomogram.

**Results:** The logistic regression analysis showed that carcinoembryonic antigen, carbohydrate antigen 199, pericolorectal fat invasion, ICAP, ICVP, and Eff-ZVP were

independent predictors in the predictive model. Based on these predictors, a quantitative nomogram was developed to predict individual LNM probability. The area under the curve (AUC) values of the nomogram were 0.876 in the training cohort and 0.852 in the validation cohort, respectively. DCA showed that our nomogram has outstanding clinical utility.

**Conclusions:** This study presents a clinical nomogram that incorporates clinical factors and DESCT parameters and can potentially be used as a clinical tool for individual preoperative prediction of LNM in patients with CRC.

**Keywords: tomography, X-ray computed, colorectal cancer, lymph node metastasis, nomogram**

## INTRODUCTION

According to the latest global cancer statistics, colorectal cancer (CRC) is ranked among the top three cancers in terms of both prevalence and mortality, and its incidence is increasing (1). Accurate preoperative evaluation of lymph node metastasis (LNM) is critical to making a precise treatment plan and evaluating patient prognosis (2, 3). Although histopathological features such as tumor differentiation and lymphatic invasion are closely related to LNM, these features are only available postoperatively and provide limited clinical guidance (4). Accurate assessment of LNM preoperatively provides valuable information for patients with CRC to choose the optimal treatment plan, thereby improving their prognosis.

Despite clinical advances, LNM evaluation remains a challenging issue for radiologists. Non-invasive radiological modalities, such as CT, magnetic resonance imaging, and endoluminal ultrasonography, have been widely utilized in the evaluation of LNM in clinical practice. However, these imaging methods cannot accurately evaluate LNM using criteria such as short-axis diameter, signal heterogeneity, shape, and boundary (2, 5, 6). Therefore, developing more sensitive diagnostic tools for the preoperative prediction of LNM in patients with CRC patients is imperative.

The application of dual-energy spectral computed tomography (DESCT) is considered a milestone in the history of CT diagnosis. DESCT expands the single parameter scanning mode of conventional CT, providing multiple quantitative parameters, such as monochromatic images at energy levels of 40~140 keV, material decomposition images (such as iodine-based or water-based decomposition images), and effective atomic number (Eff-Z) images. Based on this advantage, DESCT has been widely applied in clinical practice for such uses as CRC grading, malignant lymph node (LN) identification, neoadjuvant treatment therapy response evaluation, and microsatellite instability status evaluation (7–10). Previous studies have evaluated the value of DESCT in distinguishing metastatic and non-metastatic LNs preoperatively

in CRC (8, 11); however, the ability of DESCT for predicting LNM in primary CRC has not been evaluated. Several nomograms have been developed to predict LNM in CRC (4, 6, 12). Huang et al. developed a nomogram to predict LNM of CRC based on CT radiomics features; the nomogram showed good predictive performance in both training and validation cohorts. Li et al. proposed a clinical-radiomics nomogram with a combination of clinical risk factors and radiomics features for preoperative prediction of LNM in patients with CRC; the nomogram had moderate discrimination performance. Zhou et al. established a nomogram for LNM prediction in patients with rectal cancer based on clinical factors; the AUC of the nomogram was 0.743 in the training cohort and 0.777 in the validation cohort. However, it is not clear whether incorporating clinical risk factors and DESCT parameters in a nomogram would improve its predictive ability for LNM in patients with CRC. Therefore, the purpose of our study was to develop a clinical–DESCT nomogram that incorporated both clinical factors and DESCT parameters for individual preoperative prediction of the risk of LNM in CRC.

## MATERIALS AND METHODS

### Patients

This study was approved by the Institutional Review Board of Lanzhou University Second Hospital Medical Ethics Committee, and the requirement for informed consent was waived. We retrospectively evaluated all patients seen at our hospital between February 2015 and November 2019; eligible patients were those with pathologically confirmed CRC who underwent curative resection with LN dissection and who had received abdominal enhanced DESCT imaging before surgery. A flow diagram of the recruitment pathway, including inclusion and exclusion criteria, is shown in **Figure 1**. A total of 167 patients were identified and included in our study (85 with colon cancer; 82 with rectal cancer), and these patients were categorized into the training and validation cohorts. Clinical data and preoperative tumor serologic data were collected by reviewing the medical records of patients. Data collected were age, sex, tumor location, preoperative carbohydrate antigen 199 (CA19-9), preoperative carbohydrate antigen 125 (CA125), and preoperative carcinoembryonic antigen (CEA) levels. The status of LNM was evaluated by trained pathologists. The LN-positive group was characterized by the presence of one or more

**Abbreviations:** AP, arterial phase; CA125, carbohydrate antigen 125; CA19-9, carbohydrate antigen 199; CEA, carcinoembryonic antigen; CRC, colorectal cancer; CI, confidence interval; DCA, decision curve analysis; DESCT, dual-energy spectral CT; Eff-Z, effective atomic number; IC, iodine concentration; ICC, intraclass correlation coefficients; LNM, lymph node metastasis; PFI, pericolorectal fat invasion; ROC, receiver operating characteristic; ROI, region of interest; TGP, tumor gross pattern; VP, venous phase.
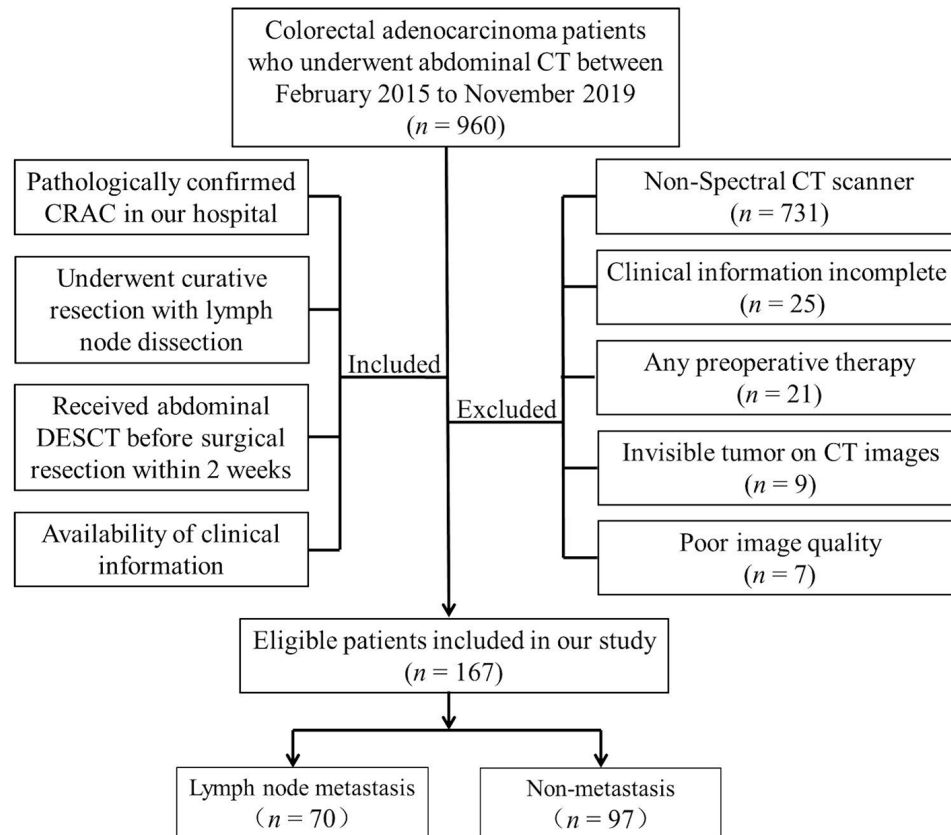
**FIGURE 1** | A flow diagram of patient recruitment, including inclusion and exclusion criteria.

metastatic LNs, while the LN-negative group was defined by normal healthy LNs.

## DESCT Imaging

All patients underwent bowel preparation before the examination. Contrast-enhanced abdominal dual-energy CT scans were performed using a Discovery CT750 HD system (GE Healthcare, Waukesha, WI, USA) in the supine position. The energy spectrum CT scanning protocol was as follows: fast tube voltage switching between 80 and 140 kVp; tube current, 350 mAs; rotation time, 0.75 s; pitch, 0.984:1; and reconstructed layer thickness, 1.25 mm. Patients were intravenously injected with iodixanol (1 ml/kg) at an injection rate of 3.5–4.5 ml/s using a high-pressure dual-cylinder injector. Arterial phase (AP) and venous phase (VP) imaging were performed 25–30 and 60–70 s after the administration of contrast agent, respectively.

## Image Postprocessing and Analysis

Raw CT imaging data were transferred to a GE ADW 4.6 workstation (GE Healthcare, Milwaukee, WI, USA). Gemstone Spectral Imaging (GSI) viewer software was used to quantitatively evaluate virtual monochrome images with a default of 70 keV, iodine-based decomposition images, and Eff-Z images. Two radiologists with more than 5 years of experience

in gastrointestinal radiology performed the image analysis. Both radiologists were blinded to the clinical and pathological data of the patients. The maximum tumor thickness was defined as the maximum diameter perpendicular to the long axis on the cross-sectional image. Pericolorectal fat invasion (PFI) was defined as the extension of the primary tumor beyond the muscularis propria and its invasion of the pericolorectal fat. Clinical tumor (cT) stage was evaluated according to the eighth edition of the American Joint Committee on Cancer Staging system (13). Two radiologists independently drew circular regions of interest (ROIs) at the maximum slice of the tumor with an average area of 97.27 mm$^2$ on 70 keV monochromatic images in the AP and VP. Previous studies have shown that 70 keV monochromatic images may provide an optimal trade-off between sensitivity and specificity for abdominal lesion analysis (14). ROIs were placed on solid areas to avoid vascular, necrotic, and cystic changes as much as possible. Tumor monochromatic CT values, iodine concentration (IC) values, and Eff-Z values were generated using the GSI viewer software package. To minimize bias, all measurements were performed three times and the average of the three values was taken as the final value. Examples of DESCT images with ROIs for evaluating quantitative measurements in two patients with CRC with and without LNM are shown in **Figures 2** and **3**, respectively.
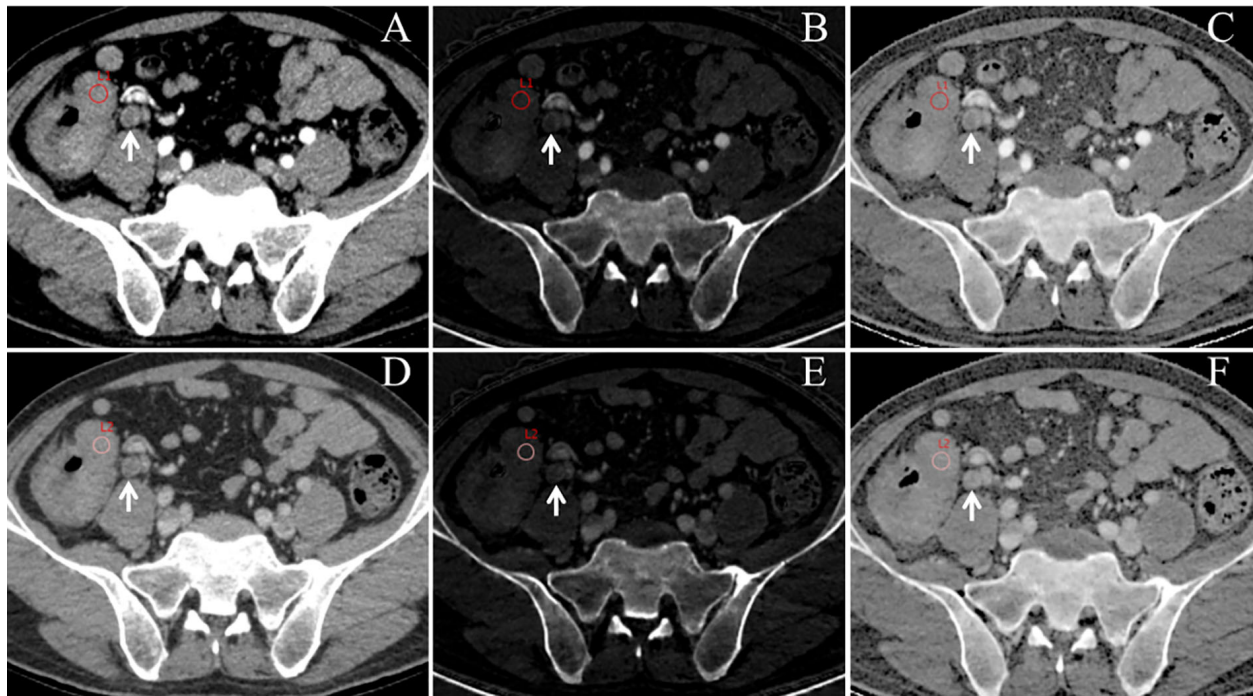
**FIGURE 2** | An example of dual-energy spectral computed tomography (DESCT) images with regions of interest (ROIs) for evaluating quantitative measurements in a 63-year-old man with ascending colon cancer that was pathologically confirmed to have lymph node metastasis (LNM). ROIs were placed in the arterial phase **(A)** and the venous phase **(D)** of the 70-keV monochromatic images. Concurrently, ROIs were copied to the arterial phase **(B)** and the venous phase **(E)** of iodine-based material decomposition images and the arterial phase **(C)** and the venous phase **(F)** of the effective atomic number images. Local lymphadenopathy is presented in front of the right psoas major (white arrow) at both the arterial phase and venous phase **(A–F)**.

## Statistical Analysis

All statistical analyses were performed using the R statistical software package (version 3.6.3; http://www.Rproject.org). Student's $t$-test or Mann–Whitney $U$ tests were used to compare continuous variables between the LN-positive and the LN-negative groups. Chi-square tests or Fisher's exact tests were used to compare categorical variables. A two-sided $p$-value <0.05 was considered statistically significant. The intraclass correlation coefficient (ICC) was used to calculate the consistency of measurements between the two radiologists. The statistically significant features in the univariate analysis were included in a multivariate logistic regression analysis. Backward stepwise selection was applied, in which the stopping rule was the likelihood ratio test with Akaike's information criterion. A multivariable logistic regression analysis was performed to select parameters with predictive significance for LNM. A quantitative and easy-to-use nomogram was built based on the final regression coefficient and designed to predict the individual probability of LNM. Receiver operating characteristic curve (ROC) analysis was used to evaluate the diagnostic capabilities of the nomogram, including calculation of the AUC value and 95% confidence interval (CI). The accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were also calculated. To verify the clinical usefulness of the nomogram, we quantified the net

benefit at different threshold probabilities in the data set using decision curve analysis (DCA).

## RESULTS

### Interobserver Agreement

Substantial interobserver agreement between the two radiologists was noted for all measurements. The ICCs for the 70-keV monochromatic CT values in the AP (CTAP), CT values in the VP (CTVP), IC values in the AP iodine concentration in the arterial phase (ICAP), IC values in the VP iodine concentration in the venous phase (ICVP), maximum tumor thickness, Eff-Z in the AP (Eff-ZAP), Eff-Z in the VP (Eff-ZVP), PFI, and cT stage were 0.912, 0.905, 0.928, 0.938, 0.922, 0.894, 0.887, 0.915, and 0.846, respectively.

### Associations Between Clinical Variables and LNM

A total of 167 patients with CRC were included in the final analysis. The LN-positive group comprised 70 patients, with an average age of $60.07 \pm 12.86$ years, and 54.3% of the patients were males. The LN-negative group comprised 97 patients, with an average age of $59.83 \pm 12.15$ years, and 60.8% of the patients were males. We used stratified sampling to categorize the study cohort
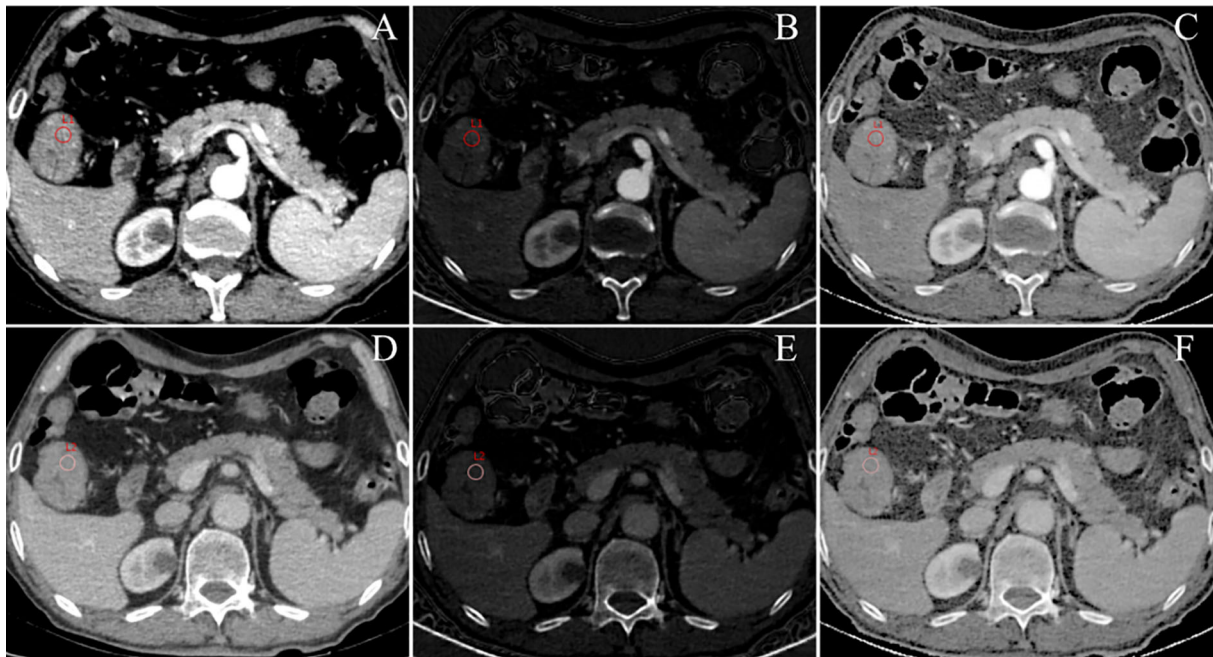
**FIGURE 3** | An example of DESCT images with ROIs for evaluating quantitative measurements in a 75-year-old man with ascending colon cancer that was pathologically confirmed to have non-metastatic lymph nodes. ROIs were placed in the arterial phase **(A)** and the venous phase **(D)** of the 70-keV monochromatic images. At the same time, ROIs were copied to the arterial phase **(B)** and the venous phase **(E)** of iodine-based material decomposition images and the arterial phase **(C)** and venous phase **(F)** of the effective atomic number images.

into a training cohort ($n = 117$) and a validation cohort ($n = 50$). The training cohort was used for model building, while the validation cohort was used for internal validation of the model. Patient and tumor characteristics in the training and validation cohorts are listed in **Table 1**.

In the training cohort, the LN-positive group showed higher CA19-9 level and PFI compared with the LN-negative group (all $p$-values <0.05; **Table 1**). The values for the DESCT parameters ICAP, ICVP, and Eff-ZVP were also significantly higher for the LN-positive group compared with those for the LN-negative group (all $p$-values <0.05; **Table 1**). There were no significant differences in other clinical factors and tumor DESCT parameters between the LN-positive group and the LN-negative group in the training cohort (all $p$-values >0.05).

## Prediction Model Analysis

Logistic regression analysis showed that CEA level, CA19-9 level, PFI, ICAP, ICVP, and Eff-ZVP were independent predictors in the predictive model. Based on these predictors, the following five predictive models were established to predict the individual probability of LNM: one clinical model, three DESCT models, and one combined clinical–DESCT model (**Table 2**). **Figure 4** shows the classification performance of the spectrum-AP model, spectrum-VP model, and spectrum-combined model. The AUC values of the spectrum-AP model, spectrum-VP model, and spectrum-combined model were 0.742 (95% CI, 0.652–0.831),

0.742 (95% CI, 0.648–0.835), and 0.786 (95% CI, 0.701–0.871), respectively, in the training cohort, and they were 0.763 (95% CI, 0.629–0.897), 0.695 (95% CI, 0.541–0.848), and 0.745 (95% CI, 0.600–0.891), respectively, in the validation cohort.

We developed a clinical–DESCT model that combined two clinical features: one morphological image feature and three DESCT parameters; the model was presented as a quantitative nomogram (**Figure 5A**). We found that the nomogram had higher prediction contributions than the clinicoradiological model or spectrum models (**Table 2** and **Figure 6**). The AUC, accuracy, sensitivity, specificity, PPV, and NPV were 0.876, 0.812, 0.755, 0.853, 0.787, and 0.829, respectively, in the training cohort, while they were 0.852, 0.760, 0.762, 0.759, 0.696, and 0.815, respectively, in the validation cohort. The DCA for the clinical–DESCT nomogram, clinicoradiological model, and spectrum-combined model is presented in **Figures 5B, C**. The clinical–DESCT nomogram was more clinically useful, i.e., it predicted the risk of LNM more accurately than the single clinicoradiological model and single spectrum-combined model in the training and validation cohorts. The DCA demonstrated that the nomogram had the highest clinical benefit when the threshold probability was 22.6%–58.3% in both training and validation cohorts. The best threshold obtained from ROC was 0.502, which falls within this interval, indicating that the nomogram had the highest performance and clinical benefit among other models in this study.

**TABLE 1** | Clinical characteristics and DESCT parameters of CRC patients [mean ± SD or no. (%)].

| Characteristics | Training cohort (*n* = 117) | | | Validation cohort (*n* = 50) | | |
|---|---|---|---|---|---|---|
| | LN metastasis (–) | LN metastasis (+) | *p*-value | LN metastasis (–) | LN metastasis (+) | *p*-value |
| Age (years) | 60.04 ± 12.09 | 60.41 ± 11.58 | 0.960 | 59.31 ± 14.47 | 59.29 ± 15.75 | 0.922 |
| Gender | | | 0.794 | | | 0.055 |
| Female | 28 (41.2) | 19 (38.8) | | 10 (34.5) | 13 (61.9) | |
| Male | 40 (58.8) | 30 (61.2) | | 19 (65.5) | 8 (38.1) | |
| Tumor location | | | 0.791 | | | 0.416 |
| Left | 46 (67.6) | 32 (65.3) | | 19 (65.5) | 16 (76.2) | |
| Right | 22 (32.4) | 17 (34.7) | | 10 (34.5) | 5 (23.8) | |
| CEA level | | | 0.055 | | | <0.001 |
| Normal | 44 (64.7) | 23 (46.9) | | 20 (69.0) | 4 (19.0) | |
| Abnormal | 24 (35.3) | 26 (53.1) | | 9 (31.0) | 17 (81.0) | |
| CA125 level | | | 0.387 | | | 0.647 |
| Normal | 64 (94.1) | 44 (89.8) | | 25 (86.2) | 19 (90.5) | |
| Abnormal | 4 (5.9) | 5 (10.2) | | 4 (13.8) | 2 (9.5) | |
| CA19-9 level | | | <0.001 | | | 0.021 |
| Normal | 62 (91.2) | 32 (65.3) | | 24 (82.8) | 11 (52.4) | |
| Abnormal | 6 (8.8) | 17 (34.7) | | 5 (17.2) | 10 (47.6) | |
| Maximum diameter (cm) | 20.75 ± 10.37 | 20.36 ± 7.09 | 0.564 | 21.69 ± 11.52 | 22.93 ± 9.51 | 0.438 |
| cT stage | | | 0.271 | | | 0.184 |
| T1–2 | 20 (29.4) | 10 (20.4) | | 7 (24.1) | 2 (9.5) | |
| T3–4 | 48 (70.6) | 39 (79.6) | | 22 (75.9) | 19 (90.5) | |
| Gross tumor pattern | | | 0.201 | | | 0.045 |
| Non-polypoid | 60 (88.2) | 39 (79.6) | | 26 (89.7) | 14 (66.7) | |
| Polypoid | 8 (11.8) | 10 (20.4) | | 3 (10.3) | 7 (33.3) | |
| Pericolorectal fat invasion | | | <0.001 | | | 0.007 |
| No | 40 (58.8) | 7 (14.3) | | 13 (44.8) | 2 (9.5) | |
| Yes | 28 (41.2) | 42 (85.7) | | 16 (55.2) | 19 (90.5) | |
| ICAP | 17.52 ± 4.53 | 20.65 ± 3.19 | <0.001 | 17.14 ± 4.28 | 20.43 ± 3.75 | 0.002 |
| ICVP | 15.15 ± 2.25 | 17.36 ± 2.77 | <0.001 | 15.35 ± 2.50 | 17.14 ± 2.85 | 0.024 |
| Eff_ZAP | 8.67 ± 0.30 | 8.79 ± 0.36 | 0.107 | 8.64 ± 0.28 | 8.82 ± 0.34 | 0.101 |
| Eff_ZVP | 8.54 ± 0.33 | 8.71 ± 0.30 | 0.004 | 8.54 ± 0.36 | 8.64 ± 0.38 | 0.132 |
| CTAP (HU) | 81.40 ± 10.63 | 80.98 ± 12.84 | 0.600 | 81.98 ± 11.87 | 80.67 ± 11.69 | 0.534 |
| CTVP (HU) | 74.00 ± 9.23 | 73.86 ± 9.48 | 0.494 | 74.39 ± 8.07 | 77.81 ± 9.94 | 0.400 |

*LN, lymph node.*

## DISCUSSION

To the best of our knowledge, this is the first study to construct a clinical–DESCT model that combined clinical risk factors and DESCT features of primary lesions for preoperative prediction of LNM in patients with CRC. Herein, we first screened the preoperatively available risk factors for independent predictors using multivariate logistic regression analysis. Then, we incorporated these clinical and DESCT risk factors into an easy-to-use nomogram to facilitate their use in clinical practice. The nomogram showed higher predictive accuracy than other clinicoradiological and spectrum models for LNM both in the training cohort and validation cohort. The DCA showed that the actual benefits of the model were excellent. The clinical application of this nomogram facilitates the individual preoperative prediction of LNM and thus helps develop more reasonable and effective therapeutic strategies.

Several previous studies identified metastatic and non-metastatic LNs in CRC using DESCT. Liu et al. (11) used energy spectrum CT to identify metastatic and non-metastatic LNs in patients with rectal cancer; they found that when combining NIC in the VP with the short-axis diameter, the overall AUC was 0.819. Al-Najami et al. (8) evaluated the value of dual-energy CT in identifying metastatic LNs in rectal cancer;

several DESCT parameters showed moderate diagnostic accuracy for LNM. However, these studies only focused on the LNs, ignoring the features of the primary tumor. Histopathological features of primary CRCs, such as tumor differentiation and lymphatic invasion, are crucial for the development of LNM (15). In the present study, we extracted the DESCT parameters from the primary tumor and found that there were significant differences in the DESCT parameters between the LN-positive and LN-negative groups in both the arterial and venous phases. DESCT parameters of arterial and venous phases showed moderate predictive accuracy for LNM.

In our study, we found that preoperative CEA level, CA19-9 level, PFI, ICAP, ICVP, and Eff-ZVP were independent risk factors for LNM. In terms of clinical features, high CEA and CA19-9 levels are an important risk factor for LNM in patients with CRC, as reported in previous studies (4, 6). Elevated CEA and CA19-9 levels indicate increased tumor aggressiveness and metastasis (6). In addition, the LN-positive group displayed a higher incidence of PFI than did the LN-negative group in our study. This finding shows that CRC with LNM has a more aggressive behavior, involving LNs as well as peripheral fat around the primary tumor. Preoperative PFI status was a qualitative feature that could be easily obtained *via* CT imaging. Our study found that PFI status was an independent

TABLE 2 | Predictive performance of different models in training and validation cohorts.

| Models | Training cohort | | | | | | Validation cohort | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | AUC | Accuracy | Sensitivity | Specificity | PPV | NPV | AUC | Accuracy | Sensitivity | Specificity | PPV | NPV |
| Clinicoradiological | 0.769 (0.689–0.848) | 0.701 (0.609–0.782) | 0.857 (0.682–0.943) | 0.588 (0.308–0.698) | 0.600 (0.544–0.623) | 0.851 (0.750–0.871) | 0.727 (0.598–0.857) | 0.640 (0.492–0.771) | 0.905 (0.692–1.000) | 0.448 (0.252–0.640) | 0.543 (0.476–0.568) | 0.867 (0.785–0.903) |
| Spectrum-AP | 0.742 (0.652–0.831) | 0.709 (0.618–0.790) | 0.898 (0.735–0.959) | 0.574 (0.326–0.706) | 0.603 (0.554–0.618) | 0.886 (0.816–0.906) | 0.763 (0.629–0.897) | 0.740 (0.597–0.854) | 0.905 (0.429–1.000) | 0.621 (0.310–0.793) | 0.633 (0.450–0.656) | 0.900 (0.818–0.920) |
| Spectrum-VP | 0.742 (0.648–0.835) | 0.735 (0.645–0.812) | 0.612 (0.347–0.735) | 0.824 (0.574–0.912) | 0.714 (0.586–0.750) | 0.747 (0.672–0.765) | 0.695 (0.541–0.848) | 0.600 (0.452–0.736) | 0.571 (0.381–0.905) | 0.621 (0.447–0.931) | 0.522 (0.421–0.633) | 0.667 (0.590–0.750) |
| Spectrum-combined | 0.786 (0.701–0.871) | 0.769 (0.682–0.842) | 0.673 (0.306–0.796) | 0.838 (0.603–0.927) | 0.750 (0.577–0.780) | 0.781 (0.719–0.798) | 0.745 (0.600–0.891) | 0.680 (0.533–0.805) | 0.571 (0.381–0.811) | 0.759 (0.483–1.000) | 0.632 (0.533–0.709) | 0.710 (0.609–0.763) |
| Nomogram | 0.876 (0.815–0.936) | 0.812 (0.729–0.878) | 0.755 (0.489–0.878) | 0.853 (0.691–0.956) | 0.787 (0.706–0.811) | 0.829 (0.797–0.844) | 0.852 (0.748–0.956) | 0.760 (0.618–0.869) | 0.762 (0.429–1.000) | 0.759 (0.552–0.931) | 0.696 (0.562–0.750) | 0.815 (0.762–0.844) |

*Clinicoradiological, fusion of clinical risks and radiological features; Spectrum-combined, fusion of spectrum-AP and spectrum-VP; Nomogram, fusion of clinical risks, radiological features, and spectrum parameters.*
*AUC, area under the curve; CI, confidence interval; PPV, positive predictive value; NPV, negative predictive value; AP, arterial phase, VP venous phase.*



FIGURE 4 | Comparison of the different spectrum models for the identification of LNM in patients with colorectal cancer in the training cohort (A) and validation cohort (B).
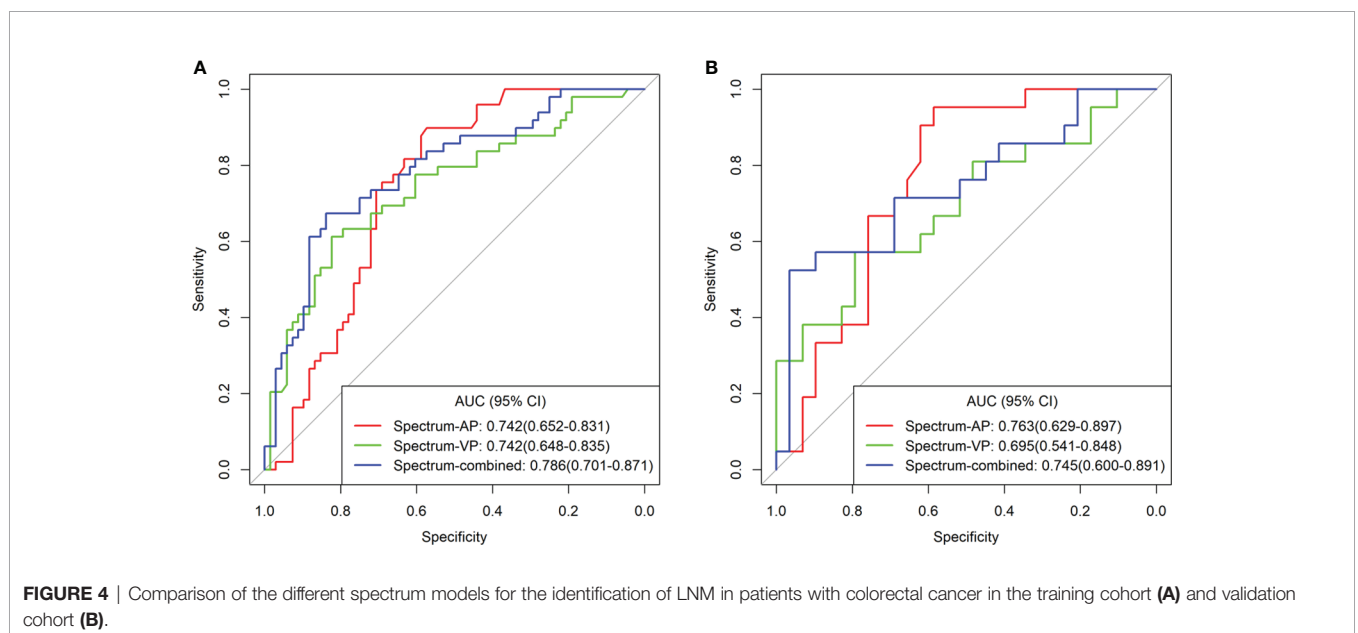
risk factor for LNM. Surprisingly, we found that DESCT parameters such as ICAP, ICVP, and Eff-ZVP were also independent predictors for LNM. These findings suggest that quantitative imaging provides significant variables for the construction of predictive nomograms. Therefore, we combined both clinical risk factors and DESCT parameters into the nomogram for preoperative prediction of LNM in patients with CRC. We found that this nomogram had higher predictive AUC and greater net benefits than single clinicoradiological and single spectrum-combined models. Hence, the clinical–DESCT combined model may be the most promising approach to predict LNM in patients with CRC.

In this study, in terms of DESCT features, ICAP and ICVP were significantly higher in the LN-positive group than in the

LN-negative group, and both ICAP and ICVP were independent predictors for LNM in the multivariate analysis. ICs can evaluate the degree of tumor vascularization as this measure can quantitatively reflect the deposition of iodine in the tissue (10). IC in the AP reflects the functional capillary density, while IC in the VP reflects iodine equilibrium in the blood vessels (16). Therefore, the ICs in the AP and the VP reflect the dynamic distribution of iodine in the tumor tissue. Tumor angiogenesis is closely related to tumor growth, progression, and metastasis (17, 18). Heterogeneity in tumor angiogenesis leads to differences in the biological behavior of tumors, such as LNM (19). High angiogenesis intensity is closely related to aggressive histopathological features, such as LNM, in CRC (20–22). In the present study, we found that the LN-positive group had
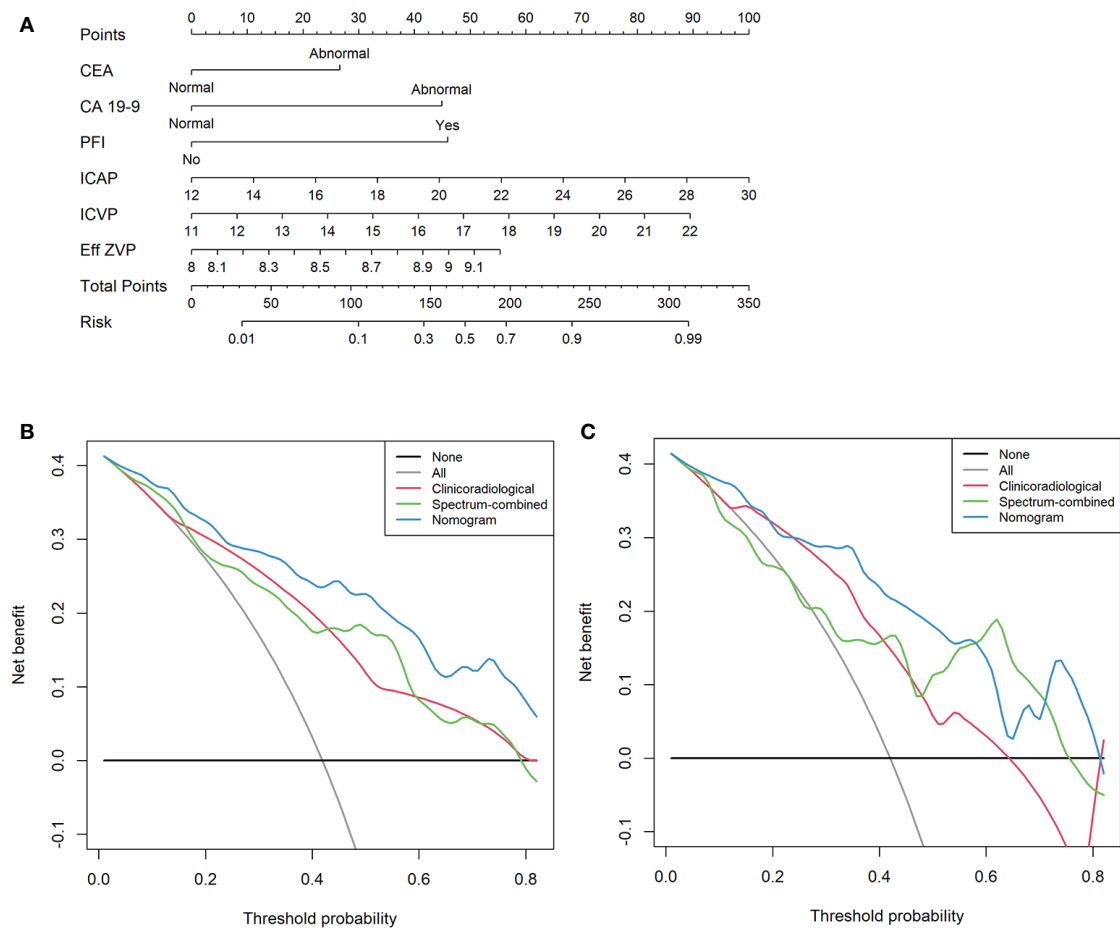
FIGURE 5 | A multiparametric clinical–DESCT nomogram for predicting the probability of LNM in CRC patients (A). Decision curve analysis (DCA) of the clinical–DESCT nomogram, clinicoradiological model, and spectrum-combined model in the training cohort (B) and validation cohort (C).
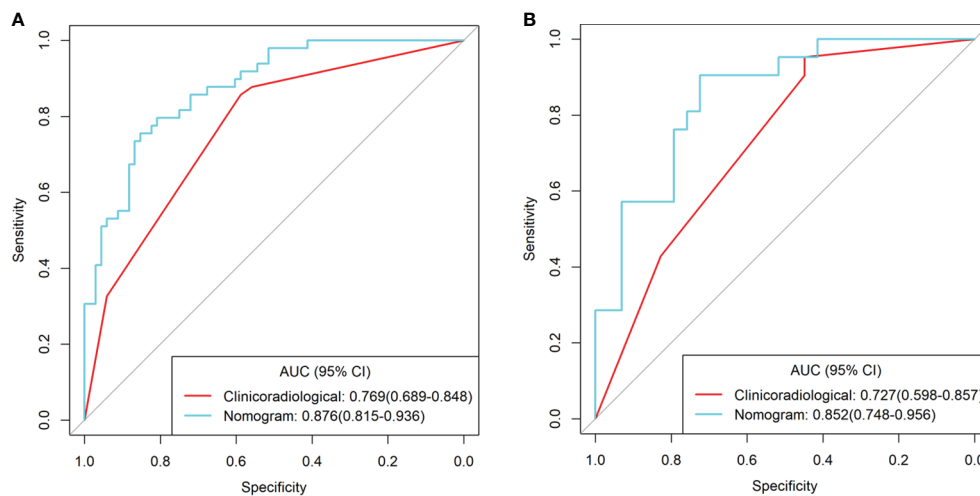


FIGURE 6 | ROC curves of the clinical–DESCT nomogram, clinicoradiological model, and spectrum-combined model for preoperative prediction of LNM in patients with colorectal cancer in the training cohort (A) and validation cohort (B).

significantly higher IC values than the LN-negative group, suggesting that the primary tumors of patients with LNM had a greater blood supply.

The Eff-Z is an indication of tissue density, i.e., the higher the density, the higher the Eff-Z (23). In the present study, we found that Eff-ZVP was statistically higher in the LN-positive group than in the LN-negative group in the training cohort. The higher Eff-Z values observed in the LN-positive group in our study might indicate a relatively compact cell structure within the tumor; this finding is consistent with previous reports (24). In a recent investigation of ADC values in early and advanced colon tumors, a significantly lower mean ADC value was observed in advanced compared with early tumors, with an optimal discrimination cutoff value of $1.179 \times 10^{-3}$ mm$^2$/s. The ADC value has a negative correlation with the tissue cell density, i.e., the higher the density of the tissue, the lower the ADC value.

In the training cohort, we found that the predictive accuracy of LNM in the spectrum-AP model was similar to that in the spectrum-VP model and increased to 0.769 when the two phases were combined. However, double DESCT scans increase the radiation dose more than single DESCT scans. Further research aimed at developing a technique to reduce the radiation dose is necessary; for example, the application of multimodel iterative reconstruction technology can reduce the radiation dose effectively and improve the image quality (25).

This study has several limitations. First, owing to the retrospective nature of the study, there was an inevitable selection bias that should be addressed in future prospective and external validation studies. Second, the number of patients in this study was small, and the results should be further validated with larger samples. Finally, we did not perform histopathologic–radiologic, one-to-one matching of LNs, as this study focused on the primary tumor and the peripheral LNs were not evaluated using DESCT.

In conclusion, we constructed a clinical–DESCT nomogram that incorporates both clinical risk factors and DESCT parameters, presenting a non-invasive and highly useful predictive tool for individual preoperative prediction of LNM in patients with CRC.

## DATA AVAILABILITY STATEMENT

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

## ETHICS STATEMENT

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

## AUTHOR CONTRIBUTIONS

Conception and design: HB, JuZ, and YuC. Collection and assembly of data: YuC and XY. Development of the methodology: JR. All authors contributed to the article and approved the submitted version.

## FUNDING

## REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer Statistics, 2020. *CA Cancer J Clin* (2020) 70(1):7–30. doi: 10.3322/caac.21590
2. Chen LD, Liang JY, Wu H, Wang Z, Li SR, Li W, et al. Multiparametric Radiomics Improve Prediction of Lymph Node Metastasis of Rectal Cancer Compared With Conventional Radiomics. *Life Sci* (2018) 208:55–63. doi: 10.1016/j.lfs.2018.07.007
3. Smith JJ, Garciaaguilar J. Advances and Challenges in Treatment of Locally Advanced Rectal Cancer. *J Clin Oncol* (2015) 33:1797–808. doi: 10.1200/JCO.2014.60.1054
4. Huang YQ, Liang CH, He L, Tian J, Liang CS, Chen X, et al. Development and Validation of a Radiomics Nomogram for Preoperative Prediction of Lymph Node Metastasis in Colorectal Cancer. *J Clin Oncol* (2016) 34:2157–64. doi: 10.1200/JCO.2015.65.9128
5. Yang L, Liu D, Fang X, Wang ZQ, Xing Y, Ma L, et al. Rectal Cancer: Can T2WI Histogram of the Primary Tumor Help Predict the Existence of Lymph Node Metastasis? *Eur Radiol* (2019) 29:6469–76. doi: 10.1007/s00330-019-06328-z
6. Li M, Zhang J, Dan Y, Yao Y, Dai W, Cai G, et al. A Clinical-Radiomics Nomogram for the Preoperative Prediction of Lymph Node Metastasis in Colorectal Cancer. *J Transl Med* (2020) 18:46. doi: 10.1186/s12967-020-02215-0
7. Gong HX, Zhang KB, Wu LM, Baigorri BF, Yin Y, Geng XC, et al. Dual Energy Spectral CT Imaging for Colorectal Cancer Grading: A Preliminary Study. *PLoS One* (2016) 11:e0147756. doi: 10.1371/journal.pone.0147756
8. Al-Najami I, Lahaye MJ, Beets-Tan RGH, Baatrup G. Dual-Energy CT can Detect Malignant Lymph Nodes in Rectal Cancer. *Eur J Radiol* (2017) 90:81–8. doi: 10.1016/j.ejrad.2017.02.005
9. Al-Najami I, Drue HC, Steele R, Baatrup G. Dual Energy CT – A Possible New Method to Assess Regression of Rectal Cancers After Neoadjuvant Treatment. *J Surg Oncol* (2017) 116:984–8. doi: 10.1002/jso.24761
10. Wu J, Lv Y, Wang N, Zhao Y, Zhang P, Liu Y, et al. The Value of Single-Source Dual-Energy CT Imaging for Discriminating Microsatellite Instability From Microsatellite Stability Human Colorectal Cancer. *Eur Radiol* (2019) 29:3782–90. doi: 10.1007/s00330-019-06144-5
11. Liu H, Yan F, Pan Z, Lin X, Luo X, Shi C, et al. Evaluation of Dual Energy Spectral CT in Differentiating Metastatic From Non-Metastatic Lymph Nodes in Rectal Cancer: Initial Experience. *Eur J Radiol* (2015) 84:228–34. doi: 10.1016/j.ejrad.2014.11.016
12. Zhou C, Liu HS, Liu XH, Zheng XB, Hu T, Liang ZX, et al. Preoperative Assessment of Lymph Node Metastasis in Clinically Node-Negative Rectal

Cancer Patients Based on a Nomogram Consisting of Five Clinical Factors. *Ann Transl Med* (2019) 7:543. doi: 10.21037/atm.2019.09.127

13. Amin MB, Greene FL, Edge SB, Compton CC, Gershenwald JE, Brookland RK, et al. The Eighth Edition AJCC Cancer Staging Manual: Continuing to Build a Bridge From a Population-Based to a More "Personalized" Approach to Cancer Staging. *CA Cancer J Clin* (2017) 67:93–9. doi: 10.3322/caac.21388

14. Schabel C, Patel B, Harring S, Duvnjak P, Ramírez-Giraldo JC, Nikolaou K, et al. Renal Lesion Characterization With Spectral CT: Determining the Optimal Energy for Virtual Monoenergetic Reconstruction. *Radiol* (2018) 287:874–83. doi: 10.1148/radiol.2018171657

15. Glasgow SC, Bleier JI, Burgart LJ, Finne CO, Lowry AC. Meta-Analysis of Histopathological Features of Primary Colorectal Cancers That Predict Lymph Node Metastases. *J Gastrointest Surg* (2012) 16:1019–28. doi: 10.1007/s11605-012-1827-4

16. Jing L, Mengjie F, Rui W, Dong D, Tian J, Liang P, et al. Diagnostic Accuracy of Dual-Energy CT-Based Nomograms to Predict Lymph Node Metastasis in Gastric Cancer. *Eur Radiol* (2018) 28:5241–9. doi: 10.1007/s00330-018-5483-2

17. Hu S, Huang W, Chen Y, Song Q, Lin X, Wang Z, et al. Spectral CT Evaluation of Interstitial Brachytherapy in Pancreatic Carcinoma Xenografts: Preliminary Animal Experience. *Eur Radiol* (2014) 24:2167–73. doi: 10.1007/s00330-014-3257-z

18. Thaiss WM, Haberland U, Kaufmann S, Spira D, Thomas C, Nikolaou K, et al. Iodine Concentration as a Perfusion Surrogate Marker in Oncology: Further Elucidation of the Underlying Mechanisms Using Volume Perfusion CT With 80 kVp. *Eur Radiol* (2016) 26:2929–36. doi: 10.1007/s00330-015-4154-9

19. Chen XH, Ren K, Liang P, Chai YR, Chen KS, Gao JB, et al. Spectral Computed Tomography in Advanced Gastric Cancer: Can Iodine Concentration non-Invasively Assess Angiogenesis? *World J Gastroenterol* (2017) 23:1666–75. doi: 10.3748/wjg.v23.i9.1666

20. Gurzu S, Jung J, Azamfirei L, Mezei T, Cîmpean AM, Szentirmay Z, et al. The Angiogenesis in Colorectal Carcinomas With and Without Lymph Node Metastases. *Rom J Morphol Embryol* (2008) 49:149–52.

21. Nakasaki T, Wada H, Shigemori C, Miki C, Gabazza EC, Nobori T, et al. Expression of Tissue Factor and Vascular Endothelial Growth Factor is Associated With Angiogenesis in Colorectal Cancer. *Am J Hematol* (2002) 69:247–54. doi: 10.1002/ajh.10061

22. Chan E. Angiogenesis in Colorectal Cancer: Antibodies. *Cancer J* (2016) 22:179–81. doi: 10.1097/PPO.0000000000000200

23. Mileto A, Allen BC, Pietryga JA, Farjat AE, Zarzour JG, Bellini D, et al. Characterization of Incidental Renal Mass With Dual-Energy CT: Diagnostic Accuracy of Effective Atomic Number Maps for Discriminating Nonenhancing Cysts From Enhancing Masses. *AJR Am J Roentgenol* (2017) 209:W221–30. doi: 10.2214/AJR.16.17325

24. Nerad E, Delli Pizzi A, Lambregts DMJ, Maas M, Wadhwani S, Bakers FCH, et al. The Apparent Diffusion Coefficient (ADC) Is a Useful Biomarker in Predicting Metastatic Colon Cancer Using the ADC-Value of the Primary Tumor. *PLoS One* (2019) 14:e0211830. doi: 10.1371/journal.pone.0211830

25. Jia Y, Zhai B, He T, Yu Y, Yu N, Duan H, et al. The Application of a New Model-Based Iterative Reconstruction in Low-Dose Upper Abdominal CT. *Acad Radiol* (2019) 26(10):e275–83. doi: 10.1016/j.acra.2018.11.020

# Predictive Value of Multiparametric MRI for Response to Single-Cycle Induction Chemo-Immunotherapy in Locally Advanced Head and Neck Squamous Cell Carcinoma

Konstantin Hellwig[1]*, Stephan Ellmann[1], Markus Eckstein[2], Marco Wiesmueller[1], Sandra Rutzner[3,4], Sabine Semrau[3,4], Benjamin Frey[3,4], Udo S. Gaipl[3,4], Antoniu Oreste Gostian[4,5], Arndt Hartmann[2,4], Heinrich Iro[4,5], Rainer Fietkau[3,4], Michael Uder[1,4], Markus Hecht[3,4†] and Tobias Bäuerle[1,4†]

[1] Institute of Radiology, University Hospital Erlangen, Erlangen, Germany, [2] Institute of Pathology, University Hospital Erlangen, Erlangen, Germany, [3] Department of Radiation Oncology, University Hospital Erlangen, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany, [4] Comprehensive Cancer Center Erlangen-European Metropolitan Region of Nuremberg (CCC ER-EMN), Erlangen, Germany, [5] Department of Otolaryngology - Head & Neck Surgery, University Hospital Erlangen, Erlangen, Germany

**Objectives:** To assess the predictive value of multiparametric MRI for treatment response evaluation of induction chemo-immunotherapy in locally advanced head and neck squamous cell carcinoma.

**Methods:** Twenty-two patients with locally advanced, histologically confirmed head and neck squamous cell carcinoma who were enrolled in the prospective multicenter phase II CheckRad-CD8 trial were included in the current analysis. In this unplanned secondary single-center analysis, all patients who received contrast-enhanced MRI at baseline and in week 4 after single-cycle induction therapy with cisplatin/docetaxel combined with the immune checkpoint inhibitors tremelimumab and durvalumab were included. In week 4, endoscopy with representative re-biopsy was performed to assess tumor response. All lesions were segmented in the baseline and restaging multiparametric MRI, including the primary tumor and lymph node metastases. The volume of interest of the respective lesions was volumetrically measured, and time-resolved mean intensities of the golden-angle radial sparse parallel-volume-interpolated gradient-echo perfusion (GRASP-VIBE) sequence were extracted. Additional quantitative parameters including the T1 ratio, short-TI inversion recovery ratio, apparent diffusion coefficient, and dynamic contrast-enhanced (DCE) values were measured. A model based on parallel random forests incorporating the MRI parameters from the baseline MRI was used to predict tumor response to therapy. Receiver operating characteristic (ROC) curves were used to evaluate the prognostic performance.

**Results:** Fifteen patients (68.2%) showed pathologic complete response in the re-biopsy, while seven patients had a residual tumor (31.8%). In all patients, the MRI-based primary tumor volume was significantly lower after treatment. The baseline DCE parameters of time to peak and wash-out were significantly different between the pathologic complete response group and the residual tumor group ($p < 0.05$). The developed model, based on parallel random forests and DCE parameters, was able to predict therapy response with a sensitivity of 78.7% (95% CI 71.24–84.93) and a specificity of 78.6% (95% CI 67.13–87.48). The model had an area under the ROC curve of 0.866 (95% CI 0.819–0.914).

**Conclusions:** DCE parameters indicated treatment response at follow-up, and a random forest machine learning algorithm based on DCE parameters was able to predict treatment response to induction chemo-immunotherapy.

Keywords: magnetic resonance imaging, dynamic contrast-enhanced imaging, DCE-MRI, head and neck cancer, immunotherapy, multiparametric MRI

# INTRODUCTION

Cancer immunotherapy is an emerging and highly promising therapeutic approach in oncology. Immune checkpoint inhibitors (ICIs), such as inhibitors targeting PD-1, PD-L1, and CTLA-4, are approved for the treatment of different cancer types such as melanoma, lung cancer, and head and neck squamous cell carcinoma (1–8). Although effective, the overall response rates for ICIs are approximately 15 to 20% for advanced melanoma and non-small-cell lung cancer, while in other tumor types the clinical response varies even more—from approximately 10% to greater than 50% of patients (9–11). Therefore, patient selection is crucial. An accurate prediction of response prior to potential ICI treatment is desirable, but this is still not established. Predictive biomarkers such as tumor PD-L1 expression, microsatellite instability status, and tumor mutational burden show an association with clinical response among different cancer types (12–17). These predictive markers require invasive tumor biopsy and the results may not be representative because of tumor heterogeneity (18). Attempts to correlate genetic heterogeneity in biopsy samples with FDG-SUV or ADC values in PET/MRI were not successful in a small cohort of patients with head and neck cancer (19). More recent approaches focus on the predictive value of peripheral blood immune cells (20).

Another challenge with ICIs is therapeutic evaluation and follow-up. Immunotherapy can cause hyperprogression, an acceleration of tumor growth after treatment, and pseudoprogression, an initial increase in tumor size followed by morphological regression; furthermore, a mixed response with shrinkage and growth of lesions has been reported (21–23). Initial morphological imaging alone may be misleading; therefore, revised Response Evaluation Criteria in Solid Tumors (RECIST) such as iRECIST and imRECIST have been established (23, 24). These criteria facilitate and standardize follow-up, but in many cases uncertainty remains. Functional imaging, e.g. ^{18}F-fluorodeoxyglucose (^{18}F-FDG) PET/CT, is commonly used to evaluate the tumor response after therapy using the standardized uptake value (SUV) and total lesion glycolysis (TLG). Response evaluation is performed by measuring the SUV normalized by lean body mass (SUL); an increase in SUL peak > 30% or the appearance of a new lesions is considered progressive disease (25). Although ^{18}F-FDG PET/CT can provide additional information after treatment with chemotherapeutics, ^{18}F-FDG accumulates in inflamed tissue and can lead to false-positive results due to pseudoprogression (26).

Dynamic contrast-enhanced (DCE) MRI is an important diagnostic tool for several tumor types, e.g. prostate and breast cancer. DCE MRI in breast cancer helps to screen high-risk patients and is used to monitor the response to therapy and detect carcinoma *in situ* (27, 28). In multiparametric MRI, the apparent diffusion coefficient (ADC), tumor volume and the difference of the DCE parameter Ktrans before and after chemoradiation are correlated with the pathologic response post chemotherapy for rectal cancer (29, 30). For head and neck cancer, DCE MRI facilitates the diagnosis of cervical lymph node metastases (31). Given the cost of treatment with ICIs, the fact that only a subset of patients responds to this therapy, and the difficulties in response evaluation, a noninvasive predictor of therapy response is essential for personalized cancer immunotherapy. Most of the studies evaluating radiomic biomarkers in head and neck squamous cell carcinoma (HNSCC) focus on CT; few used MRI or PET/CT, and a validation cohort was only used in slightly more than half of the studies (32).

This study aimed to create a predictive model for the response assessment of induction chemo-immunotherapy using multiparametric MRI in advanced HNSCC with biological validation through biopsy.

# MATERIALS AND METHODS

## Study Design, Data Source, and Image Acquisition

In this unplanned secondary single-center analysis of the CheckRad-CD8 trial, the predictive value of multiparametric

MRI was studied. Patients with locally advanced, histologically confirmed HNSCC stage III–IVb (according to TNM classification of malignant tumors 8th edition) of the oral cavity, oropharynx, hypopharynx, or supraglottic larynx received contrast-enhanced MRI at baseline and in week 4 after induction therapy with cisplatin [30 mg/m² body surface area (BSA)] on days 1–3 and docetaxel (75 mg/m² BSA) on day 1. The anti-CTLA4 ICI tremelimumab (75 mg fixed dose) and the anti-PDL1 durvalumab (1500 mg fixed dose) were administered on day 5. In week 4, endoscopy with representative re-biopsy was performed to assess pathologic response and the density of intratumoral CD8+ cells. In case no residual tumor was clinically detected, biopsies were taken from the primary tumor region. Patients with biopsies with no remaining tumor in a sufficiently covered tumor bed in the re-biopsy were scored as pathologic complete response (pCR). The results from the first interim analysis of the induction period of the CheckRad-CD8 trial on safety and efficacy were recently reported by Hecht et al. (33).

All MRI examinations were performed on a 3 Tesla MRI following the institutional reference protocol (Magnetom Vida; Siemens Healthineers, Erlangen, Germany).

- pre-contrast short-tau inversion recovery (STIR) T2-weighted (T2w) slice thickness = 4 mm, coronal
- pre-contrast T2w STIR, slice thickness = 3 mm, transversal
- pre-contrast T1-weighted (T1w) turbo spin echo (TSE), slice thickness = 3 mm, transversal
- diffusion-weighted imaging with apparent diffusion coefficient (ADC) map, slice thickness = 5 mm, transversal
- gadobutrol (Gadovist, Bayer HealthCare Pharmaceuticals) was administered to the patients (0.1 mmol/kg), followed by a 30-mL saline flush *via* a power injector at a rate of 1 mL/s
- Golden-angle radial sparse parallel (GRASP) technique (Siemens Healthineers) was applied to a transversal T1-weighted volume-interpolated gradient-echo perfusion sequence (VIBE), started 8 seconds before contrast injection, obtained for each patient with the minimum temporal offset available (2.5 s) over 338 s of total acquisition time (34).
- post-contrast T1w TSE with fat saturation (fs), slice thickness = 3 mm, transversal
- post-contrast T1w TSE fs-dixon, slice thickness = 3 mm, coronal

A total of 22 patients from October 2019 to October 2020 were included. In one case, DCE imaging could not be acquired at follow-up. All patients were male, and the median age was 61 years (interquartile range (IQR) 54–67.5 years).

## Trial Oversight

The CheckRad-CD8 trial is registered at ClinicalTrials.gov (identifier: NCT03426657). The leading institutional review board at the Friedrich-Alexander-Universität Erlangen-Nürnberg (number: 131_18 Az) and all local ethic committees approved the CheckRad-CD8 trial. All patients gave written

informed consent to all trial procedures, data protection measures, and the scientific use of imaging data.

## Definition of Pathological Complete Response

The results were correlated with either residual tumor or pathological complete response (pCR) after induction therapy before the initiation of radiation. pCR was defined as the complete absence of vital tumor cells in restaging biopsies. Biopsies were considered to represent the former tumor bed if significant regressive changes, i.e., fibrosis, bleedings, prominent chronic and active inflammation, were present.

## Image Analysis and Evaluation of Prognostic Relevance

In both baseline and follow-up multiparametric MRI, all lesions were manually segmented including the primary tumor (tumor volume) and lymph node metastases (lymph node volume) by a radiologist (KH, who was blinded to the clinical outcome) using Annotation Client (Chimaera GmbH, Erlangen, Germany) supervised by another radiologist with more than 15 years of experience in oncologic imaging (TB). The sequence used for delineation was the T1 weighted GRASP-VIBE sequence. The whole tumor volume was defined as the combined volume of the primary tumor and the lymph node metastases. The volume of the respective lesions was measured, and time-resolved mean intensities of the GRASP-VIBE sequence were extracted.

The GRASP-VIBE DCE values were analyzed with a custom-built R script. Using this script, the raw signal intensities of the first five DCE measurements were averaged and defined as a baseline with a relative enhancement set to 0. Bolus arrival was defined as the first measurement exceeding the raw signal intensity of the baseline by 5% and was set to 0 seconds. All measurements prior to the bolus arrival were omitted, and all subsequent measurements were normalized in terms of time (seconds since bolus arrival) and enhancement (relative to baseline enhancement).

The resulting data points were then fitted to a modified Brix equation (35):

$$\text{Enhancement}_{rel} = A * k_{ep} * \frac{e^{(-k_{ep} * x)} - e^{(-k_{el} * x)}}{k_{el} - k_{ep}},$$

where the relative enhancement and x (as the time in seconds) are known, while A, $k_{ep}$, and $k_{el}$ are to be determined. For this purpose, A, $k_{ep}$, and $k_{el}$ were iteratively approximated using the Levenberg–Marquardt least-squares minimization method until the algorithm converged at an optimal fit (36).

From the fitted curve, peak enhancement (PE) was determined as the maximum relative enhancement, with time to peak (TTP) defined as the corresponding x value in seconds. The maximum and minimum of the fitted curve's first derivative were respectively defined as the wash-in and wash-out.

Further, the imaging features of T1 ratio (signal intensity of the lesion as compared to autochthonous back muscle intensity), STIR ratio (signal intensity of the lesion as compared to autochthonous back muscle intensity), and the ADC value were assessed using syngo.*via* (Siemens Healthineers) in a representative region of interest without cystic changes or cavitation. The response was evaluated using the pathohistological results of the week 4 biopsy samples. No residual tumor in the biopsy was considered as a complete response. DCE maps were created using the MR Tissue4D Analysis tool in syngo.*via*.

## Statistical Analysis

Statistical analyses were performed with IBM SPSS Statistics for Windows, version 24 (IBM Corp., Armonk, NY, USA). Nonparametric testing was performed with Kruskal-Wallis test and the Mann Whitney U test adjusted for multiple testing, if needed, for independent samples. Intrarater/retest reliability was measured using intraclass correlation coefficients (ICCs) for continuous measures. Five cases were randomly selected for reassessment by the rater. The interpretation of reliability results was based on the recommendations of Koo and Mae (37). Receiver operating characteristics (ROC), with respect to their area-under-the-curve (AUCROC), were compared using DeLong's test. For all statistical tests, the level of significance was defined as p < 0.05. Confidence intervals were calculated at a confidence level of 95%.

## Predictive Modeling

The prediction of treatment response was regarded as a classification problem to be solved by a random forest algorithm, calculated in RStudio 3.4.1 (RStudio, Inc., Boston, MA, USA), using caret 6.0-81 (38). We chose a parallelized random forest as a classifier, as this algorithm is known to give stable and good results in different scenarios (39).

There were 15 parameters available as potential predictors for treatment response: A, $k_{ep}$, and $k_{el}$ from the Brix model; TTP, PE, area under the curve (AUC), wash-in, and wash-out (as parameters derived from the Brix equation); the T1 signal intensity (the raw measurement and the measurement normalized to muscle); the STIR signal intensity (raw measurement and normalized to muscle); ADC; p16-HPV-status; and patient age.

Feature selection was performed using a wrapper approach with a sequential backward selection based on parallel random forests (parRF). To account for class imbalances, the synthetic minority over-sampling technique was applied. The feature selection and training process was focused on maximizing AUCROC. The model was validated using a 10-fold cross-validation approach with 10 repeats.

The algorithm's output is twofold, providing on the one hand probability values for the class assignment (range, 0-1), and on the other hand a dichotomous classification result (response *vs.*

non-response). Hereby, the raw probability values were used to calculate the ROC curves, whereas the dichotomous classification resulted from applying a cutoff of 0.5 on the raw probability values. Sensitivity and specificity were then calculated based on the dichotomous classification results using a cutoff of 0.5. In principle, this cutoff could, however, be further adapted, e.g., to maximize the Youden-Index or to favor either sensitivity or specificity, depending on the clinical setting.

## RESULTS

Patients had histologically confirmed HNSCC stage III–IVb of the oropharynx (n = 14), hypopharynx (n = 3), or supraglottic larynx (n = 3). Multilevel disease was present in two cases. Fifteen patients (68.2%) showed pCR in the re-biopsy. Of the seven patients with residual tumor (ReTu; 31.8%), six showed an inflamed immune phenotype and one biopsy showed an immune-excluded phenotype. Ten patients had human papillomavirus (HPV)-16-positive tumor tissue samples, there was no significant difference for HPV-16 positivity (p = 0.37). Lymph node metastases were present in 12 patients (54.5%). The median initial whole tumor volume (primary tumor plus lymph node metastases) before treatment was 23.7 cm³ (IQR 14.7–40.0). There was no significant difference among patients with pCR and patients with ReTu in whole tumor volume at baseline. The median whole tumor volume after therapy was 15.2 cm³ (IQR 6.4–28.2). The initial median volume of the primary tumor was 13.9 cm³ (IQR 7.2–23.5), and after treatment it was 6.0 cm³ (IQR 2.7–21.0). The initial whole tumor volume did not differ among pCR (median 23.4 cm³) and ReTu (median 24.0 cm³) patients at baseline (see **Table 1**). The median volume of the lymph node metastases at baseline and at follow-up, if present did not differ among groups (see **Table 1**). Primary tumor volume, whole tumor volume and lymph node metastases volume did not differ among groups at baseline and at follow-up (see **Figures 1**–**3**). Representative images of a patient with pCR and a patient of the ReTu group with lymph node metastasis show DCE MRI, morphological and diffusion weighted imaging illustrating the difficulty to determine responders by morphological criteria only (**Figures 4**, **5**).

The correlation analysis of conventional MR imaging features at baseline (T1 ratio, STIR ratio, and ADC value) did not reveal any significant correlation with tumor response (see **Table 2**). Among pCR and ReTu patients, there was a significant difference in the STIR ratio measured in the primary tumor volume at follow-up (p = 0.007, see supplementary 1, raw values supplementary 2). For the pCR and ReTu group, there was no significant difference in the ADC value measured in the primary tumor volume and lymph node metastases at baseline or at follow-up (see supplementary 2). Intrareader agreement for volumetric measurements showed an ICC of 0.91 (95% CI 0.64–0.98), reflecting moderate to excellent agreement according to Koo et al. (37).

**TABLE 1 |** Response rate and tumor volume at baseline and follow up for all patients, pathologic complete response (pCR) group and residual tumor (ReTu) group.

| | All patients | pCR | ReTu |
|---|---|---|---|
| Histological complete response | 22 | 15 | 7 |
| Median initial whole tumor volume (cm$^3$, IQR 25 – 75) | 23.7 (14.7 – 40.0) | 23.4 (14.1 – 39.0) | 24.0 (15.9 – 62.1) |
| Median whole tumor volume after therapy (cm$^3$, IQR 25 – 75) | 15.2 (6.4 – 28.2) | 14.5 (6.1 – 26.9) | 15.8 (9.5 40.4) |
| Lymph node metastases | 12 | 9 | 3 |
| Median initial lymph node metastases volume (cm$^3$, IQR 25 – 75) | 18.0 (9.5 – 25.4) | 16.4 (5.7 – 23.0) | 22.9 (12.5 – 79.0) |
| Median lymph node metastases volume after therapy (cm$^3$, IQR 25 – 75) | 11.5 (4.6 – 22.2) | 11.1 (3.4 – 17.7) | 19.5 (11.6 – 91.7) |
| Median initial primary tumor volume (cm$^3$, IQR 25 – 75) | 13.9 (7.2 – 23.5) | 13.1 (7.2 – 22.3) | 15.9 (5.2 – 27.2) |
| Median primary tumor volume after therapy (cm$^3$, IQR 25 – 75) | 6.0 (2.7 – 21.0) [b] | 5.8 (2.8 – 21.3) | 9.5 (2.4 – 20.9) |

*No significant difference among patients with pathologic complete response (pCR) and patients with residual tumor (ReTu) at each timepoint (baseline or follow-up) was present, significance was defined as p < 0,05 with Kruskal-Wallis-Test corrected for multiple testing.*



**FIGURE 1 |** Tumor volume of the primary tumor in mm$^3$, outliers are marked with dots, no significant differences among groups at baseline or follow-up were present using the Kruskal-Wallis test (pCR BL, pathologic complete response at baseline; ReTu BL, residual tumor at baseline; pCR FU, pathologic complete response at follow up; ReTu FU, residual tumor at follow up).



**FIGURE 3 |** Whole tumor volume of both primary tumor and lymph node metastases in mm$^3$, outliers are marked with dots, no significant differences among groups at baseline or follow-up were present using the Kruskal-Wallis test (pCR BL, pathologic complete response at baseline; ReTu BL, residual tumor at baseline, pCR FU, pathologic complete response at follow up; ReTu FU, residual tumor at follow up).



**FIGURE 2 |** Tumor volume of the lymph node metastases in mm$^3$, outliers are marked with dots, no significant differences among groups at baseline or follow-up were present using the Kruskal-Wallis_test (pCR BL, pathologic complete response at baseline; ReTu BL, residual tumor at baseline; pCR FU, pathologic complete response at follow up; ReTu FU, residual tumor at follow up). In the pCR group 9 patients showed lymph node metastases, in the ReTu group in 3 patient lymph node metastases were present.

DCE parameters at baseline using the Brix model featured significant differences between the pCR and ReTu groups for TTP and wash-out (see **Table 2**). TTP and wash-out reached an AUCROC of 0.771 (95% CI 0.589–0.974) and 0.790 (95% CI 0.595–0.986), respectively.

Our model based on parRF, incorporating $k_{el}$ and TTP from the Brix model, was able to predict therapy response with a sensitivity of 78.7% (95% CI 71.24–84.93) and a specificity of 78.6% (95% CI 67.13–87.48). The model had an AUCROC of 0.866 (95% CI 0.819–0.914) (see **Figure 6**). Although our model showed a higher AUCROC, it failed to significantly outperform its constituent parameters $k_{el}$ and TTP.

## DISCUSSION

In the current study, the primary tumor volume, whole tumor volume and volume of lymph node metastases did not
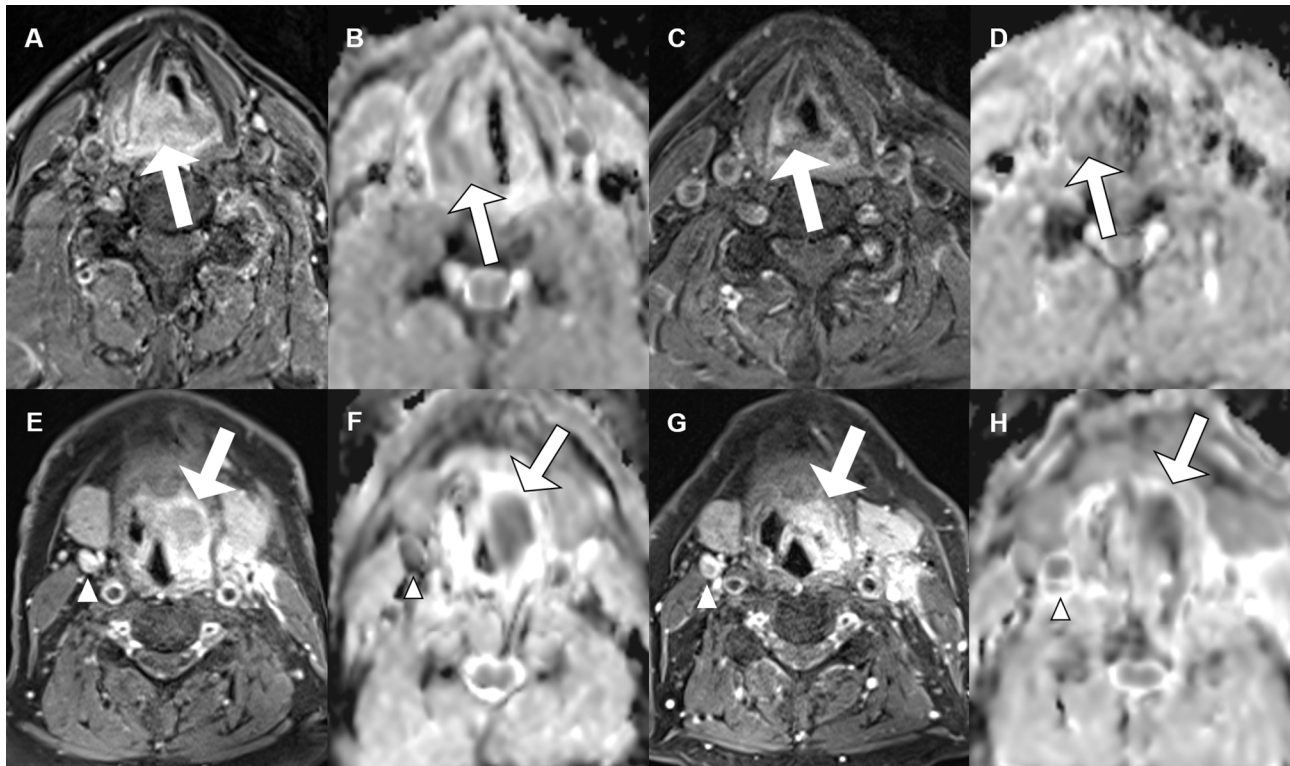
FIGURE 4 | Morphological and functional MRI images for two patients with pharyngeal carcinoma (primary tumor is marked with an arrow, lymph node metastasis with an arrow head), contrast enhanced T1 weighted GRASP-VIBE sequence **(A, C, E, G)** and ADC map **(B, D, F, H)** for both a patient with pathologic complete response (upper row) and a patient with residual tumor (lower row) at baseline [**(A, B)** and respectively **(E, F)**] and at follow-up [**(C, D)** and respectively **(G, H)**].

differ between pCR and ReTu patients. Morphological criteria using the T1 and STIR ratios did not show a consistent association with therapy response, either, suggesting that conventional imaging parameters are not capable of evaluating immunotherapy treatment success after one follow-up. Additionally, the ADC value could not be used to assess response in our dataset. Therefore, the morphological and functional ADC parameters were not reliable response prediction criteria in our dataset.

Multiparametric MRI was used to create a predictive model of tumor response in HNSCC to ICI therapy at baseline using a random forest machine learning algorithm based on DCE parameters of the primary tumor in the baseline examination. The histopathological analysis four weeks after immunotherapy served as a standard of reference for therapy response. Due to tumor heterogeneity, VOIs instead of ROIs were employed. To our knowledge, this is the first study to employ the prognostic value of DCE MRI in head and neck tumors treated with induction chemo-immunotherapy. TTP and wash-out at baseline were significantly different between the pCR and ReTu group and could therefore be used to predict which patients

would have a good response to ICI and chemotherapy treatment. The model based on parRF featured an even higher AUC (0.866) than the parameters alone, with a sensitivity of 78.7% and a specificity of 78.6%.

The introduction of immunotherapy has made the evaluation of tumor response increasingly difficult. In particular, pseudoprogression and mixed responses with both shrinkage and growth of primary tumors or metastases at follow-up are frequently observed, hampering the differentiation between remission and progression. The observation of 15 pCR in 22 treated patients reflects the possibility that induction chemo-immunotherapy may be more efficacious than classical induction chemotherapy (40).

In contrast to our results, the recent study of Borggreve et al. showed that ADC values and the $SUV_{mean}$ in $^{18}$F-FDG PET/CT can help identify pathologic complete response to neoadjuvant chemotherapy in esophageal cancer (41). These different findings might be attributed to the different tumor entities and therapy regimens under investigation.

Possible solutions for assessing tumor response could be the use of invasive techniques, such as tumor biopsy, or

TABLE 2 | Median and interquartile ranges (IQR) for DCE parameters at baseline, conventional MRI imaging features, ADC values in whole tumor volume and age with respective p-values using Wilcoxon signed rank-test and area under the curve of the Receiver Operating Characteristic (AUCROC).

| | pCR | ReTu | p-value | AUCROC |
|---|---|---|---|---|
| **A** | 1.967 | 2.126 | 0.2101 | 0.676 |
| | (IQR 1.725–2.043) | (IQR 1.968–2.208) | | |
| **kep** | 0.046 | 0.048 | 0.7309 | 0.552 |
| | (IQR 0.042–0.053) | (IQR 0.044–0.057) | | |
| **kel** | $-1.52\times10^{-4}$ | $4.48\times10^{-4}$ | 0.0659 | 0.752 |
| | (IQR $-2.8$–$3.9\times10^{-4}$) | (IQR $1.8$–$4.9\times10^{-4}$) | | |
| **TTP** | 276 | 102 | 0.04809 | 0.771 |
| | (IQR 118–280) | (IQR 93–112) | | |
| **PE** | 2.02 | 2.04 | 0.8907 | 0.524 |
| | (IQR 1.81–2.18) | (IQR 1.83–2.16) | | |
| **AUC** | 80.7 | 84.3 | 0.3322 | 0.638 |
| | (IQR 68.8–85.1) | (IQR 80.6–89.0) | | |
| **WIN** | 0.087 | 0.096 | 0.2372 | 0.667 |
| | (IQR 0.075–0.101) | (IQR 0.091–0.112) | | |
| **WOUT** | $3.06\times10^{-4}$ | $-8.990\times10^{-4}$ | 0.03194 | 0.79 |
| | (IQR $-3.78$–$6.07\times10^{-4}$) | (IQR $-9.42$–$-3.68\times10^{-4}$) | | |
| **T1** | 434 | 482 | 0.6298 | 0.571 |
| | (IQR 396–512) | (IQR 461–490) | | |
| **T1 Ratio** | 1.11 | 1.04 | 0.4069 | 0.619 |
| | (IQR 0.98–1.33) | (IQR 0.91–1.15) | | |
| **STIR** | 323 | 377 | 0.6216 | 0.571 |
| | (IQR 273–377) | (IQR 277–437) | | |
| **STIR Ratio** | 4.90 | 4.21 | 0.1624 | 0.695 |
| | (IQR 4.45–5.50) | (IQR 2.97–4.71) | | |
| **ADC** | 1029 | 850 | 0.7309 | 0.552 |
| | (IQR 803–1372) | (IQR 822–1228) | | |
| **Age** | 60 | 61 | 0.6715 | 0.562 |
| | (IQR 54–67) | (IQR 58–66) | | |

incorporating more than one follow-up imaging study according to iRECIST criteria. Although a noninvasive method for immunotherapy response evaluation remains a challenge, it is crucial to develop cost-effective and personalized estimations of tumor response to ICI treatment. Several models using radiomics have been developed to evaluate tumor response; for example, a CD8 score using contrast-enhanced CT images was associated with tumor response in patients treated with ICIs and radiotherapy (42–44). Hao et al. applied multiparametric MRI in osteosarcoma patients at baseline and follow-up after neoadjuvant chemotherapy and surgery and was able to predict event-free survival and overall response using the DCE MRI parameter Ktrans (45). Our predictive model performed slightly better than those in other studies using radiomics in CT to predict response to ICIs. For example, Ligero et al. showed a sensitivity of 75% and a specificity of 53% in tumor types of various origins, such as breast, cervix, bladder, lung, and head and neck (46). Despite of the better performance our algorithm, the results did not differ significantly from the single DCE parameters used for the model.

Overall, our findings indicate that DCE parameters are promising for predicting immunotherapy treatment responses. Our results were histopathologically validated for every patient; nevertheless, further studies in larger populations should be performed. In a clinical context, the predictive value based on DCE MRI could facilitate the optimized selection of individual treatment options for each HNSCC patient using a noninvasive approach.

There are some limitations to this study: This was a single-center study with a limited number of patients enrolled. Because of this small sample size, the algorithm's results should also be interpreted with caution. Ideally, one would have initially excluded a part of the data set as an independent test set, and trained the algorithm with the remaining data in order to validate its performance on the test set. However, with such a small number of samples, this approach was not an option. We thus decided to implement a particular cross-validation, which bears the risk of falsely estimating the algorithm's performance too optimistically.

With these limitations in mind, the predictive model performed slightly better than the sole DCE parameters, however missing the significance threshold. For these reasons, a clinical application is limited at the moment. Performance measures with a higher degree of reliability will result from a larger sample size and the inclusion of an independent test set, both planned for follow-up studies.

Further limitations include the fact that the results are related to a fixed therapy regiment consisting of ICI and chemotherapy in HNSCC. The definition of pCR was based on a biopsy sample rather than the resected tumor. However, the samples were taken from regions that were metabolically active on FDG PET/CT before and after induction therapy.
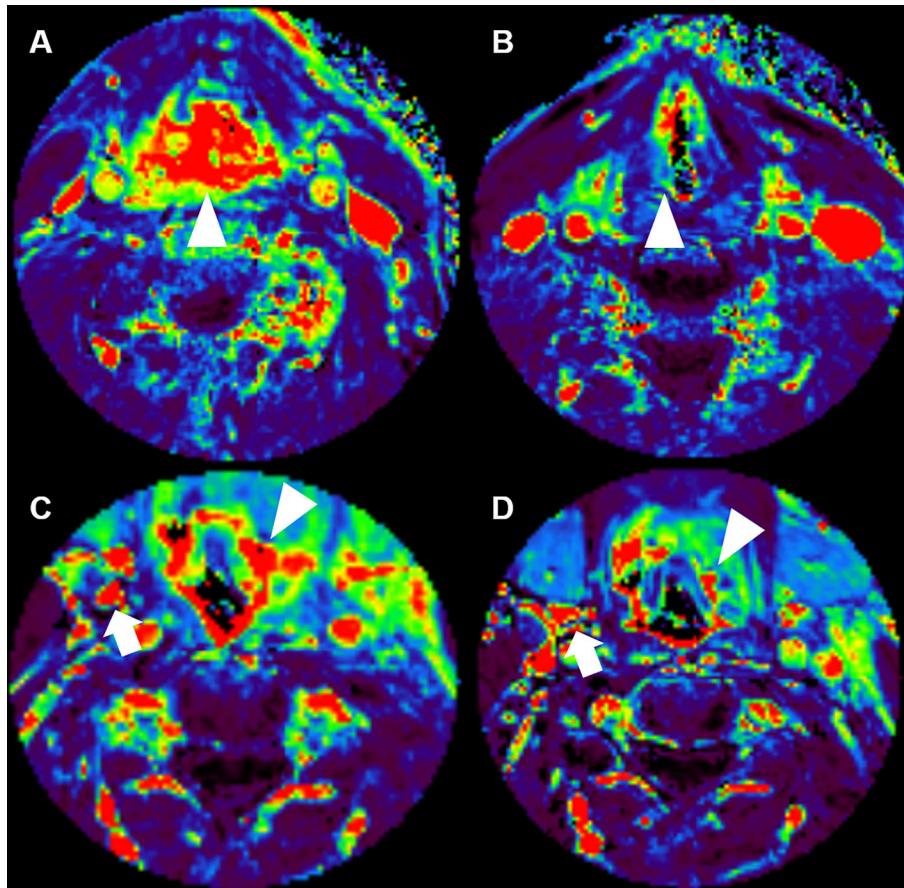
**FIGURE 5** | Corresponding positive enhancement integral maps of DCE MRI for the same patients as shown in , pathologic complete response (upper row) and residual tumor (lower row) at baseline **(A, C)** and at follow-up **(B, D)**. The area of the primary tumor is marked with an arrowhead, lymph node metastasis is marked with an arrow.
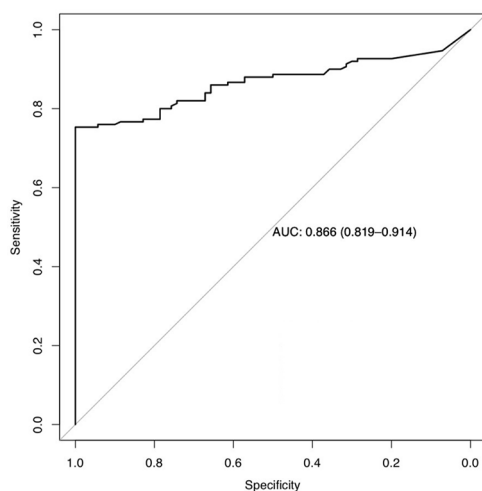


**FIGURE 6** | Area under the Receiver Operating Characteristic curve (AUC) and statistical data, in parentheses 95% confidence interval for the random forest model based on $k_{el}$ und TTP.

In conclusion, we found that tumor volume, single morphological parameters, and the ADC values are of limited use for the evaluation of treatment response to chemo-immunotherapy in locally advanced HNSCC. However, a machine learning algorithm using parallel random forests based on DCE MRI parameters was able to predict treatment response following induction chemo-immunotherapy in HNSCC patients, but did not perform better than the DCE parameters alone.

## 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 leading institutional review board at the Friedrich-Alexander-Universität Erlangen-Nürnberg. The patients/

participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

KH, SE, MH, and TB designed the study. SR, SS, AG, BF, UG, and MH investigated the patients, had trial oversight and collected the samples. ME and AH conducted the histopathological examination. KH, SE, MW, and TB performed the image analysis and carried out the statistical analysis. HI, RF, AH, and MU supervised and substantially supported the acquisition of data based on their vast experience. KH, MH, and TB drafted the manuscript. All authors reviewed the manuscript critically and provided constructive comments to improve the quality of the manuscript. All authors read and approved the final manuscript.

## FUNDING

## ACKNOWLEDGMENTS

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2021.734872/full#supplementary-material

## REFERENCES

1. Borghaei H, Paz-Ares L, Horn L, Spigel DR, Steins M, Ready NE, et al. Nivolumab *Versus* Docetaxel in Advanced Nonsquamous Non-Small-Cell Lung Cancer. *N Engl J Med* (2015) 373(17):1627–39. doi: 10.1056/NEJMoa1507643

2. Brahmer J, Reckamp KL, Baas P, Crino L, Eberhardt WE, Poddubskaya E, et al. Nivolumab *Versus* Docetaxel in Advanced Squamous-Cell Non-Small-Cell Lung Cancer. *N Engl J Med* (2015) 373(2):123–35. doi: 10.1056/NEJMoa1504627

3. Herbst RS, Baas P, Kim DW, Felip E, Perez-Gracia JL, Han JY, et al. Pembrolizumab *Versus* Docetaxel for Previously Treated, PD-L1-Positive, Advanced Non-Small-Cell Lung Cancer (KEYNOTE-010): A Randomised Controlled Trial. *Lancet* (2016) 387(10027):1540–50. doi: 10.1016/S0140-6736(15)01281-7

4. Long GV, Weber JS, Larkin J, Atkinson V, Grob JJ, Schadendorf D, et al. Nivolumab for Patients With Advanced Melanoma Treated Beyond Progression: Analysis of 2 Phase 3 Clinical Trials. *JAMA Oncol* (2017) 3 (11):1511–9. doi: 10.1001/jamaoncol.2017.1588

5. Postow MA, Chesney J, Pavlick AC, Robert C, Grossmann K, McDermott D, et al. Nivolumab and Ipilimumab *Versus* Ipilimumab in Untreated Melanoma. *N Engl J Med* (2015) 372(21):2006–17. doi: 10.1056/NEJMoa1414428

6. Ribas A, Puzanov I, Dummer R, Schadendorf D, Hamid O, Robert C, et al. Pembrolizumab *Versus* Investigator-Choice Chemotherapy for Ipilimumab-Refractory Melanoma (KEYNOTE-002): A Randomised, Controlled, Phase 2 Trial. *Lancet Oncol* (2015) 16(8):908–18. doi: 10.1016/S1470-2045(15)00083-2

7. Burtness B, Harrington KJ, Greil R, Soulieres D, Tahara M, de Castro GJr., et al. Pembrolizumab Alone or With Chemotherapy *Versus* Cetuximab With Chemotherapy for Recurrent or Metastatic Squamous Cell Carcinoma of the Head and Neck (KEYNOTE-048): A Randomised, Open-Label, Phase 3 Study. *Lancet* (2019) 394(10212):1915–28. doi: 10.1016/S0140-6736(19)32591-7

8. Ferris RL, Blumenschein GJr., Fayette J, Guigay J, Colevas AD, Licitra L, et al. Nivolumab for Recurrent Squamous-Cell Carcinoma of the Head and Neck. *N Engl J Med* (2016) 375(19):1856–67. doi: 10.1056/NEJMoa1602252

9. Schadendorf D, Hodi FS, Robert C, Weber JS, Margolin K, Hamid O, et al. Pooled Analysis of Long-Term Survival Data From Phase II and Phase III Trials of Ipilimumab in Unresectable or Metastatic Melanoma. *J Clin Oncol* (2015) 33(17):1889–94. doi: 10.1200/JCO.2014.56.2736

10. Gettinger S, Horn L, Jackman D, Spigel D, Antonia S, Hellmann M, et al. Five-Year Follow-Up of Nivolumab in Previously Treated Advanced Non-Small-Cell Lung Cancer: Results From the CA209-003 Study. *J Clin Oncol* (2018) 36 (17):1675–84. doi: 10.1200/JCO.2017.77.0412

11. Zou W, Wolchok JD, Chen L. PD-L1 (B7-H1) and PD-1 Pathway Blockade for Cancer Therapy: Mechanisms, Response Biomarkers, and Combinations. *Sci Transl Med* (2016) 8(328):328rv4. doi: 10.1126/scitranslmed.aad7118

12. Gibney GT, Weiner LM, Atkins MB. Predictive Biomarkers for Checkpoint Inhibitor-Based Immunotherapy. *Lancet Oncol* (2016) 17(12):e542–e51. doi: 10.1016/S1470-2045(16)30406-5

13. Buder-Bakhaya K, Hassel JC. Biomarkers for Clinical Benefit of Immune Checkpoint Inhibitor Treatment-A Review From the Melanoma Perspective and Beyond. *Front Immunol* (2018) 9:1474. doi: 10.3389/fimmu.2018.01474

14. Daud AI, Wolchok JD, Robert C, Hwu WJ, Weber JS, Ribas A, et al. Programmed Death-Ligand 1 Expression and Response to the Anti-Programmed Death 1 Antibody Pembrolizumab in Melanoma. *J Clin Oncol* (2016) 34(34):4102–9. doi: 10.1200/JCO.2016.67.2477

15. Tumeh PC, Harview CL, Yearley JH, Shintaku IP, Taylor EJ, Robert L, et al. PD-1 Blockade Induces Responses by Inhibiting Adaptive Immune Resistance. *Nature* (2014) 515(7528):568–71. doi: 10.1038/nature13954

16. Le DT, Uram JN, Wang H, Bartlett BR, Kemberling H, Eyring AD, et al. PD-1 Blockade in Tumors With Mismatch-Repair Deficiency. *N Engl J Med* (2015) 372(26):2509–20. doi: 10.1056/NEJMoa1500596

17. Rizvi NA, Hellmann MD, Snyder A, Kvistborg P, Makarov V, Havel JJ, et al. Cancer Immunology. Mutational Landscape Determines Sensitivity to PD-1 Blockade in Non-Small Cell Lung Cancer. *Science* (2015) 348(6230):124–8. doi: 10.1126/science.aaa1348

18. Rasmussen JH, Lelkaitis G, Hakansson K, Vogelius IR, Johannesen HH, Fischer BM, et al. Intratumor Heterogeneity of PD-L1 Expression in Head and Neck Squamous Cell Carcinoma. *Br J Cancer* (2019) 120(10):1003–6. doi: 10.1038/s41416-019-0449-y

19. Clasen K, Leibfarth S, Hilke FJ, Admard J, Winter RM, Welz S, et al. PET/MRI and Genetic Intrapatient Heterogeneity in Head and Neck Cancers. *Strahlenther Onkol* (2020) 196(6):542–51. doi: 10.1007/s00066-020-01606-y

20. Zhou JG, Donaubauer AJ, Frey B, Becker I, Rutzner S, Eckstein M, et al. Prospective Development and Validation of a Liquid Immune Profile-Based Signature (LIPS) to Predict Response of Patients With Recurrent/Metastatic Cancer to Immune Checkpoint Inhibitors. *J Immunother Cancer* (2021) 9(2): e001845. doi: 10.1136/jitc-2020-001845

21. Kanjanapan Y, Day D, Wang L, Al-Sawaihey H, Abbas E, Namini A, et al. Hyperprogressive Disease in Early-Phase Immunotherapy Trials: Clinical Predictors and Association With Immune-Related Toxicities. *Cancer* (2019) 125(8):1341–9. doi: 10.1002/cncr.31999

22. Borcoman E, Kanjanapan Y, Champiat S, Kato S, Servois V, Kurzrock R, et al. Novel Patterns of Response Under Immunotherapy. *Ann Oncol* (2019) 30 (3):385–96. doi: 10.1093/annonc/mdz003

23. Seymour L, Bogaerts J, Perrone A, Ford R, Schwartz LH, Mandrekar S, et al. iRECIST: Guidelines for Response Criteria for Use in Trials Testing Immunotherapeutics. *Lancet Oncol* (2017) 18(3):e143–52. doi: 10.1016/ S1470-2045(17)30074-8

24. Hodi FS, Ballinger M, Lyons B, Soria JC, Nishino M, Tabernero J, et al. Immune-Modified Response Evaluation Criteria In Solid Tumors (imRECIST): Refining Guidelines to Assess the Clinical Benefit of Cancer Immunotherapy. *J Clin Oncol* (2018) 36(9):850–8. doi: 10.1200/ JCO.2017.75.1644

25. Wahl RL, Jacene H, Kasamon Y, Lodge MA. From RECIST to PERCIST: Evolving Considerations for PET Response Criteria in Solid Tumors. *J Nucl Med* (2009) 50 Suppl 1:122S–50S. doi: 10.2967/jnumed.108.057307

26. Decazes P, Bohn P. Immunotherapy by Immune Checkpoint Inhibitors and Nuclear Medicine Imaging: Current and Future Applications. *Cancers (Basel)* (2020) 12(2):371. doi: 10.3390/cancers12020371

27. Turnbull LW. Dynamic Contrast-Enhanced MRI in the Diagnosis and Management of Breast Cancer. *NMR BioMed* (2009) 22(1):28–39. doi: 10.1002/nbm.1273

28. Musall BC, Abdelhafez AH, Adrada BE, Candelaria RP, Mohamed RMM, Boge M, et al. Functional Tumor Volume by Fast Dynamic Contrast-Enhanced MRI for Predicting Neoadjuvant Systemic Therapy Response in Triple-Negative Breast Cancer. *J Magn Reson Imaging* (2021) 54(1):251-260. doi: 10.1002/jmri.27557

29. Kim SH, Lee JY, Lee JM, Han JK, Choi BI. Apparent Diffusion Coefficient for Evaluating Tumour Response to Neoadjuvant Chemoradiation Therapy for Locally Advanced Rectal Cancer. *Eur Radiol* (2011) 21(5):987–95. doi: 10.1007/s00330-010-1989-y

30. Intven M, Reerink O, Philippens ME. Dynamic Contrast Enhanced MR Imaging for Rectal Cancer Response Assessment After Neo-Adjuvant Chemoradiation. *J Magn Reson Imaging* (2015) 41(6):1646–53. doi: 10.1002/ jmri.24718

31. Treutlein C, Stollberg A, Scherl C, Agaimy A, Ellmann S, Iro H, et al. Diagnostic Value of 3D Dynamic Contrast-Enhanced Magnetic Resonance Imaging in Lymph Node Metastases of Head and Neck Tumors: A Correlation Study With Histology. *Acta Radiol Open* (2020) 9(8):2058460120951966. doi: 10.1177/2058460120951966

32. Tanadini-Lang S, Balermpas P, Guckenberger M, Pavic M, Riesterer O, Vuong D, et al. Radiomic Biomarkers for Head and Neck Squamous Cell Carcinoma. *Strahlenther Onkol* (2020) 196(10):868–78. doi: 10.1007/s00066-020-01638-4

33. Hecht M, Gostian AO, Eckstein M, Rutzner S, von der Grun J, Illmer T, et al. Safety and Efficacy of Single Cycle Induction Treatment With Cisplatin/ Docetaxel/ Durvalumab/Tremelimumab in Locally Advanced HNSCC: First Results of CheckRad-Cd8. *J Immunother Cancer* (2020) 8(2):e001378. doi: 10.1136/jitc-2020-001378

34. Tomppert A, Wuest W, Wiesmueller M, Heiss R, Kopp M, Nagel AM, et al. Achieving High Spatial and Temporal Resolution With Perfusion MRI in the Head and Neck Region Using Golden-Angle Radial Sampling. *Eur Radiol* (2020) 31(4):2263–71. doi: 10.1007/s00330-020-07263-0

35. Brix G, Semmler W, Port R, Schad LR, Layer G, Lorenz WJ. Pharmacokinetic Parameters in CNS Gd-DTPA Enhanced MR Imaging. *J Comput Assist Tomogr* (1991) 15(4):621–8. doi: 10.1097/00004728-199107000-00018

36. Daniel Padfield GM. *Non-Linear Least Squares Regressions With the Levenberg-Marquardt Algorithm Using Multiple Starting Values for Increasing the Chance That the Minimum Found is the Global Minimum. 1.2.0 Ed* (2020). Available at: https://cran.r-project.org/web/packages/nls. multstart/index.html.

37. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med* (2016) 15(2):155–63. doi: 10.1016/j.jcm.2016.02.012

38. Kuhn M. *Building Predictive Models in R Using the Caret Package. 2008*, Vol. 28. (2008). p. 26. doi: 10.18637/jss.v028.i05

39. Fernandez-Delgado M, Cernadas E, Barro S. Do We Need Hundreds of Classifiers to Solve Real World Classification Problems? *J Mach Learn Res* (2014) 15:3133–81.

40. Semrau S, Gostian AO, Traxdorf M, Eckstein M, Rutzner S, von der Grun J, et al. Implementation of Double Immune Checkpoint Blockade Increases Response Rate to Induction Chemotherapy in Head and Neck Cancer. *Cancers (Basel)* (2021) 13(8):1959. doi: 10.3390/cancers13081959

41. Borggreve AS, Goense L, van Rossum PSN, Heethuis SE, van Hillegersberg R, Lagendijk JJW, et al. Preoperative Prediction of Pathologic Response to Neoadjuvant Chemoradiotherapy in Patients With Esophageal Cancer Using (18)F-FDG PET/CT and DW-MRI: A Prospective Multicenter Study. *Int J Radiat Oncol Biol Phys* (2020) 106(5):998–1009. doi: 10.1016/ j.ijrobp.2019.12.038

42. Sun KY, Hu HT, Chen SL, Ye JN, Li GH, Chen LD, et al. CT-Based Radiomics Scores Predict Response to Neoadjuvant Chemotherapy and Survival in Patients With Gastric Cancer. *BMC Cancer* (2020) 20(1):468. doi: 10.1186/ s12885-020-06970-7

43. Sun R, Limkin EJ, Vakalopoulou M, Dercle L, Champiat S, Han SR, et al. A Radiomics Approach to Assess Tumour-Infiltrating CD8 Cells and Response to Anti-PD-1 or Anti-PD-L1 Immunotherapy: An Imaging Biomarker, Retrospective Multicohort Study. *Lancet Oncol* (2018) 19(9):1180–91. doi: 10.1016/S1470-2045(18)30413-3

44. Sun R, Sundahl N, Hecht M, Putz F, Lancia A, Rouyar A, et al. Radiomics to Predict Outcomes and Abscopal Response of Patients With Cancer Treated With Immunotherapy Combined With Radiotherapy Using a Validated Signature of CD8 Cells. *J Immunother Cancer* (2020) 8(2):e001429. doi: 10.1136/jitc-2020-001429

45. Hao Y, An R, Xue Y, Li F, Wang H, Zheng J, et al. Prognostic Value of Tumoral and Peritumoral Magnetic Resonance Parameters in Osteosarcoma Patients for Monitoring Chemotherapy Response. *Eur Radiol* (2020) 31 (5):3518–29. doi: 10.1007/s00330-020-07338-y

46. Ligero M, Garcia-Ruiz A, Viaplana C, Villacampa G, Raciti MV, Landa J, et al. A CT-Based Radiomics Signature Is Associated With Response to Immune Checkpoint Inhibitors in Advanced Solid Tumors. *Radiology* (2021) 299 (1):109–19. doi: 10.1148/radiol.2021200928

# Bone Metastases Are Measurable: The Role of Whole-Body MRI and Positron Emission Tomography

Daniela E. Oprea-Lager[1,2*], Matthijs C.F. Cysouw[2], Ronald Boellaard[2], Christophe M. Deroose[1,3,4], Lioe-Fee de Geus-Oei[5,6], Egesta Lopci[7], Luc Bidaut[1,8], Ken Herrmann[9], Laure S. Fournier[1,10,11], Tobias Bäuerle[12], Nandita M. deSouza[1,11,13] and Frederic E. Lecouvet[1,14]

[1] Imaging Group, European Organisation of Research and Treatment in Cancer (EORTC), Brussels, Belgium, [2] Department of Radiology and Nuclear Medicine, Cancer Center Amsterdam, Amsterdam University Medical Center, Vrije Universiteit Amsterdam, Amsterdam, Netherlands, [3] Nuclear Medicine, University Hospitals Leuven, Leuven, Belgium, [4] Nuclear Medicine & Molecular Imaging, Department of Imaging and Pathology, KU Leuven, Leuven, Belgium, [5] Department of Radiology, Leiden University Medical Center, Leiden, Netherlands, [6] Biomedical Photonic Imaging Group, University of Twente, Enschede, Netherlands, [7] Nuclear Medicine Unit, IRCCS – Humanitas Research Hospital, Milan, Italy, [8] College of Science, University of Lincoln, Lincoln, United Kingdom, [9] Department of Nuclear Medicine, University of Duisburg-Essen, and German Cancer Consortium (DKTK)-University Hospital Essen, Essen, Germany, [10] Paris Cardiovascular Research Center (PARCC), Institut National de la Santé et de la Recherche Médicale (INSERM), Radiology Department, Assistance Publique-Hôpitaux de Paris (AP-HP), Hopital europeen Georges Pompidou, Université de Paris, Paris, France, [11] European Imaging Biomarkers Alliance (EIBALL), European Society of Radiology, Vienna, Austria, [12] Institute of Radiology, University Hospital Erlangen, Friedrich-Alexander University Erlangen-Nürnberg, Erlangen, Germany, [13] Division of Radiotherapy and Imaging, The Institute of Cancer Research and Royal Marsden NHS Foundation Trust, London, United Kingdom, [14] Department of Radiology, Institut de Recherche Expérimentale et Clinique (IREC), Cliniques Universitaires Saint Luc, Université Catholique de Louvain (UCLouvain), Brussels, Belgium

Metastatic tumor deposits in bone marrow elicit differential bone responses that vary with the type of malignancy. This results in either sclerotic, lytic, or mixed bone lesions, which can change in morphology due to treatment effects and/or secondary bone remodeling. Hence, morphological imaging is regarded unsuitable for response assessment of bone metastases and in the current Response Evaluation Criteria In Solid Tumors 1.1 (RECIST1.1) guideline bone metastases are deemed unmeasurable. Nevertheless, the advent of functional and molecular imaging modalities such as whole-body magnetic resonance imaging (WB-MRI) and positron emission tomography (PET) has improved the ability for follow-up of bone metastases, regardless of their morphology. Both these modalities not only have improved sensitivity for visual detection of bone lesions, but also allow for objective measurements of bone lesion characteristics. WB-MRI provides a global assessment of skeletal metastases and for a one-step "all-organ" approach of metastatic disease. Novel MRI techniques include diffusion-weighted imaging (DWI) targeting highly cellular lesions, dynamic contrast-enhanced MRI (DCE-MRI) for quantitative assessment of bone lesion vascularization, and multiparametric MRI (mpMRI) combining anatomical and functional sequences. Recommendations for a homogenization of MRI image acquisitions and generalizable response criteria have been developed. For PET, many metabolic and molecular radiotracers are available, some targeting tumor characteristics not confined to cancer type (e.g. ^{18}F-FDG) while

other targeted radiotracers target specific molecular characteristics, such as prostate specific membrane antigen (PSMA) ligands for prostate cancer. Supporting data on quantitative PET analysis regarding repeatability, reproducibility, and harmonization of PET/CT system performance is available. Bone metastases detected on PET and MRI can be quantitatively assessed using validated methodologies, both on a whole-body and individual lesion basis. Both have the advantage of covering not only bone lesions but visceral and nodal lesions as well. Hybrid imaging, combining PET with MRI, may provide complementary parameters on the morphologic, functional, metabolic and molecular level of bone metastases in one examination. For clinical implementation of measuring bone metastases in response assessment using WB-MRI and PET, current RECIST1.1 guidelines need to be adapted. This review summarizes available data and insights into imaging of bone metastases using MRI and PET.

## INTRODUCTION

Bone is a common site of secondary tumor deposits because, in addition to its rigid, calcified, outer cortex, it has a richly vascular inner marrow of bony trabeculae, stroma, haematopoeitic tissue and fat (1). Within bone, it is the crucial balance between osteoblastic and osteoclastic elements that maintains its functional strength and rigidity. Metastatic deposits elicit differential responses from the osteblastic and osteoclastic components, which vary with the type of malignancy and result in strikingly different appearances on imaging (2). In some cases, the tumor incites a predominantly osteoblastic response with a resulting increase in calcified sclerotic matrix, as in prostate and breast cancer (3). In other tumor types, the metastasis causes bony destruction (osteoclastic response) without exciting an osteoblastic response, so that metastases (e.g. kidneys, thyroid, lungs) appear lytic and expansile (3). Finally, the tumor cells can simply invade the marrow without influence on the mineral content of the bone (i.e. radio-occult metastases). In many instances there is a mixture of sclerotic, lytic and radio-occult types. As treatment response is often accompanied by an increase in bony sclerosis ("flare response"), it can be difficult to differentiate it from an osteoblastic response to the tumor itself (4). Moreover, once deformed by the presence of metastases, the rigid form of the bony skeleton does not usually remodel sufficiently after treatment to distinguish untreated from treated tumor. Therefore, on morphological imaging, especially X-ray based, evaluation of response to treatment of bone metastases remains difficult.

RECIST were presented more than 2 decades ago and rely principally on unidimensional size measurements (5). Nowadays, RECIST forms the mainstay of response evaluation of solid tumors to treatment and is universally used in clinical trials of solid tumors. Index lesions with well-defined margins, discernable from adjacent parenchyma are required for reproducible measurements, and specific modifications are set out for some tissues (short-axis measurements for lymph nodes, bi-dimensional measurements for brain lesions). However, because of the blastic response of bone to tumor or to treatment, and of the rigid nature of calcified

bone where deformity of the cortex persists after treatment, bone lesions were considered unmeasurable by RECIST. Modifications to RECIST (i.e., RECIST 1.1) stated that bone metastases with soft tissue masses > 10 mm could be considered measurable index lesions (6). Nevertheless, as reduction of the soft tissue component renders the lesions unmeasurable by these criteria again, there remains a critical unmet need for a means of quantifying bone lesions and their response to treatment.

The advent of imaging modalities providing information about tissue microstructure or its metabolism has accelerated the identification of skeletal metastases. $^{18}$F-fluorodeoxyglucose ($^{18}$F-FDG) PET/CT identifies secondary deposits within bone because of their increased glucose turnover. Its whole-body coverage and increasingly widespread availability has made it of primary importance in cancer staging, particularly in patients where the tumor pathology or molecular profile indicates a high metastatic risk (7, 8). Additionally, techniques such as WB-MRI with DWI have a high sensitivity for identifying highly cellular lesions such as tumors and have been incorporated routinely into the staging of some tumor types such as myeloma (9–11). Dynamic contrast-enhanced MRI (DCE-MRI) for quantitatively assessing vascularization within bone marrow in patients with multiple myeloma was found to be of prognostic significance for these patients (12, 13). While these techniques have their own limitations, they are not hampered by what makes bone lesions unmeasurable by RECIST 1.1 (i.e. radio-occult appearance, sclerotic response and persistent bone deformity on healing). The purpose of this manuscript is to review the MRI and PET techniques available for measuring bone metastases, their opportunities and challenges, and their applicability in various tumor types.

## DIFFERENT CANCERS – DIFFERENT TYPES OF BONE METASTASES

At present, the incidence of bone metastases is 65-75% in advanced metastatic breast cancer, 65-75% in prostate cancer, 60% in thyroid cancer, 30-40% in lung cancer, 40% in bladder

cancer, 20-25% in renal cell carcinoma and 14-45% in melanoma (14).

Bone metastases can be classified as osteolytic, osteoblastic, radio-occult, or as a mixed type. Osteolytic metastases are characterized by destruction of normal bone and osteoblastic/sclerotic metastases are characterized by deposition of new bone. Radio-occult lesions have no impact on the mineral content of the bone. Osteolytic lesions are predominantly present in multiple myeloma, renal cell carcinoma, melanoma, non-small cell lung cancer (NSCLC), non-Hodgkin lymphoma (NHL), thyroid cancer, Langerhans-cell histiocytosis and breast cancer, and osteoblastic lesions are present in prostate cancer, neuroendocrine tumors, small-cell lung cancer (SCLC), Hodgkin lymphoma and medulloblastoma (14). Mixed lesions can be found in gastrointestinal cancers and squamous cancers, and 15-20% of bone metastases of breast cancer can be either osteoblastic or mixed (14). Radio-occult lesions can be present in virtually all tumor types. The mechanisms responsible for the impact of metastatic tumor growth on the mineral content of the skeleton are complex and involve the stimulation of osteoclasts and osteoblasts by tumor cells expressing factors. The resulting imbalance between resorption and production of bone matrix subsequently leads to osteoclastic, osteoblastic, or mixed metastatic disease (2).

In osteolytic lesions, bone destruction is primarily mediated by osteoclasts and, in later stages, ischemia can play a role due to the compression of the vasculature (15). Parathyroid hormone-related peptide (PTHrP) induces osteoblasts to produce a receptor activator of nuclear factor κB ligand, which stimulates osteoclast maturation, and thereby plays a critical role in the development of osteolytic lesions. Increased osteoclast activity leads to bone resorption that exceeds the reparative ability of osteoblasts (16). It releases factors from the bone matrix that stimulate PTHrP, thereby creating a vicious cycle. In osteoblastic lesions, osteoblast generation is influenced by transforming growth factor, bone morphogenic proteins (BMP), and endothelin-1 (17). Tumor-derived growth factors stimulate primarily osteoblasts rather than osteoclasts, resulting in deposition of excess abnormal bone. PTHrP can be cleaved by prostate-specific antigen (PSA), resulting in an osteoblastic reaction and decreased bone reabsorption. Furthermore, osteoblast differentiation is influenced by core binding factor alpha1, also known as Runx-2 (14). Osteoblast activity may also increase as a reparative process in successfully treated bone metastases, which can be visible on molecular imaging as the so-called "flare phenomenon" and can cause lesions to become denser on radiographs or CT scans (18).

After the tumor cells have left the primary tumor and are in circulation, the bone tumor microenvironment needs to provide a fertile ground (the soil), for the survival and growth of metastatic cancer cells (the seed) (19). Vascular adhesion and extravasation need to occur, and the tumor cells have to remain at the metastatic site. Subsequently, chemo-attractive and adhesion molecules play an important role in the retention of the tumor cells in the bone marrow vasculature. In turn, tumor cells use equivalent molecules, such as chemokines, integrins, osteopontin, bone sialoprotein and type I collagen for organ colonization (20). The microenvironment supports cancer cell survival and growth by producing promoting factors that may contribute to bone metastases development. Subsequently, epithelial-mesenchymal transition occurs, which enables epithelial cells to migrate to a new environment. While this occurs mainly during embryogenesis, in cancer cells this process denotes the invasive phenotype (21).

Sex-associated differences exist in bone metastasis formation from breast-, lung- and prostate cancer. In breast cancer, estrogen influences the bone microenvironment by creating and conditioning a favorable niche for colonization of breast cancer cells. Patients with estrogen receptor α positive (ER+) tumors have bone metastases three times more often than do patients with ER- tumors (22). In lung cancer, it is reported that females more often have bone metastases due to a more favorable bone microenvironment for metastasis formation. In prostate cancer patients, a decrease in the androgen-to-estrogen balance results in bone metastasis formation, with a potentially important role for ERβ that may be similar to that in breast cancer. Androgens as well as estrogens have an influence on osteoblast proliferation and on bone resorbing osteoclasts. In both males and females, estrogens have a dominant effect on bone maintenance and can directly inhibit osteoclasts. Furthermore, androgens directly contribute to male periosteal bone expansion, mineralization, and trabecular bone maintenance (23).

The time from primary diagnosis to the development of bone metastasis can range from months to decades. This implies that tumor cells can lay dormant for significant periods of time after they leave the primary site. It has been shown that the bone is an important reservoir for dormant tumor cells. The best-illustrated cases for clinical dormancy are in breast cancer, where ER+ patients show late recurrences, sometimes decades after removal of the primary tumor. Latent bone metastasis formation likely depends on estrogen regulation, and it is significantly higher in ER+ cases (24).

Bone metastases have unique disease-specific characteristics, such as longevity, fracture healing rates, local and systemic disease progression, and sensitivity to adjuvant treatments. Bone metastases from lung cancer and renal cancer can also show acral distribution (25). Patients with bone metastases of lung cancer historically showed a median survival of approximately 6 months (14). Treatment options for patients with identifiable mutations include immunotherapy and epidermal growth factor receptor tyrosine kinase inhibitors, with evidently improved survival benefit (26). Bone metastases of lung cancer are, in general, sensitive to radiation therapy (27).

The median survival of breast cancer patients with bone-only metastasis is 36 months (28). The medical treatment of breast cancer depends on the hormone receptor and HER-2/neu status and is different for premenopausal and postmenopausal women (25). Furthermore, pain reduction can be achieved, and skeletal-related events and the development of new skeletal lesions can be prevented by the use of bisphosphonates or denosumab, due to their ability to limit bone resorption. Bone metastases of breast cancer are in general radiosensitive, resulting in a lower proportion of surgical treatments (29).

Men with prostate cancer, a good performance status, and bone-only disease have a median duration of disease control after androgen blockade of 4 years and a median survival of 53 months (14). Bone metastases of prostate cancer have a predilection for the axial skeleton, resulting in an increased risk for spinal cord compression (25). However, due to the osteoblastic nature of the metastases, skeletal-related events are relatively uncommon. Also, bone metastases of prostate cancer tend to be radiosensitive, which allows a higher proportion of nonsurgical treatment. In case of a pathologic fracture, healing rates are higher than for most other metastatic carcinomas (29). Treatment with Radium-223, a calcium-mimetic and alpha-emitter that selectively binds to areas of increased bone turnover, results in significantly prolonged OS in patients who had castration-resistant prostate cancer and bone metastases (30).

## LESSONS LEARNED FROM EXPERIMENTAL IMAGING

Quantitative imaging of bone metastases beyond morphology has been studied in preclinical studies on the functional and molecular level using MRI and PET. In these studies, quantitative biomarkers in skeletal lesions were assessed and validated with the underlying histology. Thereby, DCE-MRI parameters in bone metastatic lesions from breast cancer associated with blood volume and vessel permeability were correlated with vessel maturity, while the apparent diffusion coefficient (ADC) from DWI was associated with tumor cellularity as assessed by cell nuclei staining (31). Treatment monitoring in an animal model of osteolytic breast cancer could be performed reliably using DCE-MRI and $^{18}$F-FDG PET, while therapy response could be detected through functional and metabolic techniques earlier than through morphological imaging (32, 33). Integration of parameters from DCE-MRI and $^{18}$F-FDG PET by machine learning algorithms enabled the detection of early pathologic processes in the bone marrow preceding morphologic changes in bone structure (34). Thus, parameters from functional and metabolic MR and PET imaging are powerful tools to quantify pathophysiologic processes during colonization of bone marrow and to determine response to treatment of skeletal metastasis.

On the molecular level, PET is the method of choice to determine molecular structures expressed in bone metastases, such as integrins alpha$_v$beta$_{3/5}$ or the chemokine receptor CXCR4 (35, 36). Although a major limitation of MRI is the lack of sensitivity when compared to PET, a strategy of signal amplification using a pair of enzymes and an appropriate reducing substrate was presented recently to non-invasively assess epidermal growth factor receptor (EGFR) expression in MRI (37). Besides MRI and PET, other imaging modalities may also be used to determine molecular information in bone metastases, such as ultrasound with its high spatial resolution and unique contrast characteristics of gas-filled microbubbles for enabling the assessment of intra-vascular targets such as vascular endothelial growth factor receptor-2 (VEGFR-2) expressed in bone metastases (38). Thus, molecular imaging strategies for molecular characterization of skeletal lesions have been developed for PET but also for MRI and ultrasound, which are suitable for clinical translation in the near future.

## MAGNETIC RESONANCE IMAGING (MRI)

### From Axial Spine-MRI to Whole Body-MRI With Diffusion-Weighted Imaging

Since the early 1990s, bone marrow MRI has been developed to overcome the limitations of bone scintigraphy and CT for the assessment of bone metastatic disease, showing an unparalleled sensitivity to the replacement of the bone marrow by neoplastic cells (39, 40).

Axial skeleton MRI (AS-MRI) examinations was first developed as a tool used for the detection of bone marrow replacement by neoplastic foci and their quantification (40, 41). Coverage of the "axial skeleton", i.e. the whole spine, bony pelvis and proximal femurs, already probes more than 80% of the red marrow containing areas where metastatic disease is observed, and has limited risk to miss isolated peripheral metastatic disease (39, 42).

Whole body MRI (WB-MRI) was later developed for a global assessment of skeletal metastases and for a one-step "all-organ" approach of metastatic disease. The morphologic T1, fat saturated T2/STIR sequences were first used, and were later complemented with functional DWI sequences (42). The "fluid sensitive-fat saturated" T2-like sequences are now preferably acquired using the Dixon method, that not only provides fat-saturated T2 or STIR equivalent "water only" images, but also "fat only" images providing T1-like information and highly sensitive detection of focal lesions on a background of fatty marrow, questioning the residual need for T1 images (43). This T2 Dixon approach can now be extended to whole body examinations: using T2 Dixon sequences as an alternative to the addition of T1 and STIR drastically decreases the acquisition times of anatomical WB-MRI studies (44). Additionally, the Dixon technique offers the possibility to calculate the marrow fat fraction (FF) and generate fat fraction maps. This quantitative approach is gaining interest along with ADC measurements as a biomarker for response evaluation. Indeed, the fat proportion is expected to increase in focal and diffuse marrow infiltration in response to treatment (45).

### Principles, Advantages and Weaknesses

Classic morphologic MRI sequences detect metastases based on the decrease in normal marrow components, mainly fat cell, and on their replacement by neoplastic cells which may present different biochemical composition properties and variable influence on the adjacent bone structure (46).

DWI sequences detect metastatic foci based on the alteration of the movements of water molecules through tissues. In the bone marrow, early infiltration by neoplastic cells is responsible for a decrease in the free movements of water and ADC (47). DWI sequences provide a functional dimension to MRI examinations, as diffusion parameters mainly probe membrane

integrity, cell viability and tissue density, and allow a quantitative approach of these parameters. It also largely contributes to the detection and response evaluation in extraskeletal organs involved by the metastatic disease (11, 48, 49).

The detection of neoplastic tissue using MRI does not rely on activation of osteoblasts/clasts and subsequent sclerosis/lysis developed on bone trabeculae, which causes delay in the diagnosis of bone infiltration by radiographs, CT and bone scintigraphy. Unlike PET, MRI does not rely on the avidity of the tumoral tissue for a given radioactive tracer, which largely varies according the primary cancer and also according to the disease stage in the same cancer (50). This provides a "universal" dimension to MRI for the detection and follow-up of metastatic disease.

A major strength of MRI is the detailed morphologic analysis of bones, which allows distinction of benign versus malignant fractures, assessment of extraosseous spread and (sometimes preclinical) impingement on neurologic structures, and monitoring of these complications after initiation of targeted or systemic treatment (51).

As main weaknesses, some benign bone lesions may mimic neoplastic foci and should be identified based on the correlation of DWI and morphologic sequences and on ADC measurements (52, 53). In late stages of the disease, treated lesions and scar tissue within the bone marrow may complicate the detection and size measurements of active metastases, especially on morphologic sequences. DWI sequences and ADC maps then become cardinal for response assessment (54–56).

Another potential limitation of MRI is a benign increase in marrow cellularity of the red bone marrow during the treatment, in response to various factors among which are marrow stimulating drugs, potentially resulting in a diffuse "pseudoprogression" (57). This can be prevented by avoiding the use of MRI during and shortly after the use of these drugs.

## Measurement of Response

Bone marrow MRI is currently used daily in clinical practice and clinical trials to assess the response to treatment of bone only and bone predominant metastatic disease, using several approaches with different complexity (18, 58). Recommendations for a homogenization of MRI image acquisitions and generalizable response criteria have been developed (55). The harmonization of quantitative DWI acquisitions and ADC calculations has been addressed by the United Kingdom Quantitative WB-DWI Technical Workgroup (59).

## Size and Number

Metastatic disease to the bone marrow may present as a focal or a diffuse pattern. Evolution from a normal appearing marrow to a focal or diffuse pattern, increase in number and size of focal lesions will indicate disease progression (60). A decrease in focal lesion number and size, return from diffuse or focal patterns of marrow infiltration to a normal marrow appearance will indicate response (**Figures 1**, **2**).

RECIST-like criteria can be transposed to bone marrow metastases. Simple size measurements of bone metastases on morphologic sequences in (a limited sample of) bone metastases

allows objective assessment of response, especially in early disease. This approach can be used on morphologic sequences and on high b value DWI sequences. In prostate cancer, this approach more than doubles the proportion of patients with measurable metastatic disease, previously limited to those patients with quantifiable abdominal lymph nodes (41).

## Non-Quantitative Features

Additional "qualitative" signs may be used for response assessment on MR images. The progressive appearance of a "fatty halo" of high signal on T1-weighted images at the periphery of regressing focal lesions indicates responsive disease (60). Conversely, the disappearance of a peripheral "cellular" of high signal intensity on T2-weighted images, representing active or aggressive disease, also represents an early sign of response, whereas its re-appearance suggests disease relapse. The appearance of malignant vertebral compression fractures, and appearance/progression of extraosseous/epidural spread unambiguously indicate progressive disease (60).

## Quantitative Functional and Multiparametric Approaches

The quantitative approach can be directed either to individual lesions or to the whole-skeleton using ADC measurements and mapping derived from DWI sequences and fat fraction (FF) measurements derived from Dixon acquisitions. This approach becomes cardinal in advanced metastatic disease where previously treated lesions and scar tissue complicate the size measurements of active lesions on morphologic sequences.

Response to treatment is associated with an early increase in ADC values within individual lesions (61). At a later stage, responsive bone metastases present a decrease in ADC values together with a decreased signal on high b-value images due to recolonization by normal bone marrow. A sharper decrease in signal intensity and ADC is related to the sclerotic transformation of treated lesions, which is also observed on anatomic sequences. A total diffusion volume can be derived from WB DWI sequences for a global quantification of the metastatic burden and its follow-up under treatment (62, 63). The FF presents an early increase in focal and diffuse metastatic infiltration in response to treatment) (45).

Multiparametric MRI by definition combines anatomical and at least two functional sequences. The multiparametric WB-MRI approach used for the quantitative evaluation of bone lesions combines anatomical T1 and STIR sequences (potentially replaced by single T2 Dixon acquistions), FF measurements, and functional DWI sequences along with ADC maps.

The METastasis Reporting and Data System for prostate cancer (MET-RADS-P) guidelines were designed in prostate cancer, in an international initiative to standardize WB-MRI protocols and most importantly to provide multiparametric response evaluation criteria for bone, node, and visceral lesions (55). These criteria combine quantitative approaches of ADC and FF within bone marrow metastases, RECIST-like size criteria transposed to bone lesions, and RECIST criteria for node and visceral lesions follow-up. They allow categorization of the disease response or progression on a 5-point Likert scale. The method also offers the possibility to record the heterogeneity of response within metastases and
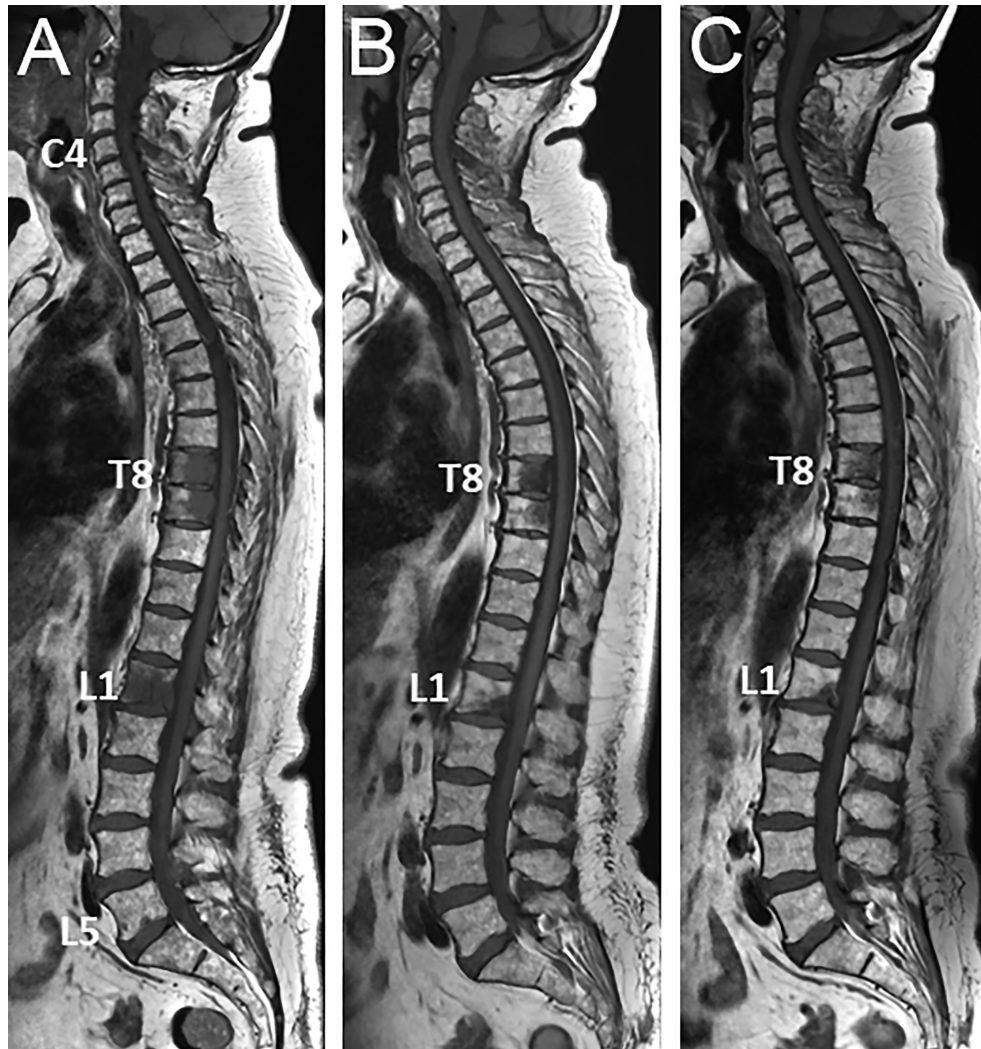
**FIGURE 1** | 53 year-old woman with newly diagnosed metastatic breast cancer (grade II ductal carcinoma, ER 8, PR8, KI 67 5%, HER2 neu 2+): spinal MRI findings at diagnosis of bone metastases and during treatment. **(A)** Baseline sagittal T1-weighted MR image of the whole spine shows multiple foci of low signal intensity of the bone marrow, typical for bone metastases (posterior arch of C4, vertebral bodies of T8, T9, L1, tiny foci in L5). **(B)** Corresponding MR image obtained 2-m later after combined treatment including a selective estrogen receptor degrader (SERD) and palbociclib shows significant decrease in size of all lesions, and disappearance of the small L5 foci. **(C)** Follow-up MR image obtained 2-m later shows further decrease in size of all lesions, with measurable decrease in lesion size and reappearance of fatty marrow at the periphery and within the lesions, again indicating frank response to treatment.

categorizes the response as "discordant" if some bone lesions or soft-tissue are progressing, while others are stable or are responding, and vice-versa. The reproducibility of the technique as well as its use by readers with various experience have been validated (64). The same criteria may be transposed for WB-MRI studies performed for lesion follow-up and response assessment in bone-only or bone-predominant metastatic disease from other primary cancers.

## Target Cancers

The objective parameters extracted from AS-MRI and WB-MRI/ DWI are increasingly used to assess response of bone metastases to treatment in a large number of primary cancers.

In prostate cancer, AS-MRI and later WB-MRI were introduced after demonstration of their superiority to bone scintigraphy for detection of bone metastases and for a one step staging of bone and lymph node involvement (40, 65, 66). The current roles of WB-MRI to assess metastatic disease have been recently illustrated and compared to other techniques (44). PSMA-PET/CT is most likely the current most sensitive technique for the detection of low volume metastatic disease and for therapeutic decision (curative versus systemic treatment) in newly diagnosed prostate cancer and at the biochemical recurrence stage. WB-MRI is an optimal non-irradiating alternative for polymetastatic disease detection and follow-up
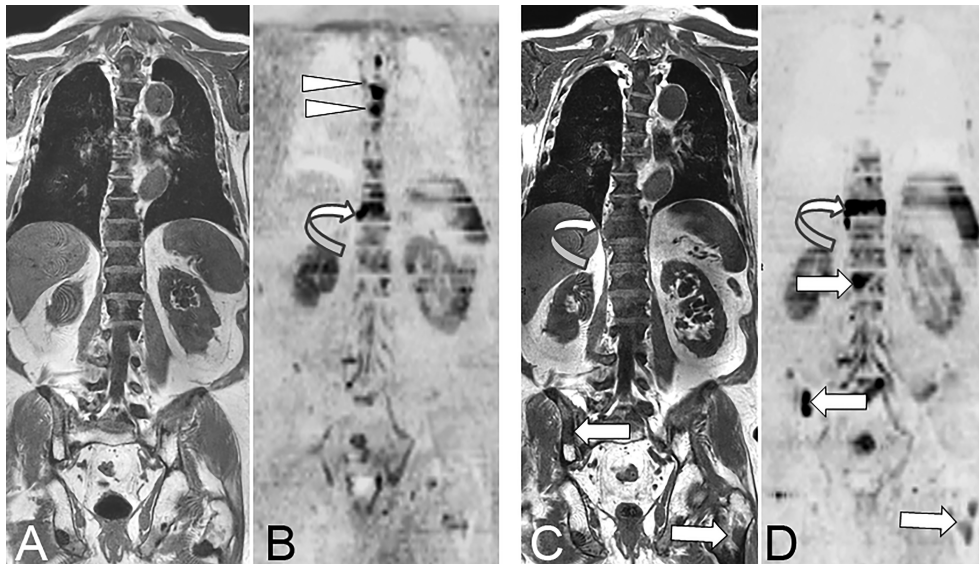
**FIGURE 2** | 73 year-old man with advanced prostate cancer. Comparison of pre- and post-treatment (enzalutamide) WB-MRI/DWI. Baseline coronal T1-weighted MR image **(A)** shows diffuse bone marrow infiltration within the spine, responsible for diffuse low signal intensity of the bone marrow, and related to advanced metastatic disease after several lines of treatment. The pelvic bones show higher signal of the bone marrow indicating a fatty content due to previous irradiation. Several focal lesions of low signal intensity are visible within the pelvis and left proximal femur. Baseline DWI MR image **(B**; B = 1000 s/mm$^2$, inverted grey scale) shows high signal intensity foci typical for active bone metastases within the T4, T5 (arrowheads) and T10 (curved arrow) vertebrae. Follow-up T1-weighted MR image **(C)** shows no evident change of the spinal bone marrow, but increase in the right paraspinal extension of the T10 metastasis (curved arrow), and a new lesion within the right posterior iliac crest (arrow). Follow-up DWI MR image **(D)** shows disappearance of the midthoracic vertebral lesions, but increase in size and right paraspinal extension of the T10 vertebral lesion (curved arrow), and appearance of new lesions within the L1 vertebral body, the right iliac crest and left proximal femur (arrows). The observation of concurrent signs of disease response and progression is frequent, especially in advanced stages of metastatic cancer.

under systemic treatment (**Figure 2**). WB-MRI might become the first choice in advanced disease, castration-resistant prostate cancer (CRPC), as PSMA-PET/CT might be confounded by androgen blockade (AB) treatments which induce short term upregulation of PSMA expression and long term downregulation of this expression, limiting the possibility of following metastatic prostate cancer lesions at this stage (67, 68).

In breast cancer, AS-MRI and WB-MRI were also introduced to overcome the limitations of bone scintigraphy (BS) and CT for the detection of bone metastases and evaluation of their response to treatment (**Figure 1A**) (54, 69, 70). WB-MRI progressively becomes a key imaging modality for the evaluation of response in bone only/predominant metastatic breast cancer for the follow-up of treatment response (71). In patients with advanced breast cancer treated with systemic treatment of metastatic disease and followed-up with WB-MRI in addition to other imaging modalities (CT, BS, TAP-CT or PET/CT), WB-MRI discloses progressive disease earlier than the reference examination and provides decisive information for changes in treatment in more than 50% of patients (72–74). Of note, WB-MRI shows a frequent discrepancy between response as assessed locally within the primary cancer and within metastases, and disease progression is identified earlier in distant disease compared to local disease assessment (75).

There is a consistently increasing number of indications of WB-MRI for bone and visceral metastases detection in various primary cancers, sometimes relying on the design of disease- or patient-tailored MRI studies (coverage of lung, liver, and brain, with specific sequences according to primary cancer). WB-MRI can for example be proposed in this indication in lung, thyroid, kidney and colorectal cancers, in melanoma, myxoid liposarcoma, Ewing sarcoma or osteosarcoma. The detection of bone metastases using the same technique substantiates its use for the subsequent evaluation of the response of bone lesions to treatment (76).

# POSITRON EMISSION TOMOGRAPHY (PET)

## Quantitative Assessment of Bone Metastases on PET

Traditionally, PET is used for staging of many cancer types because of its high sensitivity for visual detection of metastatic disease, typically using $^{18}$F-FDG as radiotracer. In 2009, novel qualitative and quantitative approaches to metabolic tumor response assessment, solely applicable for $^{18}$F-FDG PET, were proposed (77). The purpose was to overcome the limitations of morphologic imaging alone-based criteria (e.g. RECIST, RECIST1.1) and to capitalize the benefit of using newer cancer therapies. The framework for PET Response Criteria in Solid Tumors (PERCIST), version 1.0, was meant to serve as an example for use in clinical trials and in structured quantitative clinical reporting (77).

In current practice, however, the quantitative nature of PET is often unexploited. Especially in the case of bone metastases that are deemed non-measurable by RECIST 1.1, quantification of radiotracer uptake might prove crucial for assessing bone disease through changes in the viability or molecular processes of tumor cells instead of lesion morphology. A further advantage is that quantitative PET assessment can be performed on a per-lesion basis, as well as on a whole-body level.

Parameters that can be extracted from routinely acquired static whole-body PET images have been validated for many tracers in different cancer types (78–82). In general, these parameters can be divided into those based on (83): i) tracer uptake intensity (e.g. standardized uptake values, SUVs), ii) metabolically active tumor volumes (MATV), and iii) a combination of both, representing the total tracer uptake in a tumor. Typical SUV metrics are the mean uptake ($SUV_{mean}$), the maximum uptake (highest voxel value; $SUV_{max}$), or the peak uptake (highest average value of a $1cm^3$ sphere; $SUV_{peak}$) within an identified lesion. Depending on specific radiotracer kinetics, uptake may need to be normalized to background activity in e.g. liver or blood (81). Metrics combining lesion volume and tracer uptake, such as total lesion glycolysis (TLG) for $^{18}$F-FDG, seem especially promising for objective longitudinal assessment of bone metastases load, as they provide information on the total amount of viable tumor tissue within a bone lesion both on an individual lesion and patient-basis (84, 85).

## Target Cancers
### Prostate Cancer
In metastatic prostate cancer, osteoblastic or mixed bone lesions with minor soft tissue component are frequently observed,

challenging accurate RECIST1.1-based follow-up for these patients. With the recent introduction of several PET-tracers targeting the PSMA (**Figure 3**), detection of prostate cancer lesions has significantly improved (86, 87). In 2018, guidelines for standardized interpretation of PSMA PET images (PROMISE) were proposed (88). Quantitative parameters for evaluation of treatment response using PSMA PET/CT, besides well-known maximum standardized uptake values ($SUV_{max}$), have been proposed including PSMA tumor volume (PSMA-TV) and total lesion PSMA expression (TL-PSMA) (85). Initial studies evaluating metrics such as PSMA-TV and TL-PSMA for metabolic response assessment have shown promising results, some of these through a 'PSMA-modified' RECIST or PERCIST classification system. Importantly, several studies reported an association of these PSMA PET parameters with overall survival (OS) during treatment with radioligand therapy (RLT) with $^{177}$Lu-PSMA (89–91). A recent systematic review summarized the available evidence for using quantitative PSMA parameters versus serum PSA in assessing response for castration-resistant prostate cancer (92).

In parallel to ER-targeted PET imaging in breast cancer with $16\alpha$-$^{18}$F-fluoro-$17\beta$-estradiol ([$^{18}$F]FES), androgen receptor (AR)-targeted PET imaging in prostate cancer is possible using $^{18}$F-fluorodihydrotestosterone ($^{18}$F-FDHT; **Figure 3**), which binds the intracellular AR in prostate cancer cells (93). This enables quantitative assessment of AR-expression in bone metastases, both for response monitoring and prognostic purposes (80, 93, 94). $^{18}$F-FDHT cannot be used during treatment with drugs that directly block the AR (95, 96). For PSMA-ligands and $^{18}$F-FDHT, technical validation studies assessing tracer pharmacokinetics and repeatability have been
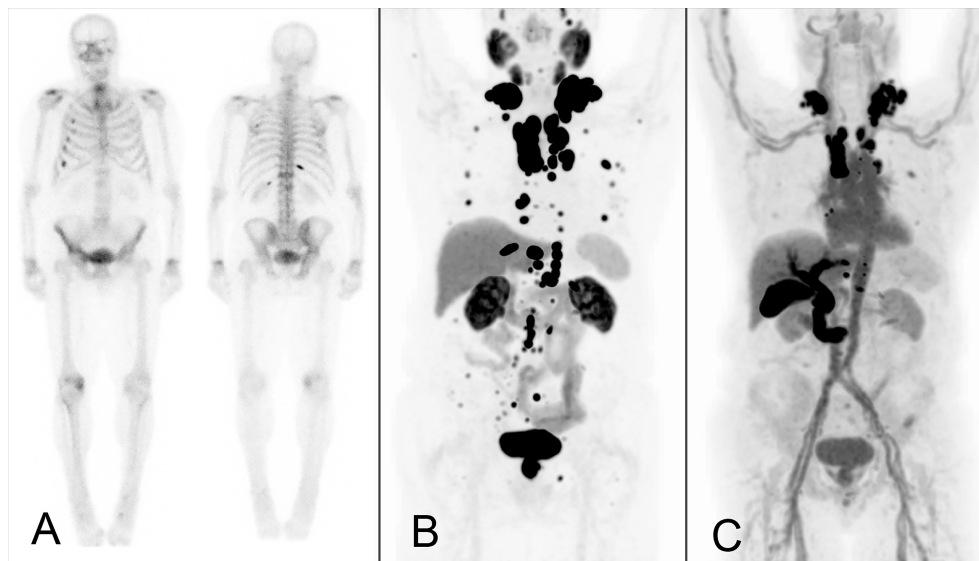


**FIGURE 3** | Example of a male patient with bone and lymph node metastases from castration-resistant prostate cancer who underwent bone scintigraphy **(A)** and $^{18}$F-DCFPyL **(B)** and $^{18}$F-FDHT PET **(C)** for research purposes. Bone scintigraphy demonstrated several rib metastases. A large number of additional (measurable) bone metastases were observed on $^{18}$F-DCFPyl PET, with additional lymph node metastases detected on-par. Disconcordant AR-expression visualized on $^{18}$F-FDHT PET.

performed, enabling their clinical use in (quantitative) response assessment (80, 81, 94, 97, 98).

The PET response assessment approach for bone metastases in prostate cancer using PSMA PET may be extended to other targeted PET tracers, such as 18F-NaF and 18F-FDHT in prostate cancer, 18F-FES in breast cancer and 18F-FDG and 68Ga-fibroblast activation protein inhibitors (68Ga-FAPI) in a multitude of cancer types (94, 99–102).

## Lung Cancer

The skeleton is the most common site of distant metastasis in lung cancer. Approximately 30% to 40% of the patients with advanced cancer will develop bone metastases, which represent 10% of disease recurrence even in early stage operable lung cancer (15, 103, 104). 18F-FDG PET/CT plays a key role in the diagnostic work-up of lung cancer, being fundamental especially at diagnosis and during staging/restaging (105). Consequently, all clinical guidelines support the use of the modality for the assessment of advanced disease (106–110), given the high diagnostic accuracy in depicting distant metastases for which 18F-FDG PET/CT results superior to other conventional imaging (111–115).

Recent meta-analysis data comparing [18F]FDG PET/CT with WB-MRI show similar performances for staging NSCLC, i.e. area under the curve (AUC) 0.95 for PET versus 0.93 for MRI (116). The performance was also similar in case of SCLC patients (117). When considering only bone metastases, dedicated meta-analyses in lung cancer have proven PET/CT is superior to other modalities, with a pooled sensitivity for [18F]FDG PET/CT, MRI and bone scintigraphy (BS) of 92%, 77% and 86%, respectively, associated to a pooled specificity of 98%, 92% and 88%, respectively (118). Depending on cancer type, there is also an associated impact in patient management that ranges from 12%-40% of the cases (105, 111, 115, 118).

## Breast Cancer

The use of 18F-FDG PET/CT in breast cancer faces more conflicting indications based on major clinical guidelines (111, 119, 120). While staging in advanced or suspicious metastatic breast cancer is widely supported, initial preoperative staging is regarded of limited value. Still, the results of a recent meta-analysis in 4276 patients prove that the use of 18F-FDG PET for initial evaluation of breast cancer leads to a change in staging and management in 25% and 18% of patients, respectively (121). With younger age, clinical stage III to IV and histologic grade II to III were significantly associated with a greater proportion of changes. These results are most likely attributable to the superior diagnostic accuracy of 18F-FDG PET/CT compared with other modalities (122, 123). In particular, the pooled sensitivity and specificity of whole-body 18F-FDG PET and PET/CT are reported to be 99% and 95%, respectively, compared to 57% and 88% for conventional imaging studies (8, 124).

Approximately 70%–80% of breast cancers express hormone receptors (HR), i.e. ERα and/or progesterone receptors (PR) (125). Thanks to the use of [18F]FES PET, breast cancer metastases can be characterized non-invasively also for ER status reaching a pooled sensitivity and specificity of 78% and

98%, respectively (126, 127). The information obtained by [18F] FES PET can be used also to predict the response to hormonal therapy in patients with locally advanced or metastatic breast cancer. For this purpose, SUV cut-off values can be applied, for example 1.5 and 2.0, demonstrating pooled sensitivities and specificities for response prediction of 63.9% vs. 66.7%, and 28.6% vs. 62.1%, respectively (127). In newly diagnosed ER-positive breast cancer, moreover, [18F]FES PET shows a sensitivity of 90.8% versus 82.8% for 18F-FDG PET/CT, thus potentially leading to a change in patient management in 26.3% of the cases (128).

Besides overexpression of hormone receptors, a proportion of breast cancer tumors is known to show expression of human epidermal growth factor receptors 2 (HER2) (129). In recent years, whole body HER2-targeted PET imaging has proven to be a valuable tool, both for the identification of patients suitable for anti-HER2 therapy and monitoring therapeutic efficacy (130–135). HERs can be targeted by several inhibitors that directly block the receptors on HER-expressing tumor cells or interfere with their signaling pathways (135). HER2-targeted PET imaging with 64Cu- or 89Zr-labeled antibodies is effective but typically requires late time points acquisitions due to the antibody and radio-isotope properties (132, 133). 68Ga-labeled affibody molecules targeting HER2 allow for routine same-day PET imaging, thereby improving the clinical utility of HER2-targeted imaging, and have yielded promising initial results (130, 131, 136). More clinical data on the use of HER2-targeted molecular imaging in breast cancer patients is required before future clinical use.

## Challenges and Opportunities in PET

Absolute measurements of tumor lesion PET metrics are inherently dependent on the method used for tumor delineation (137). Several segmentation methods have been proposed, most semi-automatic and relatively easy to apply, requiring a good repeatability and reproducibility basis in order to detect small changes during response monitoring (138, 139). Software packages are often vendor-supplied and differences between several methods have been well evaluated (139–141).

Evaluating longitudinal changes in tracer uptake on PET typically requires patients to be scanned on preferably the same PET/CT system using the same image reconstruction protocol (142–144). Still, in PET the technical uncertainties can be easily mitigated by harmonization of PET/CT system performances between and within clinical centers. The latter is achieved by the EARL accreditation program showing that harmonization is feasible and is a prerequisite for a high reproducibility of quantitative reads (145, 146).

Besides technical challenges, biological aspects need to be considered when using PET for measuring bone lesions and response to treatment in a clinical setting. The optimal timing of disease assessment will depend on the specific treatment type a patient is receiving, such as radiotherapy, RLT, chemotherapy, or other targeted drugs. For example, systemic cytotoxic or antihormonal treatments may elicit so called 'flare' phenomena, potentially precluding the use of PET early during treatment follow-

up (147–150). This can be avoided by adhering to clinical guidelines and not performing PET too soon after treatment initiation.

Recent and ongoing technical advances have given rise to several new opportunities in PET imaging. PET initially was a stand-alone modality, but has moved on to become a hybrid imaging modality (with CT and MRI). Even more recently, the novel 'total body' PET systems have become available (151, 152). These total body (or 'long axial field of view') PET systems can be used to perform PET imaging in a single field-of-view instead of multiple bed positions, with typical FOV from skull apex to mid-thighs (151–153). Not only does this severely shorten the required acquisition time (a large benefit for patients with often painful bone metastases), but this is also accompanied by a large increase in system sensitivity which is expected to improve lesion detection rates (153). Moreover, the total body PET might enable quantitative parameters incorporating radiotracer dynamics, such as whole body Patlak (154), to be extracted and parametric images to be generated.

Advances in computer science have made the routine use of artificial intelligence (AI) in medical imaging analysis possible (155). A common application of AI in PET lies in the analysis and modeling of radiomics features. Radiomics pertain to large volumes of data on tumor shape, size, metabolism and texture that can be extracted from PET-positive lesions, providing an image-based tumor phenotype (155–157). The Imaging Biomarker Standardization Initiative has harmonized performance of radiomics software packages to allow for its robust and reproducible use (157). Recently, consensus recommendations for considerations on the use of radiomics (both PET, CT, and MRI) in clinical trials have been proposed (158). Deep learning techniques, which do not require extraction of predefined features seem particularly promising for segmentation purposes of PET-avid bone metastases (159, 160). For PSMA PET, a deep learning algorithm for automated analysis of PET images ('aPROMISE') has been developed (161).

## HYBRID IMAGING (PET/MRI)

The unique potential of hybrid imaging, as reviewed by Schmidkonz and colleagues, lies in the assessment of complementary parameters on the morphologic, functional, metabolic and molecular levels of bone metastases from different modalities in a single examination (162). When combining PET with CT in a PET/CT study, the CT component enables assessment of bone morphology and osseous destruction, while MRI in a PET/MRI hybrid study will offer superior soft tissue contrast. Due to the (still) novelty and increased cost and complexity of PET/MRI, this technique currently is primarily compared to PET/CT for assessing the respective potential of these two imaging approaches for evaluating bone metastases.

When comparing the performance of $^{18}$F-FDG PET/CT with $^{18}$F-FDG PET/MRI for the assessment of malignant bone lesions, the overall performance of PET/MRI has been found to be equivalent to PET/CT for the detection and characterization of bone lesions when these hybrid techniques were performed sequentially (163). However, in PET/MRI, lesion delineation and allocation of PET-positive findings were found to be superior to PET/CT (163). Samarin and colleagues reported

similar results from a comparison of $^{18}$F-FDG PET/CT with $^{18}$F-FDG PET/MRI in 24 patients with bone metastases from different primary tumors (164). The overall detection rate was not significantly different between PET/CT and PET/MRI, but the latter provided higher reader confidence and improved conspicuity as compared with PET/CT (164).

In a prospective comparison of the diagnostic accuracy of $^{18}$F-FDG PET/MRI and CT, PET/MRI was significantly better than CT for the detection of bone metastases in patients with newly diagnosed breast cancer (165). Also, in a particular series of 109 breast cancer patients, PET/MR demonstrated an improved sensitivity over $^{18}$F-FDG PET/CT alone, where the sensitivity of PET/MR and PET/CT were 96% and 85%, respectively (166). In men with biochemical recurrence of prostate cancer following curative therapy, $^{68}$Ga-PSMA-11 PET/MRI demonstrated a high detection rate especially for recurrent disease with low PSA values, but included all sites of local or distant recurrence including lymph nodes and bone (167). In 26 patients with prostate cancer, $^{68}$Ga-PSMA-11 PET/MRI and PET/CT performed equally regarding the PET component for detection of bone metastases, while two PET-positive skeletal metastases could be confirmed on contrast MRI, but not on CT (168).

An interesting approach for patients with both osteolytic and osteoblastic metastases from breast or prostate cancer was proposed by Sonni and colleagues (169). Combining $^{18}$F-FDG and Na$^{18}$F in PET/MRI was superior for the detection of skeletal metastases as compared to whole-body bone scintigraphy (169). This approach includes in an innovative manner both a radiotracer (Na$^{18}$F) for the assessment of primarily osteoblastic activity in osteoblastic lesions, and another tracer ($^{18}$F-FDG) for assessing increased glucose metabolism in the soft tissue component of predominantly osteolytic metastases. Based on the data and results referenced above, PET/MRI appears rather superior to PET/CT for the detection of metastatic bone lesions, but it lacks the morphologic information of bone and osteoblastic bone formation derived from the CT component, which might be mitigated some through innovative approaches as reported by Sonni and co-workers.

## BEYOND RECIST AND PERCIST

In 2009, the PET Response Criteria in Solid Tumors (PERCIST) were introduced for $^{18}$F-FDG PET (77). Later on, along with the detailed describing of the $^{18}$F-FDG PET requirements to allow quantitative expression of the changes in PET measurements and assessment of overall treatment response, a Simplified Guide to PERCIST 1.0 was published (170). The PERCIST criteria enable avid bone target lesions to be selected based on their metabolic activity, and response to be measured objectively based on the changes in metabolic activity even in the absence of an evident anatomic change. PERCIST, however, only considers the change in uptake of a single target lesion when assessing response, which is the lesion with the highest SUV$_{peak}$ value normalized for lean body mass (SUL$_{peak}$). New lesions, in the bone or elsewhere, result in progressive disease by definition. The target lesion may or may not

be within the bone, but all bone lesions have to be considered in the selection of target lesion. Of note, there is no impact of changes in volume of lesions, only the uptake concentration is considered. Compared to RECIST, PERCIST represents a major step forward for bone assessment as it considers bone lesions equally to any other lesions anywhere else in the body.

The PERCIST approach focusses on the remaining hottest lesion and has similarities with the therapy response criteria for lymphoma, where the most active remaining lesions play a dominant role (171). This "hottest lesion" centric approach is well tailored to therapies with curative intent, but it might miss a beneficial effect in non-curative therapies, where tumor control and tumor bulk reduction are clinically relevant achievements. A recent approach in image analysis is about abandoning the selection of target lesions, as determined on baseline or post-therapy scans, and aiming to take the entire tumor bulk in consideration. The high contrast of modern oncological PET tracers [e.g., $^{18}$F-FDG, somatostatin receptor (SSTR) and PSMA ligands, $^{18}$F-DOPA, $^{18}$F-MFBG (172)] permits straightforward three-dimensional segmentation of lesions; by segmenting all lesions, the total tumor burden can then be obtained. This type of analysis does not distinguish between bone and non-bone lesions and thus puts bone metastases on par with other metastases.

There is evidence that the baseline metabolic tumor volume (MTV) is an important prognostic factor, e.g. in NHL, NSCLC and multiple myeloma, as well as in prostate cancer patients treated with the bone-seeking agent radium-223 (173–176). Furthermore, MTV can be combined with metrics of tumor distance within a patient to not only represent volume, but also dissemination for better reflecting prognosis, as shown in NHL (175). Evaluation of the changes in MTV and/or TLG have been shown to outperform PERCIST-based approaches in tumors with frequent bone lesions, such as Ewing sarcoma and osteosarcoma (177–180). Volumetric determination on PET is not hampered by bone/soft tissue interfaces, taking the total tumor burden into account in combination with the metabolic activity. Total tumor burden analyses can be combined with specific organ segmentation either based on PET or CT, e.g. for spleen (for lymphoma) and bone, to generate organ-specific tumor burden (181, 182). Furthermore, the segmentation leading to total tumor burden or organ-specific tumor burden can also be used as a mask to determine specific radiomic features, which can provide even more information for response evaluation (180).

Although promising, some challenges remain to the application of tumoral volumes for routine therapy response monitoring: (i) lack of standardization of uptake thresholds for PET-positive tumor delineation; (ii) still too time consuming for clinical routine; (iii) no prospectively defined response criteria. Especially regarding i and ii, it is expected that advances in tumor segmentation, e.g. with further automatization of the segmentation process and contributions from deep learning-based AI algorithms, will increase the robustness, accuracy and feasibility of total tumor segmentation to routine clinical practice levels (141, 183). Based on results from current and ongoing studies, automated tumor segmentation should actually be one of the key expected improvements from AI applications to PET (and other) imaging (160).

At any rate, similar analyses can be developed for non-FDG tracers, e.g. with SSTR ligands in neuroendocrine tumors and PSMA ligands in prostate cancer, Na$^{18}$F in breast and prostate cancer patients (184–187). For evaluation of response on PSMA PET, consensus criteria have recently been proposed with specific cut-off values for both uptake and volume (188).

A specific case of total tumor burden imaging that is worthy to mention is the use of the bone scan index (BSI), which is a metric based on 2D planar bone scintigraphy that reflects the fraction of bone showing increased turnover due to metastatic invasion (189). It has been proposed two decades ago as a metric for tumor burden and response assessment in metastatic prostate cancer (189). Changes in BSI under treatment have been shown to correlate with OS in patients with metastatic castrate-resistant prostate cancer (mCRPC) treated with a range of therapies (190). This has been corroborated in multicenter trials in mCRCP patients treated with abiraterone acetate and with radium-223 dichloride (191, 192). Although promising, it is expected that the shift from 2D to 3D imaging and the increasing use of novel PET tracers that can pick up lesions outside of the bone as well as bone lesions (e.g. PSMA ligands) will eventually displace the currently widespread adoption of BSI for therapy response. Accordingly, similar but PET-based metrics from PSMA and/or Na$^{18}$F PET will likely outperform and replace BSI.

## CONCLUSION

Modern imaging with PET and MRI allows bone metastases to be detected and assessed both before and after therapy, without the drawbacks of X-ray based imaging techniques (e.g. radiographs, CT). These techniques assess bone metastases within the same framework, as metastases in other organs. They further allow total tumor burden to be assessed within a single imaging session, and also the development of response criteria that include the bone, thus filling a critical gap in the RECIST1.1 framework. The EORTC, PERCIST and recent PSMA PET criteria are examples of criteria that take bone metastases in consideration, on-par with extra-osseous lesions. PET and/or MRI can detect and characterize bone metastases of various types (e.g. lytic, sclerotic, radio-occult or mixed) independently from the bone density changes. In contrast with CT, they are not affected by changes in bone mineralization induced by the tumor(s), and are not dependent on soft-tissue components (as required by RECIST 1.1). Whole body MRI including modern techniques such as DWI, DCE-MRI and mpMRI can provide both detailed information on anatomical structures as well as functional information on individual lesions and whole body tumor burden. Modern PET imaging is performed on hybrid cameras, with CT (from PET/CT) allowing assessment of the bone mineral content (including fractures), while MRI (from PET/MRI) can more often clarify a correlate for the lesions observed on PET. Total tumor burden, incorporating bone metastases on par with other metastases, is an attractive approach to be applied in most PET tracers. While advances in algorithms and deep-learning contributions are expected to permit the determination of total tumor burden

metrics in actual clinical routine before and after therapy, response criteria through total tumor burden assessment are currently developed, taking into consideration the tracer, therapy and underlying cancer type.

## AUTHOR CONTRIBUTIONS

## REFERENCES

1. Moreira CA, Dempster DW, Baron R. Anatomy and Ultrastructure of Bone - Histogenesis, Growth and Remodeling. In: Feingold KR, Anawalt B, Boyce A, Chrousos G, de Herder WW, Dhatariya K, et al, editors. *Endotext*. South Dartmouth (MA): MDText.com, Inc. (2000).

2. Guise TA, Mohammad KS, Clines G, Stebbins EG, Wong DH, Higgins LS, et al. Basic Mechanisms Responsible for Osteolytic and Osteoblastic Bone Metastases. *Clin Cancer Res* (2006) 12(20 Pt 2):6213s–6s. doi: 10.1158/1078-0432.CCR-06-1007

3. Clezardin P. Pathophysiology of Bone Metastases From Solid Malignancies. *Joint Bone Spine* (2017) 84(6):677–84. doi: 10.1016/j.jbspin.2017.05.006

4. Messiou C, Cook G, Reid AH, Attard G, Dearnaley D, de Bono JS, et al. The CT Flare Response of Metastatic Bone Disease in Prostate Cancer. *Acta Radiol* (2011) 52(5):557–61. doi: 10.1258/ar.2011.100342

5. Therasse P, Arbuck SG, Eisenhauer EA, Wanders J, Kaplan RS, Rubinstein L, et al. New Guidelines to Evaluate the Response to Treatment in Solid Tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada. *J Natl Cancer Inst* (2000) 92(3):205–16. doi: 10.1093/jnci/92.3.205

6. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New Response Evaluation Criteria in Solid Tumors: Revised RECIST Guideline (Version 1.1). *Eur J Cancer* (2009) 45(2):228–47. doi: 10.1016/j.ejca.2008.10.026

7. Spiro SG, Buscombe J, Cook G, Eisen T, Gleeson F, O'Brien M, et al. Ensuring the Right PET Scan for the Right Patient. *Lung Cancer* (2008) 59(1):48–56. doi: 10.1016/j.lungcan.2007.07.026

8. Paydary K, Seraj SM, Zadeh MZ, Emamzadehfard S, Shamchi SP, Gholami S, et al. The Evolving Role of FDG-PET/CT in the Diagnosis, Staging, and Treatment of Breast Cancer. *Mol Imaging Biol* (2019) 21(1):1–10. doi: 10.1007/s11307-018-1181-3

9. Stecco A, Trisoglio A, Soligo E, Berardo S, Sukhovei L, Carriero A. Whole-Body MRI With Diffusion-Weighted Imaging in Bone Metastases: A Narrative Review. *Diagnostics (Basel)* (2018) 8(3):45. doi: 10.3390/diagnostics8030045

10. Giles SL, Messiou C, Collins DJ, Morgan VA, Simpkin CJ, West S, et al. Whole-Body Diffusion-Weighted MR Imaging for Assessment of Treatment Response in Myeloma. *Radiology* (2014) 271(3):785–94. doi: 10.1148/radiol.13131529

11. Lecouvet FE, Van Nieuwenhove S, Jamar F, Lhommel R, Guermazi A, Pasoglou VP. Whole-Body Mr Imaging: The Novel, "Intrinsically Hybrid," Approach to Metastases, Myeloma, Lymphoma, in Bones and Beyond. *PET Clin* (2018) 13(4):505–22. doi: 10.1016/j.cpet.2018.05.006

12. Hillengass J, Ritsch J, Merz M, Wagner B, Kunz C, Hielscher T, et al. Increased Microcirculation Detected by Dynamic Contrast-Enhanced Magnetic Resonance Imaging Is of Prognostic Significance in Asymptomatic Myeloma. *Br J Haematol* (2016) 174(1):127–35. doi: 10.1111/bjh.14038

13. Merz M, Moehler TM, Ritsch J, Bauerle T, Zechmann CM, Wagner B, et al. Prognostic Significance of Increased Bone Marrow Microcirculation in Newly Diagnosed Multiple Myeloma: Results of a Prospective DCE-MRI Study. *Eur Radiol* (2016) 26(5):1404–11. doi: 10.1007/s00330-015-3928-4

14. Macedo F, Ladeira K, Pinho F, Saraiva N, Bonito N, Pinto L, et al. Bone Metastases: An Overview. *Oncol Rev* (2017) 11(1):321. doi: 10.4081/oncol.2017.321

15. Coleman RE. Metastatic Bone Disease: Clinical Features, Pathophysiology and Treatment Strategies. *Cancer Treat Rev* (2001) 27(3):165–76. doi: 10.1053/ctrv.2000.0210

16. Southby J, Kissin MW, Danks JA, Hayman JA, Moseley JM, Henderson MA, et al. Immunohistochemical Localization of Parathyroid Hormone-Related Protein in Human Breast Cancer. *Cancer Res* (1990) 50(23):7710–6.

17. Keller ET, Zhang J, Cooper CR, Smith PC, McCauley LK, Pienta KJ, et al. Prostate Carcinoma Skeletal Metastases: Cross-Talk Between Tumor and Bone. *Cancer Metastasis Rev* (2001) 20(3-4):333–49. doi: 10.1023/A:1015599831232

18. Cook GJR, Goh V. Molecular Imaging of Bone Metastases and Their Response to Therapy. *J Nucl Med* (2020) 61(6):799–806. doi: 10.2967/jnumed.119.234260

19. Papotti M, Kalebic T, Volante M, Chiusa L, Bacillo E, Cappia S, et al. Bone Sialoprotein is Predictive of Bone Metastases in Resectable Non-Small-Cell Lung Cancer: A Retrospective Case-Control Study. *J Clin Oncol* (2006) 24(30):4818–24. doi: 10.1200/JCO.2006.06.1952

20. Sipkins DA, Wei X, Wu JW, Runnels JM, Cote D, Means TK, et al. *In Vivo* Imaging of Specialized Bone Marrow Endothelial Microdomains for Tumor Engraftment. *Nature* (2005) 435(7044):969–73. doi: 10.1038/nature03703

21. Fornetti J, Welm AL, Stewart SA. Understanding the Bone in Cancer Metastasis. *J Bone Miner Res* (2018) 33(12):2099–113. doi: 10.1002/jbmr.3618

22. Wei S, Li Y, Siegal GP, Hameed O. Breast Carcinomas With Isolated Bone Metastases Have Different Hormone Receptor Expression Profiles Than Those With Metastases to Other Sites or Multiple Organs. *Ann Diagn Pathol* (2011) 15(2):79–83. doi: 10.1016/j.anndiagpath.2010.06.010

23. Farach-Carson MC, Lin SH, Nalty T, Satcher RL. Sex Differences and Bone Metastases of Breast, Lung, and Prostate Cancers: Do Bone Homing Cancers Favor Feminized Bone Marrow? *Front Oncol* (2017) 7:163. doi: 10.3389/fonc.2017.00163

24. Pan H, Gray R, Braybrooke J, Davies C, Taylor C, McGale P, et al. 20-Year Risks of Breast-Cancer Recurrence After Stopping Endocrine Therapy at 5 Years. *N Engl J Med* (2017) 377(19):1836–46. doi: 10.1056/NEJMoa1701830

25. Johnson CN, Gurich RWJr., Pavey GJ, Thompson MJ. Contemporary Management of Appendicular Skeletal Metastasis by Primary Tumor Type. *J Am Acad Orthop Surg* (2019) 27(10):345–55. doi: 10.5435/JAAOS-D-17-00749

26. Maung TZ, Ergin HE, Javed M, Inga EE, Khan S. Immune Checkpoint Inhibitors in Lung Cancer: Role of Biomarkers and Combination Therapies. *Cureus* (2020) 12(5):e8095. doi: 10.7759/cureus.8095

27. Lutz S, Balboni T, Jones J, Lo S, Petit J, Rich SE, et al. Palliative Radiation Therapy for Bone Metastases: Update of an ASTRO Evidence-Based Guideline. *Pract Radiat Oncol* (2017) 7(1):4–12. doi: 10.1016/j.prro.2016.08.001

28. Wang R, Zhu Y, Liu X, Liao X, He J, Niu L. The Clinicopathological Features and Survival Outcomes of Patients With Different Metastatic Sites in Stage IV Breast Cancer. *BMC Cancer* (2019) 19(1):1091. doi: 10.1186/s12885-019-6311-z

29. Wang X, Yang KH, Wanyan P, Tian JH. Comparison of the Efficacy and Safety of Denosumab Versus Bisphosphonates in Breast Cancer and Bone Metastases Treatment: A Meta-Analysis of Randomized Controlled Trials. *Oncol Lett* (2014) 7(6):1997–2002. doi: 10.3892/ol.2014.1982

30. Parker C, Nilsson S, Heinrich D, Helle SI, O'Sullivan JM, Fossa SD, et al. Alpha Emitter Radium-223 and Survival in Metastatic Prostate Cancer. *N Engl J Med* (2013) 369(3):213–23. doi: 10.1056/NEJMoa1213755

31. Merz M, Seyler L, Bretschi M, Semmler W, Bauerle T. Diffusion-Weighted Imaging and Dynamic Contrast-Enhanced MRI of Experimental Breast Cancer Bone Metastases–A Correlation Study With Histology. *Eur J Radiol* (2015) 84(4):623–30. doi: 10.1016/j.ejrad.2015.01.002

32. Bauerle T, Merz M, Komljenovic D, Zwick S, Semmler W. Drug-Induced Vessel Remodeling in Bone Metastases as Assessed by Dynamic Contrast Enhanced Magnetic Resonance Imaging and Vessel Size Imaging: A Longitudinal *In Vivo* Study. *Clin Cancer Res* (2010) 16(12):3215–25. doi: 10.1158/1078-0432.CCR-09-2932

33. Merz M, Komljenovic D, Zwick S, Semmler W, Bauerle T. Sorafenib Tosylate and Paclitaxel Induce Anti-Angiogenic, Anti-Tumor and Anti-Resorptive

Effects in Experimental Breast Cancer Bone Metastases. *Eur J Cancer* (2011) 47(2):277–86. doi: 10.1016/j.ejca.2010.08.019

34. Ellmann S, Seyler L, Evers J, Heinen H, Bozec A, Prante O, et al. Prediction of Early Metastatic Disease in Experimental Breast Cancer Bone Metastasis by Combining PET/CT and MRI Parameters to a Model-Averaged Neural Network. *Bone* (2019) 120:254–61. doi: 10.1016/j.bone.2018.11.008

35. Muhlhausen U, Komljenovic D, Bretschi M, Leotta K, Eisenhut M, Semmler W, et al. A Novel PET Tracer for the Imaging of Alphavbeta3 and Alphavbeta5 Integrins in Experimental Breast Cancer Bone Metastases. *Contrast Media Mol Imaging* (2011) 6(6):413–20. doi: 10.1002/cmmi.435

36. Hennrich U, Seyler L, Schafer M, Bauder-Wust U, Eisenhut M, Semmler W, et al. Synthesis and *In Vitro* Evaluation of 68Ga-DOTA-4-Fbn-TN14003, a Novel Tracer for the Imaging of CXCR4 Expression. *Bioorg Med Chem* (2012) 20(4):1502–10. doi: 10.1016/j.bmc.2011.12.052

37. Bauerle T, Gupta S, Zheng S, Seyler L, Leporati A, Marosfoi M, et al. Multimodal Bone Metastasis-Associated Epidermal Growth Factor Receptor Imaging in an Orthotopic Rat Model. *Radiol Imaging Cancer* (2021) 3(4): e200069. doi: 10.1148/rycan.2021200069

38. Heinen H, Seyler L, Popp V, Hellwig K, Bozec A, Uder M, et al. Morphological, Functional, and Molecular Assessment of Breast Cancer Bone Metastases by Experimental Ultrasound Techniques Compared With Magnetic Resonance Imaging and Histological Analysis. *Bone* (2021) 144:115821. doi: 10.1016/j.bone.2020.115821

39. Traill ZC, Talbot D, Golding S, Gleeson FV. Magnetic Resonance Imaging Versus Radionuclide Scintigraphy in Screening for Bone Metastases. *Clin Radiol* (1999) 54(7):448–51. doi: 10.1016/S0009-9260(99)90830-9

40. Lecouvet FE, Geukens D, Stainier A, Jamar F, Jamart J, d'Othee BJ, et al. Magnetic Resonance Imaging of the Axial Skeleton for Detecting Bone Metastases in Patients With High-Risk Prostate Cancer: Diagnostic and Cost-Effectiveness and Comparison With Current Detection Strategies. *J Clin Oncol* (2007) 25(22):3281–7. doi: 10.1200/JCO.2006.09.2940

41. Tombal B, Rezazadeh A, Therasse P, Van Cangh PJ, Vande Berg B, Lecouvet FE. Magnetic Resonance Imaging of the Axial Skeleton Enables Objective Measurement of Tumor Response on Prostate Cancer Bone Metastases. *Prostate* (2005) 65(2):178–87. doi: 10.1002/pros.20280

42. Lecouvet FE, Simon M, Tombal B, Jamart J, Vande Berg BC, Simoni P. Whole-Body MRI (WB-MRI) Versus Axial Skeleton MRI (as-MRI) to Detect and Measure Bone Metastases in Prostate Cancer (Pca). *Eur Radiol* (2010) 20(12):2973–82. doi: 10.1007/s00330-010-1879-3

43. Maeder Y, Dunet V, Richard R, Becce F, Omoumi P. Bone Marrow Metastases: T2-Weighted Dixon Spin-Echo Fat Images can Replace T1-Weighted Spin-Echo Images. *Radiology* (2018) 286(3):948–59. doi: 10.1148/radiol.2017170325

44. Van Nieuwenhove S, Van Damme J, Padhani AR, Vandecaveye V, Tombal B, Wuts J, et al. Whole-Body Magnetic Resonance Imaging for Prostate Cancer Assessment: Current Status and Future Directions. *J Magn Reson Imaging* (2020). doi: 10.1002/jmri.27485

45. Perez-Lopez R, Nava Rodrigues D, Figueiredo I, Mateo J, Collins DJ, Koh DM, et al. Multiparametric Magnetic Resonance Imaging of Prostate Cancer Bone Disease: Correlation With Bone Biopsy Histological and Molecular Features. *Invest Radiol* (2018) 53(2):96–102. doi: 10.1097/RLI.0000000000000415

46. Daffner RH, Lupetin AR, Dash N, Deeb ZL, Sefczek RJ, Schapiro RL. MRI in the Detection of Malignant Infiltration of Bone Marrow. *AJR Am J Roentgenol* (1986) 146(2):353–8. doi: 10.2214/ajr.146.2.353

47. Messiou C, Collins DJ, Giles S, de Bono JS, Bianchini D, de Souza NM. Assessing Response in Bone Metastases in Prostate Cancer With Diffusion Weighted MRI. *Eur Radiol* (2011) 21(10):2169–77. doi: 10.1007/s00330-011-2173-8

48. Pfannenberg C, Schwenzer N. [Whole-Body Staging of Malignant Melanoma: Advantages, Limitations and Current Importance of PET-CT, Whole-Body MRI and PET-MRI]. *Radiologe* (2015) 55(2):120–6. doi: 10.1007/s00117-014-2762-z

49. Schmidt GP, Reiser MF, Baur-Melnyk A. Whole-Body MRI for the Staging and Follow-Up of Patients With Metastasis. *Eur J Radiol* (2009) 70(3):393–400. doi: 10.1016/j.ejrad.2009.03.045

50. Lecouvet FE. Whole-Body Mr Imaging: Musculoskeletal Applications. *Radiology* (2016) 279(2):345–65. doi: 10.1148/radiol.2016142084

51. Switlyk MD, Hole KH, Skjeldal S, Hald JK, Knutstad K, Seierstad T, et al. MRI and Neurological Findings in Patients With Spinal Metastases. *Acta Radiol* (2012) 53(10):1164–72. doi: 10.1258/ar.2012.120442

52. Lecouvet FE, Vande Berg BC, Malghem J, Omoumi P, Simoni P. Diffusion-Weighted MR Imaging: Adjunct or Alternative to T1-Weighted MR Imaging for Prostate Carcinoma Bone Metastases? *Radiology* (2009) 252(2):624. doi: 10.1148/radiol.2522090263

53. Winfield JM, Poillucci G, Blackledge MD, Collins DJ, Shah V, Tunariu N, et al. Apparent Diffusion Coefficient of Vertebral Haemangiomas Allows Differentiation From Malignant Focal Deposits in Whole-Body Diffusion-Weighted MRI. *Eur Radiol* (2018) 28(4):1687–91. doi: 10.1007/s00330-017-5079-2

54. Woolf DK, Padhani AR, Makris A. Assessing Response to Treatment of Bone Metastases From Breast Cancer: What Should be the Standard of Care? *Ann Oncol* (2015) 26(6):1048–57. doi: 10.1093/annonc/mdu558

55. Padhani AR, Lecouvet FE, Tunariu N, Koh DM, De Keyzer F, Collins DJ, et al. Metastasis Reporting and Data System for Prostate Cancer: Practical Guidelines for Acquisition, Interpretation, and Reporting of Whole-Body Magnetic Resonance Imaging-Based Evaluations of Multiorgan Involvement in Advanced Prostate Cancer. *Eur Urol* (2017) 71(1):81–92. doi: 10.1016/j.eururo.2016.05.033

56. Bauerle T, Semmler W. Imaging Response to Systemic Therapy for Bone Metastases. *Eur Radiol* (2009) 19(10):2495–507. doi: 10.1007/s00330-009-1443-1

57. Padhani AR, Gogbashian A. Bony Metastases: Assessing Response to Therapy With Whole-Body Diffusion MRI. *Cancer Imaging.* (2011) 11 Spec No A:S129–45. doi: 10.1102/1470-7330.2011.9034

58. Lecouvet FE, Talbot JN, Messiou C, Bourguet P, Liu Y, de Souza NM, et al. Monitoring the Response of Bone Metastases to Treatment With Magnetic Resonance Imaging and Nuclear Medicine Techniques: A Review and Position Statement by the European Organisation for Research and Treatment of Cancer Imaging Group. *Eur J Cancer* (2014) 50(15):2519–31. doi: 10.1016/j.ejca.2014.07.002

59. Barnes A, Alonzi R, Blackledge M, Charles-Edwards G, Collins DJ, Cook G, et al. UK Quantitative WB-DWI Technical Workgroup: Consensus Meeting Recommendations on Optimisation, Quality Control, Processing and Analysis of Quantitative Whole-Body Diffusion-Weighted Imaging for Cancer. *Br J Radiol* (2018) 91(1081):20170577. doi: 10.1259/bjr.20170577

60. Lecouvet FE, Larbi A, Pasoglou V, Omoumi P, Tombal B, Michoux N, et al. MRI for Response Assessment in Metastatic Bone Disease. *Eur Radiol* (2013) 23(7):1986–97. doi: 10.1007/s00330-013-2792-3

61. Padhani AR, Koh DM. Diffusion MR Imaging for Monitoring of Treatment Response. *Magn Reson Imaging Clin N Am* (2011) 19(1):181–209. doi: 10.1016/j.mric.2010.10.004

62. Blackledge MD, Collins DJ, Tunariu N, Orton MR, Padhani AR, Leach MO, et al. Assessment of Treatment Response by Total Tumor Volume and Global Apparent Diffusion Coefficient Using Diffusion-Weighted MRI in Patients With Metastatic Bone Disease: A Feasibility Study. *PloS One* (2014) 9(4):e91779. doi: 10.1371/journal.pone.0091779

63. Yamamoto S, Yoshida S, Ishii C, Takahara T, Arita Y, Fukushima H, et al. Metastatic Diffusion Volume Based on Apparent Diffusion Coefficient as a Prognostic Factor in Castration-Resistant Prostate Cancer. *J Magn Reson Imaging* (2021) 54(2):401–8. doi: 10.1002/jmri.27596

64. Pricolo P, Ancona E, Summers P, Abreu-Gomez J, Alessi S, Jereczek-Fossa BA, et al. Whole-Body Magnetic Resonance Imaging (WB-MRI) Reporting With the Metastasis Reporting and Data System for Prostate Cancer (MET-RADS-P): Inter-Observer Agreement Between Readers of Different Expertise Levels. *Cancer Imaging* (2020) 20(1):77. doi: 10.1186/s40644-020-00350-x

65. Messiou C, Cook G, deSouza NM. Imaging Metastatic Bone Disease From Carcinoma of the Prostate. *Br J Cancer* (2009) 101(8):1225–32. doi: 10.1038/sj.bjc.6605334

66. Lecouvet FE, El Mouedden J, Collette L, Coche E, Danse E, Jamar F, et al. Can Whole-Body Magnetic Resonance Imaging With Diffusion-Weighted Imaging Replace Tc 99m Bone Scanning and Computed Tomography for Single-Step Detection of Metastases in Patients With High-Risk Prostate Cancer? *Eur Urol* (2012) 62(1):68–75. doi: 10.1016/j.eururo.2012.02.020

67. Afshar-Oromieh A, Debus N, Uhrig M, Hope TA, Evans MJ, Holland-Letz T, et al. Impact of Long-Term Androgen Deprivation Therapy on PSMA

Ligand PET/CT in Patients With Castration-Sensitive Prostate Cancer. *Eur J Nucl Med Mol Imaging* (2018) 45(12):2045–54. doi: 10.1007/s00259-018-4079-z

68. Liu T, Wu LY, Fulton MD, Johnson JM, Berkman CE. Prolonged Androgen Deprivation Leads to Downregulation of Androgen Receptor and Prostate-Specific Membrane Antigen in Prostate Cancer Cells. *Int J Oncol* (2012) 41 (6):2087–92. doi: 10.3892/ijo.2012.1649

69. Schmidt GP, Baur-Melnyk A, Haug A, Heinemann V, Bauerfeind I, Reiser MF, et al. Comprehensive Imaging of Tumor Recurrence in Breast Cancer Patients Using Whole-Body MRI at 1.5 and 3 T Compared to FDG-PET-CT. *Eur J Radiol* (2008) 65(1):47–58. doi: 10.1016/j.ejrad.2007.10.021

70. Nakanishi K, Kobayashi M, Nakaguchi K, Kyakuno M, Hashimoto N, Onishi H, et al. Whole-Body MRI for Detecting Metastatic Bone Tumor: Diagnostic Value of Diffusion-Weighted Images. *Magn Reson Med Sci* (2007) 6(3):147–55. doi: 10.2463/mrms.6.147

71. Azad GK, Taylor BP, Green A, Sandri I, Swampillai A, Harries M, et al. Prediction of Therapy Response in Bone-Predominant Metastatic Breast Cancer: Comparison of [(18)F] Fluorodeoxyglucose and [(18)F]-Fluoride PET/CT With Whole-Body MRI With Diffusion-Weighted Imaging. *Eur J Nucl Med Mol Imaging* (2019) 46(4):821–30. doi: 10.1007/s00259-018-4223-9

72. Kosmin M, Makris A, Joshi PV, Ah-See ML, Woolf D, Padhani AR. The Addition of Whole-Body Magnetic Resonance Imaging to Body Computerised Tomography Alters Treatment Decisions in Patients With Metastatic Breast Cancer. *Eur J Cancer* (2017) 77:109–16. doi: 10.1016/j.ejca.2017.03.001

73. Zugni F, Ruju F, Pricolo P, Alessi S, Iorfida M, Colleoni MA, et al. The Added Value of Whole-Body Magnetic Resonance Imaging in the Management of Patients With Advanced Breast Cancer. *PloS One* (2018) 13(10):e0205251. doi: 10.1371/journal.pone.0205251

74. Kosmin M, Padhani AR, Gogbashian A, Woolf D, Ah-See ML, Ostler P, et al. Comparison of Whole-Body MRI, CT, and Bone Scintigraphy for Response Evaluation of Cancer Therapeutics in Metastatic Breast Cancer to Bone. *Radiology* (2020) 297(3):622–9. doi: 10.1148/radiol.2020192683

75. Kosmin M, Padhani AR, Sokhi H, Thijssen T, Makris A. Patterns of Disease Progression in Patients With Local and Metastatic Breast Cancer as Evaluated by Whole-Body Magnetic Resonance Imaging. *Breast* (2018) 40:82–4. doi: 10.1016/j.breast.2018.04.019

76. Jacobs MA, Macura KJ, Zaheer A, Antonarakis ES, Stearns V, Wolff AC, et al. Multiparametric Whole-Body MRI With Diffusion-Weighted Imaging and ADC Mapping for the Identification of Visceral and Osseous Metastases From Solid Tumors. *Acad Radiol* (2018) 25(11):1405–14. doi: 10.1016/j.acra.2018.02.010

77. Wahl RL, Jacene H, Kasamon Y, Lodge MA. From RECIST to PERCIST: Evolving Considerations for PET Response Criteria in Solid Tumors. *J Nucl Med* (2009) 50 Suppl 1:122S–50S. doi: 10.2967/jnumed.108.057307

78. Frings V, Yaqub M, Hoyng LL, Golla SS, Windhorst AD, Schuit RC, et al. Assessment of Simplified Methods to Measure 18F-FLT Uptake Changes in EGFR-Mutated Non-Small Cell Lung Cancer Patients Undergoing EGFR Tyrosine Kinase Inhibitor Treatment. *J Nucl Med* (2014) 55(9):1417–23. doi: 10.2967/jnumed.114.140913

79. Verwer EE, Oprea-Lager DE, van den Eertwegh AJ, van Moorselaar RJ, Windhorst AD, Schwarte LA, et al. Quantification of 18F-Fluorocholine Kinetics in Patients With Prostate Cancer. *J Nucl Med* (2015) 56(3):365–71. doi: 10.2967/jnumed.114.148007

80. Kramer GM, Yaqub M, Vargas HA, Schuit RC, Windhorst AD, van den Eertwegh AJM, et al. Assessment of Simplified Methods for Quantification of (18)F-FDHT Uptake in Patients With Metastatic Castration-Resistant Prostate Cancer. *J Nucl Med* (2019) 60(9):1221–7. doi: 10.2967/jnumed.118.220111

81. Jansen BHE, Yaqub M, Voortman J, Cysouw MCF, Windhorst AD, Schuit RC, et al. Simplified Methods for Quantification of (18)F-Dcfpyl Uptake in Patients With Prostate Cancer. *J Nucl Med* (2019) 60(12):1730–5. doi: 10.2967/jnumed.119.227520

82. Schwartz J, Grkovski M, Rimner A, Schoder H, Zanzonico PB, Carlin SD, et al. Pharmacokinetic Analysis of Dynamic (18)F-Fluoromisonidazole PET Data in Non-Small Cell Lung Cancer. *J Nucl Med* (2017) 58(6):911–9. doi: 10.2967/jnumed.116.180422

83. Zaidi H, Karakatsanis N. Towards Enhanced PET Quantification in Clinical Oncology. *Br J Radiol* (2018) 91(1081):20170508. doi: 10.1259/bjr.20170508

84. Larson SM, Erdi Y, Akhurst T, Mazumdar M, Macapinlac HA, Finn RD, et al. Tumor Treatment Response Based on Visual and Quantitative Changes in Global Tumor Glycolysis Using PET-FDG Imaging. The Visual Response Score and the Change in Total Lesion Glycolysis. *Clin Positron Imaging* (1999) 2(3):159–71. doi: 10.1016/s1095-0397(99)00016-3

85. Schmuck S, von Klot CA, Henkenberens C, Sohns JM, Christiansen H, Wester HJ, et al. Initial Experience With Volumetric (68)Ga-PSMA I&T PET/CT for Assessment of Whole-Body Tumor Burden as a Quantitative Imaging Biomarker in Patients With Prostate Cancer. *J Nucl Med* (2017) 58 (12):1962–8. doi: 10.2967/jnumed.117.193581

86. Calais J, Ceci F, Eiber M, Hope TA, Hofman MS, Rischpler C, et al. (18)F-Fluciclovine PET-CT and (68)Ga-PSMA-11 PET-CT in Patients With Early Biochemical Recurrence After Prostatectomy: A Prospective, Single-Centre, Single-Arm, Comparative Imaging Trial. *Lancet Oncol* (2019) 20(9):1286–94. doi: 10.1016/S1470-2045(19)30415-2

87. Afshar-Oromieh A, Zechmann CM, Malcher A, Eder M, Eisenhut M, Linhart HG, et al. Comparison of PET Imaging With a (68)Ga-Labelled PSMA Ligand and (18)F-Choline-Based PET/CT for the Diagnosis of Recurrent Prostate Cancer. *Eur J Nucl Med Mol Imaging* (2014) 41(1):11–20. doi: 10.1007/s00259-013-2525-5

88. Eiber M, Herrmann K, Calais J, Hadaschik B, Giesel FL, Hartenbach M, et al. Prostate Cancer Molecular Imaging Standardized Evaluation (PROMISE): Proposed Mitnm Classification for the Interpretation of PSMA-Ligand PET/CT. *J Nucl Med* (2018) 59(3):469–78. doi: 10.2967/jnumed.117.198119

89. Prasad V, Huang K, Prasad S, Makowski MR, Czech N, Brenner W. In Comparison to PSA, Interim Ga-68-PSMA PET/CT Response Evaluation Based on Modified RECIST 1.1 After 2(Nd) Cycle Is Better Predictor of Overall Survival of Prostate Cancer Patients Treated With (177)Lu-PSMA. *Front Oncol* (2021) 11:578093. doi: 10.3389/fonc.2021.578093

90. Grubmuller B, Senn D, Kramer G, Baltzer P, D'Andrea D, Grubmuller KH, et al. Response Assessment Using (68)Ga-PSMA Ligand PET in Patients Undergoing (177)Lu-PSMA Radioligand Therapy for Metastatic Castration-Resistant Prostate Cancer. *Eur J Nucl Med Mol Imaging* (2019) 46(5):1063–72. doi: 10.1007/s00259-018-4236-4

91. Rosar F, Hau F, Bartholoma M, Maus S, Stemler T, Linxweiler J, et al. Molecular Imaging and Biochemical Response Assessment After a Single Cycle of [(225)Ac]Ac-PSMA-617/[(177)Lu]Lu-PSMA-617 Tandem Therapy in Mcrpc Patients Who Have Progressed on [(177)Lu]Lu-PSMA-617 Monotherapy. *Theranostics* (2021) 11(9):4050–60. doi: 10.7150/thno.56211

92. Han S, Woo S, Kim YI, Lee JL, Wibmer AG, Schoder H, et al. Concordance Between Response Assessment Using Prostate-Specific Membrane Antigen PET and Serum Prostate-Specific Antigen Levels After Systemic Treatment in Patients With Metastatic Castration Resistant Prostate Cancer: A Systematic Review and Meta-Analysis. *Diagnostics (Basel)* (2021) 11 (4):663. doi: 10.3390/diagnostics11040663

93. Fox JJ, Gavane SC, Blanc-Autran E, Nehmeh S, Gonen M, Beattie B, et al. Positron Emission Tomography/Computed Tomography-Based Assessments of Androgen Receptor Expression and Glycolytic Activity as a Prognostic Biomarker for Metastatic Castration-Resistant Prostate Cancer. *JAMA Oncol* (2018) 4(2):217–24. doi: 10.1001/jamaoncol.2017.3588

94. Vargas HA, Kramer GM, Scott AM, Weickhardt A, Meier AA, Parada N, et al. Reproducibility and Repeatability of Semiquantitative (18)F-Fluorodihydrotestosterone Uptake Metrics in Castration-Resistant Prostate Cancer Metastases: A Prospective Multicenter Study. *J Nucl Med* (2018) 59 (10):1516–23. doi: 10.2967/jnumed.117.206490

95. Scher HI, Beer TM, Higano CS, Anand A, Taplin ME, Efstathiou E, et al. Antitumor Activity of MDV3100 in Castration-Resistant Prostate Cancer: A Phase 1-2 Study. *Lancet* (2010) 375(9724):1437–46. doi: 10.1016/S0140-6736(10)60172-9

96. Rathkopf DE, Morris MJ, Fox JJ, Danila DC, Slovin SF, Hager JH, et al. Phase I Study of ARN-509, a Novel Antiandrogen, in the Treatment of Castration-Resistant Prostate Cancer. *J Clin Oncol* (2013) 31(28):3525–30. doi: 10.1200/JCO.2013.50.1684

97. Jansen BHE, Cysouw MCF, Vis AN, van Moorselaar RJA, Voortman J, Bodar YJL, et al. Repeatability of Quantitative (18)F-Dcfpyl PET/CT

Measurements in Metastatic Prostate Cancer. *J Nucl Med* (2020) 61(9):1320–5. doi: 10.2967/jnumed.119.236075

98. Pollard JH, Raman C, Zakharia Y, Tracy CR, Nepple KG, Ginader T, et al. Quantitative Test-Retest Measurement of (68)Ga-PSMA-HBED-CC in Tumor and Normal Tissue. *J Nucl Med* (2020) 61(8):1145–52. doi: 10.2967/jnumed.119.236083

99. Mammatas LH, Venema CM, Schroder CP, de Vet HCW, van Kruchten M, Glaudemans A, et al. Visual and Quantitative Evaluation of [(18)F]FES and [(18)F]FDHT PET in Patients With Metastatic Breast Cancer: An Interobserver Variability Study. *EJNMMI Res* (2020) 10(1):40. doi: 10.1186/s13550-020-00627-z

100. Sheikhbahaei S, Jones KM, Werner RA, Salas-Fragomeni RA, Marcus CV, Higuchi T, et al. (18)F-Naf-PET/CT for the Detection of Bone Metastasis in Prostate Cancer: A Meta-Analysis of Diagnostic Accuracy Studies. *Ann Nucl Med* (2019) 33(5):351–61. doi: 10.1007/s12149-019-01343-y

101. van Kruchten M, de Vries EGE, Brown M, de Vries EFJ, Glaudemans A, Dierckx R, et al. PET Imaging of Oestrogen Receptors in Patients With Breast Cancer. *Lancet Oncol* (2013) 14(11):e465–75. doi: 10.1016/S1470-2045(13)70292-4

102. 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(6):801–5. doi: 10.2967/jnumed.119.227967

103. Schirrmeister H, Arslandemir C, Glatting G, Mayer-Steinacker R, Bommer M, Dreinhofer K, et al. Omission of Bone Scanning According to Staging Guidelines Leads to Futile Therapy in non-Small Cell Lung Cancer. *Eur J Nucl Med Mol Imaging* (2004) 31(7):964–8. doi: 10.1007/s00259-004-1492-2

104. Brown JE, Cook RJ, Major P, Lipton A, Saad F, Smith M, et al. Bone Turnover Markers as Predictors of Skeletal Complications in Prostate Cancer, Lung Cancer, and Other Solid Tumors. *J Natl Cancer Inst* (2005) 97(1):59–69. doi: 10.1093/jnci/dji002

105. Ung YC, Maziak DE, Vanderveen JA, Smith CA, Gulenchyn K, Lacchetti C, et al. 18Fluorodeoxyglucose Positron Emission Tomography in the Diagnosis and Staging of Lung Cancer: A Systematic Review. *J Natl Cancer Inst* (2007) 99(23):1753–67. doi: 10.1093/jnci/djm232

106. Planchard D, Popat S, Kerr K, Novello S, Smit EF, Faivre-Finn C, et al. Metastatic non-Small Cell Lung Cancer: ESMO Clinical Practice Guidelines for Diagnosis, Treatment and Follow-Up. *Ann Oncol* (2018) 29(Suppl 4):iv192–237. doi: 10.1093/annonc/mdy275

107. Postmus PE, Kerr KM, Oudkerk M, Senan S, Waller DA, Vansteenkiste J, et al. Early and Locally Advanced non-Small-Cell Lung Cancer (NSCLC): ESMO Clinical Practice Guidelines for Diagnosis, Treatment and Follow-Up. *Ann Oncol* (2017) 28(suppl_4):iv1–iv21. doi: 10.1093/annonc/mdx222

108. Fruh M, De Ruysscher D, Popat S, Crino L, Peters S, Felip E, et al. Small-Cell Lung Cancer (SCLC): ESMO Clinical Practice Guidelines for Diagnosis, Treatment and Follow-Up. *Ann Oncol* (2013) 24 Suppl 6:vi99–105. doi: 10.1093/annonc/mdt178

109. NCCN Clinical Practice Guidelines in Oncology (NCCN Guidelines®). Non-Small Cell Lung Cancer. Version 5.2021. Available at: www.nccn.org.

110. NCCN Clinical Practice Guidelines in Oncology (NCCN Guidelines®). Small Cell Lung Cancer. Version 1.2022. Available at: www.nccn.org.

111. Quartuccio N, Salem A, Laudicella R, Spataro A, Chiaravalloti A, Caobelli F, et al. The Role of 18F-Fluorodeoxyglucose PET/CT in Restaging Patients With Small Cell Lung Cancer: A Systematic Review. *Nucl Med Commun* (2021) 42(8):839–45. doi: 10.1097/MNM.0000000000001407

112. Chang MC, Chen JH, Liang JA, Lin CC, Yang KT, Cheng KY, et al. Meta-Analysis: Comparison of F-18 Fluorodeoxyglucose-Positron Emission Tomography and Bone Scintigraphy in the Detection of Bone Metastasis in Patients With Lung Cancer. *Acad Radiol* (2012) 19(3):349–57. doi: 10.1016/j.acra.2011.10.018

113. Liu N, Ma L, Zhou W, Pang Q, Hu M, Shi F, et al. Bone Metastasis in Patients With non-Small Cell Lung Cancer: The Diagnostic Role of F-18 FDG PET/CT. *Eur J Radiol* (2010) 74(1):231–5. doi: 10.1016/j.ejrad.2009.01.036

114. Wu Y, Li P, Zhang H, Shi Y, Wu H, Zhang J, et al. Diagnostic Value of Fluorine 18 Fluorodeoxyglucose Positron Emission Tomography/Computed Tomography for the Detection of Metastases in non-Small-Cell Lung Cancer Patients. *Int J Cancer* (2013) 132(2):E37–47. doi: 10.1002/ijc.27779

115. Martucci F, Pascale M, Valli MC, Pesce GA, Froesch P, Giovanella L, et al. Impact of (18)F-FDG PET/CT in Staging Patients With Small Cell Lung

Cancer: A Systematic Review and Meta-Analysis. *Front Med (Lausanne)* (2019) 6:336. doi: 10.3389/fmed.2019.00336

116. Machado Medeiros T, Altmayer S, Watte G, Zanon M, Basso Dias A, Henz Concatto N, et al. 18F-FDG PET/CT and Whole-Body MRI Diagnostic Performance in M Staging for Non-Small Cell Lung Cancer: A Systematic Review and Meta-Analysis. *Eur Radiol* (2020) 30(7):3641–9. doi: 10.1007/s00330-020-06703-1

117. Li J, Zhou H, Zhang X, Song F, Pang X, Wei Z. A Two-Way Comparison of Whole-Body 18FDG PET-CT and Whole-Body Contrast-Enhanced MRI for Distant Metastasis Staging in Patients With Malignant Tumors: A Meta-Analysis of 13 Prospective Studies. *Ann Palliat Med* (2020) 9(2):247–55. doi: 10.21037/apm.2020.02.30

118. Qu X, Huang X, Yan W, Wu L, Dai K. A Meta-Analysis of (1)(8)FDG-PET-CT, (1)(8)FDG-PET, MRI and Bone Scintigraphy for Diagnosis of Bone Metastases in Patients With Lung Cancer. *Eur J Radiol* (2012) 81(5):1007–15. doi: 10.1016/j.ejrad.2011.01.126

119. Cardoso F, Kyriakides S, Ohno S, Penault-Llorca F, Poortmans P, Rubio IT, et al. Early Breast Cancer: ESMO Clinical Practice Guidelines for Diagnosis, Treatment and Follow-Updagger. *Ann Oncol* (2019) 30(8):1194–220. doi: 10.1093/annonc/mdz173

120. Cardoso F, Paluch-Shimon S, Senkus E, Curigliano G, Aapro MS, Andre F, et al. 5th ESO-ESMO International Consensus Guidelines for Advanced Breast Cancer (ABC 5). *Ann Oncol* (2020) 31(12):1623–49. doi: 10.1016/j.annonc.2020.09.010

121. Han S, Choi JY. Impact of 18F-FDG PET, PET/CT, and PET/MRI on Staging and Management as an Initial Staging Modality in Breast Cancer: A Systematic Review and Meta-Analysis. *Clin Nucl Med* (2021) 46(4):271–82. doi: 10.1097/RLU.0000000000003502

122. Houssami N, Costelloe CM. Imaging Bone Metastases in Breast Cancer: Evidence on Comparative Test Accuracy. *Ann Oncol* (2012) 23(4):834–43. doi: 10.1093/annonc/mdr397

123. Rong J, Wang S, Ding Q, Yun M, Zheng Z, Ye S. Comparison of 18 FDG PET-CT and Bone Scintigraphy for Detection of Bone Metastases in Breast Cancer Patients. A meta-analysis. *Surg Oncol* (2013) 22(2):86–91. doi: 10.1016/j.suronc.2013.01.002

124. Sun Z, Yi YL, Liu Y, Xiong JP, He CZ. Comparison of Whole-Body PET/PET-CT and Conventional Imaging Procedures for Distant Metastasis Staging in Patients With Breast Cancer: A Meta-Analysis. *Eur J Gynaecol Oncol* (2015) 36(6):672–6.

125. Rugo HS, Rumble RB, Macrae E, Barton DL, Connolly HK, Dickler MN, et al. Endocrine Therapy for Hormone Receptor-Positive Metastatic Breast Cancer: American Society of Clinical Oncology Guideline. *J Clin Oncol* (2016) 34(25):3069–103. doi: 10.1200/JCO.2016.67.1487

126. Mintun MA, Welch MJ, Siegel BA, Mathias CJ, Brodack JW, McGuire AH, et al. Breast Cancer: PET Imaging of Estrogen Receptors. *Radiology* (1988) 169(1):45–8. doi: 10.1148/radiology.169.1.3262228

127. Kurland BF, Wiggins JR, Coche A, Fontan C, Bouvet Y, Webner P, et al. Whole-Body Characterization of Estrogen Receptor Status in Metastatic Breast Cancer With 16alpha-18F-Fluoro-17beta-Estradiol Positron Emission Tomography: Meta-Analysis and Recommendations for Integration Into Clinical Applications. *Oncologist* (2020) 25(10):835–44. doi: 10.1634/theoncologist.2019-0967

128. Liu C, Gong C, Liu S, Zhang Y, Zhang Y, Xu X, et al. (18)F-FES PET/CT Influences the Staging and Management of Patients With Newly Diagnosed Estrogen Receptor-Positive Breast Cancer: A Retrospective Comparative Study With (18)F-FDG PET/CT. *Oncologist* (2019) 24(12):e1277–85. doi: 10.1634/theoncologist.2019-0096

129. Pauletti G, Dandekar S, Rong H, Ramos L, Peng H, Seshadri R, et al. Assessment of Methods for Tissue-Based Detection of the HER-2/Neu Alteration in Human Breast Cancer: A Direct Comparison of Fluorescence *in Situ* Hybridization and Immunohistochemistry. *J Clin Oncol* (2000) 18 (21):3651–64. doi: 10.1200/JCO.2000.18.21.3651

130. Zhou N, Liu C, Guo X, Xu Y, Gong J, Qi C, et al. Impact of (68)Ga-NOTA-MAL-MZHER2 PET Imaging in Advanced Gastric Cancer Patients and Therapeutic Response Monitoring. *Eur J Nucl Med Mol Imaging* (2021) 48 (1):161–75. doi: 10.1007/s00259-020-04898-5

131. Sorensen 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

132. Gebhart G, Lamberts LE, Wimana Z, Garcia C, Emonts P, Ameye L, et al. Molecular Imaging as a Tool to Investigate Heterogeneity of Advanced HER2-Positive Breast Cancer and to Predict Patient Outcome Under Trastuzumab Emtansine (T-DM1): The ZEPHIR Trial. *Ann Oncol* (2016) 27(4):619–24. doi: 10.1093/annonc/mdv577

133. Mortimer JE, Bading JR, Park JM, Frankel PH, Carroll MI, Tran TT, et al. Tumor Uptake of (64)Cu-DOTA-Trastuzumab in Patients With Metastatic Breast Cancer. *J Nucl Med* (2018) 59(1):38–43. doi: 10.2967/jnumed.117.193888

134. Kramer-Marek G, Bernardo M, Kiesewetter DO, Bagci U, Kuban M, Aras O, et al. PET of HER2-Positive Pulmonary Metastases With 18F-ZHER2:342 Affibody in a Murine Model of Breast Cancer: Comparison With 18F-FDG. *J Nucl Med* (2012) 53(6):939–46. doi: 10.2967/jnumed.111.100354

135. Kramer-Marek G, Oyen WJ. Targeting the Human Epidermal Growth Factor Receptors With Immuno-PET: Imaging Biomarkers From Bench to Bedside. *J Nucl Med* (2016) 57(7):996–1001. doi: 10.2967/jnumed.115.169540

136. Sandstrom 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

137. Boellaard R, Krak NC, Hoekstra OS, Lammertsma AA. Effects of Noise, Image Resolution, and ROI Definition on the Accuracy of Standard Uptake Values: A Simulation Study. *J Nucl Med* (2004) 45(9):1519–27.

138. Kolinger GD, Vallez Garcia D, Kramer GM, Frings V, Smit EF, de Langen AJ, et al. Repeatability of [(18)F]FDG PET/CT Total Metabolic Active Tumor Volume and Total Tumor Burden in NSCLC Patients. *EJNMMI Res* (2019) 9(1):14. doi: 10.1186/s13550-019-0481-1

139. Frings V, van Velden FH, Velasquez LM, Hayes W, van de Ven PM, Hoekstra OS, et al. Repeatability of Metabolically Active Tumor Volume Measurements With FDG PET/CT in Advanced Gastrointestinal Malignancies: A Multicenter Study. *Radiology* (2014) 273(2):539–48. doi: 10.1148/radiol.14132807

140. Hartrampf PE, Heinrich M, Seitz AK, Brumberg J, Sokolakis I, Kalogirou C, et al. Metabolic Tumor Volume From PSMA PET/CT Scans of Prostate Cancer Patients During Chemotherapy-Do Different Software Solutions Deliver Comparable Results? *J Clin Med* (2020) 9(5):1390. doi: 10.3390/jcm9051390

141. Gafita A, Bieth M, Kronke M, Tetteh G, Navarro F, Wang H, et al. Qpsma: Semiautomatic Software for Whole-Body Tumor Burden Assessment in Prostate Cancer Using (68)Ga-PSMA11 PET/CT. *J Nucl Med* (2019) 60(9):1277–83. doi: 10.2967/jnumed.118.224055

142. Kaalep A, Burggraaff CN, Pieplenbosch S, Verwer EE, Sera T, Zijlstra J, et al. Quantitative Implications of the Updated EARL 2019 PET-CT Performance Standards. *EJNMMI Phys* (2019) 6(1):28. doi: 10.1186/s40658-019-0257-8

143. Devriese J, Beels L, Maes A, Van de Wiele C, Pottel H. Impact of PET Reconstruction Protocols on Quantification of Lesions That Fulfil the PERCIST Lesion Inclusion Criteria. *EJNMMI Phys* (2018) 5(1):35. doi: 10.1186/s40658-018-0235-6

144. Quak E, Le Roux PY, Hofman MS, Robin P, Bourhis D, Callahan J, et al. Harmonizing FDG PET Quantification While Maintaining Optimal Lesion Detection: Prospective Multicentre Validation in 517 Oncology Patients. *Eur J Nucl Med Mol Imaging* (2015) 42(13):2072–82. doi: 10.1007/s00259-015-3128-0

145. Kaalep A, Sera T, Rijnsdorp S, Yaqub M, Talsma A, Lodge MA, et al. Feasibility of State of the Art PET/CT Systems Performance Harmonisation. *Eur J Nucl Med Mol Imaging* (2018) 45(8):1344–61. doi: 10.1007/s00259-018-3977-4

146. Kaalep A, Sera T, Oyen W, Krause BJ, Chiti A, Liu Y, et al. EANM/EARL FDG-PET/CT Accreditation - Summary Results From the First 200 Accredited Imaging Systems. *Eur J Nucl Med Mol Imaging* (2018) 45(3):412–22. doi: 10.1007/s00259-017-3853-7

147. Aggarwal R, Wei X, Kim W, Small EJ, Ryan CJ, Carroll P, et al. Heterogeneous Flare in Prostate-Specific Membrane Antigen Positron Emission Tomography Tracer Uptake With Initiation of Androgen Pathway Blockade in Metastatic Prostate Cancer. *Eur Urol Oncol* (2018) 1(1):78–82. doi: 10.1016/j.euo.2018.03.010

148. Krupitskaya Y, Eslamy HK, Nguyen DD, Kumar A, Wakelee HA. Osteoblastic Bone Flare on F18-FDG PET in non-Small Cell Lung Cancer (NSCLC) Patients Receiving Bevacizumab in Addition to Standard Chemotherapy. *J Thorac Oncol* (2009) 4(3):429–31. doi: 10.1097/JTO.0b013e3181989e12

149. De Giorgi U, Caroli P, Burgio SL, Menna C, Conteduca V, Bianchi E, et al. Early Outcome Prediction on 18F-Fluorocholine PET/CT in Metastatic Castration-Resistant Prostate Cancer Patients Treated With Abiraterone. *Oncotarget* (2014) 5(23):12448–58. doi: 10.18632/oncotarget.2558

150. Conteduca V, Poti G, Caroli P, Russi S, Brighi N, Lolli C, et al. Flare Phenomenon in Prostate Cancer: Recent Evidence on New Drugs and Next Generation Imaging. *Ther Adv Med Oncol* (2021) 13:1758835920987654. doi: 10.1177/1758835920987654

151. Cherry SR, Jones T, Karp JS, Qi J, Moses WW, Badawi RD. Total-Body PET: Maximizing Sensitivity to Create New Opportunities for Clinical Research and Patient Care. *J Nucl Med* (2018) 59(1):3–12. doi: 10.2967/jnumed.116.184028

152. Vandenberghe S, Moskal P, Karp JS. State of the Art in Total Body PET. *EJNMMI Phys* (2020) 7(1):35. doi: 10.1186/s40658-020-00290-2

153. Alberts I, Hunermund JN, Prenosil G, Mingels C, Bohn KP, Viscione M, et al. Clinical Performance of Long Axial Field of View PET/CT: A Head-to-Head Intra-Individual Comparison of the Biograph Vision Quadra With the Biograph Vision PET/CT. *Eur J Nucl Med Mol Imaging* (2021) 48(8):2395–404. doi: 10.1007/s00259-021-05282-7

154. Fahrni G, Karakatsanis NA, Di Domenicantonio G, Garibotto V, Zaidi H. Does Whole-Body Patlak (18)F-FDG PET Imaging Improve Lesion Detectability in Clinical Oncology? *Eur Radiol* (2019) 29(9):4812–21. doi: 10.1007/s00330-018-5966-1

155. Visvikis D, Cheze Le Rest C, Jaouen V, Hatt M. Artificial Intelligence, Machine (Deep) Learning and Radio(Geno)Mics: Definitions and Nuclear Medicine Imaging Applications. *Eur J Nucl Med Mol Imaging* (2019) 46(13):2630–7. doi: 10.1007/s00259-019-04373-w

156. Sollini M, Cozzi L, Antunovic L, Chiti A, Kirienko M. PET Radiomics in NSCLC: State of the Art and a Proposal for Harmonization of Methodology. *Sci Rep* (2017) 7(1):358. doi: 10.1038/s41598-017-00426-y

157. Zwanenburg A, Vallieres M, Abdalah MA, Aerts H, Andrearczyk V, Apte A, et al. The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-Based Phenotyping. *Radiology* (2020) 295(2):328–38. doi: 10.1148/radiol.2020191145

158. Fournier L, Costaridou L, Bidaut L, Michoux N, Lecouvet FE, de Geus-Oei LF, et al. Incorporating Radiomics Into Clinical Trials: Expert Consensus Endorsed by the European Society of Radiology on Considerations for Data-Driven Compared to Biologically Driven Quantitative Biomarkers. *Eur Radiol* (2021) 31(8):6001–12. doi: 10.1007/s00330-020-07598-8

159. Arabi H, AkhavanAllaf A, Sanaat A, Shiri I, Zaidi H. The Promise of Artificial Intelligence and Deep Learning in PET and SPECT Imaging. *Phys Med* (2021) 83:122–37. doi: 10.1016/j.ejmp.2021.03.008

160. Seifert R, Weber M, Kocakavuk E, Rischpler C, Kersting D. Artificial Intelligence and Machine Learning in Nuclear Medicine: Future Perspectives. *Semin Nucl Med* (2021) 51(2):170–7. doi: 10.1053/j.semnuclmed.2020.08.003

161. Nickols N, Anand A, Johnsson K, Brynolfsson J, Borrelli P, Juarez J, et al. Apromise: A Novel Automated-PROMISE Platform to Standardize Evaluation of Tumor Burden in (18)F-Dcfpyl (PSMA) Images of Veterans With Prostate Cancer. *J Nucl Med* (2021). doi: 10.2967/jnumed.120.261863

162. Schmidkonz C, Ellmann S, Ritt P, Roemer FW, Guermazi A, Uder M, et al. Hybrid Imaging (PET-Computed Tomography/PET-MR Imaging) of Bone Metastases. *PET Clin* (2019) 14(1):121–33. doi: 10.1016/j.cpet.2018.08.003

163. Eiber M, Takei T, Souvatzoglou M, Mayerhoefer ME, Furst S, Gaertner FC, et al. Performance of Whole-Body Integrated 18F-FDG PET/MR in Comparison to PET/CT for Evaluation of Malignant Bone Lesions. *J Nucl Med* (2014) 55(2):191–7. doi: 10.2967/jnumed.113.123646

164. Samarin A, Hullner M, Queiroz MA, Stolzmann P, Burger IA, von Schulthess G, et al. 18F-FDG-PET/MR Increases Diagnostic Confidence in Detection of Bone Metastases Compared With 18F-FDG-PET/CT. *Nucl Med Commun* (2015) 36(12):1165–73. doi: 10.1097/MNM.0000000000000387

165. Bruckmann NM, Kirchner J, Umutlu L, Fendler WP, Seifert R, Herrmann K, et al. Prospective Comparison of the Diagnostic Accuracy of 18F-FDG PET/MRI, MRI, CT, and Bone Scintigraphy for the Detection of Bone Metastases in the Initial Staging of Primary Breast Cancer Patients. *Eur Radiol* (2021) 31(11):8714–24. doi: 10.1007/s00330-021-07956-0

166. Catalano OA, Nicolai E, Rosen BR, Luongo A, Catalano M, Iannace C, et al. Comparison of CE-FDG-PET/CT With CE-FDG-PET/MR in the Evaluation of Osseous Metastases in Breast Cancer Patients. *Br J Cancer* (2015) 112(9):1452–60. doi: 10.1038/bjc.2015.112

167. Joshi A, Roberts MJ, Perera M, Williams E, Rhee H, Pryor D, et al. The Clinical Efficacy of PSMA PET/MRI in Biochemically Recurrent Prostate Cancer Compared With Standard of Care Imaging Modalities and Confirmatory Histopathology: Results of a Single-Centre, Prospective Clinical Trial. *Clin Exp Metastasis* (2020) 37(4):551–60. doi: 10.1007/s10585-020-10043-1

168. Freitag MT, Radtke JP, Hadaschik BA, Kopp-Schneider A, Eder M, Kopka K, et al. Comparison of Hybrid (68)Ga-PSMA PET/MRI and (68)Ga-PSMA PET/CT in the Evaluation of Lymph Node and Bone Metastases of Prostate Cancer. *Eur J Nucl Med Mol Imaging* (2016) 43(1):70–83. doi: 10.1007/s00259-015-3206-3

169. Sonni I, Minamimoto R, Baratto L, Gambhir SS, Loening AM, Vasanawala SS, et al. Simultaneous PET/MRI in the Evaluation of Breast and Prostate Cancer Using Combined Na[(18)F] F and [(18)F]FDG: A Focus on Skeletal Lesions. *Mol Imaging Biol* (2020) 22(2):397–406. doi: 10.1007/s11307-019-01392-9

170. JH O, Lodge MA, Wahl RL. Practical PERCIST: A Simplified Guide to PET Response Criteria in Solid Tumors 1.0. *Radiology* (2016) 280(2):576–84. doi: 10.1148/radiol.2016142043

171. Cheson BD, Fisher RI, Barrington SF, Cavalli F, Schwartz LH, Zucca E, et al. Recommendations for Initial Evaluation, Staging, and Response Assessment of Hodgkin and Non-Hodgkin Lymphoma: The Lugano Classification. *J Clin Oncol* (2014) 32(27):3059–68. doi: 10.1200/JCO.2013.54.8800

172. Pauwels E, Celen S, Vandamme M, Leysen W, Baete K, Bechter O, et al. Improved Resolution and Sensitivity of [(18)F]MFBG PET Compared With [(123)I]MIBG SPECT in a Patient With a Norepinephrine Transporter-Expressing Tumor. *Eur J Nucl Med Mol Imaging* (2021) 48(1):313–5. doi: 10.1007/s00259-020-04830-x

173. Terao T, Machida Y, Narita K, Kuzume A, Tabata R, Tsushima T, et al. Total Diffusion Volume in MRI vs. Total Lesion Glycolysis in PET/CT for Tumor Volume Evaluation of Multiple Myeloma. *Eur Radiol* (2021) 31(8):6136–44. doi: 10.1007/s00330-021-07687-2

174. Bauckneht M, Capitanio S, Donegani MI, Zanardi E, Miceli A, Murialdo R, et al. Role of Baseline and Post-Therapy 18F-FDG PET in the Prognostic Stratification of Metastatic Castration-Resistant Prostate Cancer (Mcrpc) Patients Treated With Radium-223. *Cancers (Basel)* (2019) 12(1):31. doi: 10.3390/cancers12010031

175. Cottereau AS, Meignan M, Nioche C, Capobianco N, Clerc J, Chartier L, et al. Risk Stratification in Diffuse Large B-Cell Lymphoma Using Lesion Dissemination and Metabolic Tumor Burden Calculated From Baseline PET/CT(Dagger). *Ann Oncol* (2021) 32(3):404–11. doi: 10.1016/j.annonc.2020.11.019

176. Apostolova I, Ego K, Steffen IG, Buchert R, Wertzel H, Achenbach HJ, et al. The Asphericity of the Metabolic Tumor Volume in NSCLC: Correlation With Histopathology and Molecular Markers. *Eur J Nucl Med Mol Imaging* (2016) 43(13):2360–73. doi: 10.1007/s00259-016-3452-z

177. El-Hennawy G, Moustafa H, Omar W, Elkinaai N, Kamel A, Zaki I, et al. Different (18) F-FDG PET Parameters for the Prediction of Histological Response to Neoadjuvant Chemotherapy in Pediatric Ewing Sarcoma Family of Tumors. *Pediatr Blood Cancer* (2020) 67(11):e28605. doi: 10.1002/pbc.28605

178. Annovazzi A, Ferraresi V, Anelli V, Covello R, Vari S, Zoccali C, et al. [(18)F] FDG PET/CT Quantitative Parameters for the Prediction of Histological Response to Induction Chemotherapy and Clinical Outcome in Patients With Localised Bone and Soft-Tissue Ewing Sarcoma. *Eur Radiol* (2021) 31(9):7012–21. doi: 10.1007/s00330-021-07841-w

179. Byun BH, Kong CB, Lim I, Kim BI, Choi CW, Song WS, et al. Early Response Monitoring to Neoadjuvant Chemotherapy in Osteosarcoma Using Sequential (1)(8)F-FDG PET/CT and MRI. *Eur J Nucl Med Mol Imaging* (2014) 41(8):1553–62. doi: 10.1007/s00259-014-2746-2

180. Song H, Jiao Y, Wei W, Ren X, Shen C, Qiu Z, et al. Can Pretreatment (18)F-FDG PET Tumor Texture Features Predict the Outcomes of Osteosarcoma Treated by Neoadjuvant Chemotherapy? *Eur Radiol* (2019) 29(7):3945–54. doi: 10.1007/s00330-019-06074-2

181. Takahashi MES, Mosci C, Souza EM, Brunetto SQ, de Souza C, Pericole FV, et al. Computed Tomography-Based Skeletal Segmentation for Quantitative PET Metrics of Bone Involvement in Multiple Myeloma. *Nucl Med Commun* (2020) 41(4):377–82. doi: 10.1097/MNM.0000000000001165

182. Koizumi M, Motegi K, Umeda T. A Novel Biomarker, Active Whole Skeletal Total Lesion Glycolysis (WS-TLG), as a Quantitative Method to Measure Bone Metastatic Activity in Breast Cancer Patients. *Ann Nucl Med* (2019) 33(7):502–11. doi: 10.1007/s12149-019-01359-4

183. Capobianco N, Meignan M, Cottereau AS, Vercellino L, Sibille L, Spottiswoode B, et al. Deep-Learning (18)F-FDG Uptake Classification Enables Total Metabolic Tumor Volume Estimation in Diffuse Large B-Cell Lymphoma. *J Nucl Med* (2021) 62(1):30–6. doi: 10.2967/jnumed.120.242412

184. Pauwels E, Van Binnebeek S, Vandecaveye V, Baete K, Vanbilloen H, Koole M, et al. Inflammation-Based Index and (68)Ga-DOTATOC PET-Derived Uptake and Volumetric Parameters Predict Outcome in Neuroendocrine Tumor Patients Treated With (90)Y-DOTATOC. *J Nucl Med* (2020) 61(7):1014–20. doi: 10.2967/jnumed.119.236935

185. Okudan B, Coskun N, Seven B, Atalay MA, Yildirim A, Gortan FA. Assessment of Volumetric Parameters Derived From 68Ga-PSMA PET/CT in Prostate Cancer Patients With Biochemical Recurrence: An Institutional Experience. *Nucl Med Commun* (2021) 42(11):1254–60. doi: 10.1097/MNM.0000000000001459

186. Santos A, Mattiolli A, Carvalheira JB, Ferreira U, Camacho M, Silva C, et al. PSMA Whole-Body Tumor Burden in Primary Staging and Biochemical Recurrence of Prostate Cancer. *Eur J Nucl Med Mol Imaging* (2021) 48(2):493–500. doi: 10.1007/s00259-020-04981-x

187. Arvola S, Jambor I, Kuisma A, Kemppainen J, Kajander S, Seppanen M, et al. Comparison of Standardized Uptake Values Between (99m)Tc-HDP SPECT/CT and (18)F-Naf PET/CT in Bone Metastases of Breast and Prostate Cancer. *EJNMMI Res* (2019) 9(1):6. doi: 10.1186/s13550-019-0475-z

188. Fanti S, Goffin K, Hadaschik BA, Herrmann K, Maurer T, MacLennan S, et al. Consensus Statements on PSMA PET/CT Response Assessment Criteria in Prostate Cancer. *Eur J Nucl Med Mol Imaging* (2021) 48(2):469–76. doi: 10.1007/s00259-020-04934-4

189. Imbriaco M, Larson SM, Yeung HW, Mawlawi OR, Erdi Y, Venkatraman ES, et al. A New Parameter for Measuring Metastatic Bone Involvement by Prostate Cancer: The Bone Scan Index. *Clin Cancer Res* (1998) 4(7):1765–72.

190. Dennis ER, Jia X, Mezheritskiy IS, Stephenson RD, Schoder H, Fox JJ, et al. Bone Scan Index: A Quantitative Treatment Response Biomarker for Castration-Resistant Metastatic Prostate Cancer. *J Clin Oncol* (2012) 30(5):519–24. doi: 10.1200/JCO.2011.36.5791

191. Reza M, Ohlsson M, Kaboteh R, Anand A, Franck-Lissbrant I, Damber JE, et al. Bone Scan Index as an Imaging Biomarker in Metastatic Castration-Resistant Prostate Cancer: A Multicentre Study Based on Patients Treated With Abiraterone Acetate (Zytiga) in Clinical Practice. *Eur Urol Focus* (2016) 2(5):540–6. doi: 10.1016/j.euf.2016.02.013

192. Anand A, Tragardh E, Edenbrandt L, Beckman L, Svensson JH, Thellenberg C, et al. Assessing Radiographic Response to (223)Ra With an Automated Bone Scan Index in Metastatic Castration-Resistant Prostate Cancer Patients. *J Nucl Med* (2020) 61(5):671–5. doi: 10.2967/jnumed.119.231100

# Prostate Cancer Gleason Score From Biopsy to Radical Surgery: Can Ultrasound Shear Wave Elastography and Multiparametric Magnetic Resonance Imaging Narrow the Gap?

Cheng Wei[1], Yilong Zhang[2], Xinyu Zhang[3], Wael Ageeli[1,4], Magdalena Szewczyk-Bieda[5], Jonathan Serhan[5], Jennifer Wilson[6], Chunhui Li[2] and Ghulam Nabi[1]*

[1] Division of Imaging Sciences and Technology, School of Medicine, University of Dundee, Dundee, United Kingdom, [2] School of Science and Engineering, University of Dundee, Dundee, United Kingdom, [3] Division of Population Health and Genomics, University of Dundee, Dundee, United Kingdom, [4] Diagnostic Radiology Department, College of Applied Medical Sciences, Jazan University, Jazan, Saudi Arabia, [5] Department of Clinical Radiology, Ninewells Hospital, Dundee, United Kingdom, [6] Department of Pathology, Ninewells Hospital, Dundee, United Kingdom

**Objectives:** To investigate the impact of ultrasound shear wave elastography (USWE) and multiparametric magnetic resonance imaging (mpMRI) in predicting a change in biopsy-assigned Gleason Score (GS) after radical surgery for localised prostate cancer (PCa).

**Method:** A total of 212 men opting for laparoscopic radical prostatectomy (LRP) between September 2013 and June 2017 were recruited into this study. All the participants had 12-core transrectal ultrasound (TRUS) biopsies and imaging using USWE and mpMRI before radical surgery. The predictive accuracy for imaging modalities was assessed in relation to upgrading and downgrading of PCa GS between the biopsies and radical prostatectomy using Student's t-test and multivariable logistic regression analyses. A decision analysis curve was constructed assessing the impact of nomogram on clinical situations using different thresholds of upgrading probabilities.

**Results:** Most GS 6 diseases on biopsies were upgraded on radical surgery (37/42, 88.1%). Major downgrading was seen in GS 8 category of disease (14/35; 37.1%), whereas no alteration was observed in GS 7 on biopsies in most men (55/75; 73.3%). In univariate analysis, higher preoperative prostate-specific antigen (PSA) (p = 0.001), higher prostate-specific antigen density (PSAD) (p = 0.002), stiffer USWE lesions (p = 0.009), and higher prostate imaging–reporting and data system (PIRADS) (p = 0.002) on mpMRI were significant predictors of upgrading. In multivariate logistic regression analyses, only PSA (p = 0.016) and USWE-measured tissue stiffness (p = 0.029) showed statistical significance in predicting upgrading.

**Conclusions:** Measurement of tissue stiffness using USWE in clinically localised PCa can predict upgrading of GS and has the potential to improve patient management options.

Keywords: prostate cancer, ultrasound shear wave elastography, multiparametric MRI, PIRADS, radical prostatectomy, prostate biopsy

# HIGHLIGHTS

○ Ultrasound shear wave elastography can significantly predict upgrading of biopsy-assigned Gleason Score in prostate cancer following radical surgery.

○ Ultrasound shear wave elastography is an emerging technology based on measurement of tissue stiffness.

○ USWE-measured tissue stiffness can impact decision analysis based on different probabilities of Gleason Score upgrading from biopsies to radical surgery in prostate cancer.

# INTRODUCTION

The histological Gleason Score (GS) obtained using 12-core transrectal ultrasound (TRUS) biopsy informs risk stratification and counselling of clinically localised prostate cancer (PCa) patients regarding various treatment options. The approach has significant limitations, as a large discrepancy exists between biopsy and postoperative radical prostatectomy GS in approximately 40% of all localised PCa patients especially for those with biopsy GS 6 disease (1–3). GS upgrading after radical surgery is also associated with poor disease prognostic factors such as extracapsular extension (ECE) and higher rates of biochemical recurrences (4). Thus, predicting GS prior to treatment of PCa becomes crucial, and the role of imaging as marker is less understood. The role of imaging in the detection and characterisation of PCa is now well-established (4). Prebiopsy multiparametric magnetic resonance imaging (mpMRI) has been widely applied to increase biopsy accuracy, particularly over the last decade (5–8). However, there is still a burgeoning interest in investigating the role that imaging can play in predicting underestimated GS in biopsies. This will help in accurately assessing prognosis, treatment selection, and decision-making.

Recently, ultrasound shear wave elastography (USWE) has emerged as a promising imaging modality in the detection and characterisation of localised PCa (9–11). USWE can assess tissue stiffness of the whole prostate including cancerous tissue. USWE measures the shear wave speed generated by specialised ultrasound transducers through the target organs. Under imaging, the speed of these scattered shear waves is shown as a colour-coded dynamic map of tissue stiffness (presented as Young's modulus) in real time (12, 13). The USWE-based imaging approach not only provides characterisation of clinically significant PCa (9) but also predicts biochemical recurrence on follow-up (14).

Although previous studies have focused on other multifactorial analyses and nomograms to predict GS change after radical surgery (3, 15–21), USWE or mpMRI, key imaging

modalities have seldom been considered as potential imaging markers to predict GS upgrading or downgrading in PCa (22).

The aim of this study was to assess the impact of imaging markers [tissue stiffness using USWE and prostate imaging–reporting and data system (PIRADS) using mpMRI] in predicting a change in biopsy-assigned GS after radical surgery for localised prostate cancer. Furthermore, we aimed to quantify the additional benefits that imaging information may bring to the already known and reported clinicopathological parameters through the construction of nomogram and decision-analysis curves.

# MATERIALS AND METHODS

## Study Cohort

Two hundred and twelve patients opting for laparoscopic radical prostatectomy between September 2013 and June 2017 were recruited into this study. All patients were confirmed to have PCa on 12-core TRUS biopsies. In brief, transrectal ultrasound imaging of prostates was performed. After measuring size of prostate gland, local anaesthetic agent was infiltrated from the base to apex. Prostate was divided into 12 regions (as per our protocol) including lateral and paramedian regions. Each region was biopsied and sent for histopathology. Participants were then scanned using two imaging modalities: mpMRI and USWE preoperatively. The images from mpMRI were assessed, and abnormal areas were classified using PIRADS score by two uro-radiologists. The USWE images were analysed, and a quantitative cancer stiffness estimation in kilopascals (kPa) was made. Patients' age at the time of radical surgery, prostate-specific antigen (PSA), prostate weight, prostate-specific antigen density (PSAD), biopsy GS, number of positive cores, maximum of cancer in cores, clinical stage, and postoperatively pathological GS were collected and analysed. Prostate specimens were sectioned in a 3D printed patient specific mould and analysed by two experienced study pathologists including the co-author (JW with more than 5 years' experience) (23, 24). **Figure 1** shows the flow of participants recruited to the study. **Table 1** presents baseline patient characteristics.

## USWE Protocol and Acquisition

All USWE images were obtained using a transrectal endocavitory ultrasound transducer (SuperSonic Imagine, Aix en Provence, France) with patients being in either lithotomy or lateral position the day before the scheduled radical surgery. USWE mode was activated, and prostate gland elastograms were obtained from the cranial to caudal direction for each lobe of the prostate. All regions were scanned as described in our previously published protocol (11). Guidelines for clinical practice have been framed based on data emanating from many centres (25). Each patient's prostate gland was scanned transrectally; USWE images were acquired in transverse planes from the base to apex with a gap of 4–6 mm. The most suspicious cancer located in the planes was marked and reconstructed offline into 3D images. Suspicious areas for cancer were scanned by rotating the transducer in

**FIGURE 1** | Flow chart of study procedure. **(A)** TRUS biopsy result with two positive biopsy cores. **(B)** One suspicious lesion in peripheral zone is shown in MRI (left) and SWE (right) images. **(C)** Post-prostatectomy specimen sectioning in steps (23). **(D)** Histopathology photo after analysing.

different directions to confirm abnormalities and to perform measurements of their sizes. Three stiffness measurements of shear wave speed in m/s or Young's modulus in kPa using pseudo-colour maps were obtained independently by three researchers. The ratio between abnormal and normal areas were also recorded (**Figure 1B**).

## MRI Protocol and PIRAD Score

MRI scan of each patient was performed using 3T scanners (TIM Trio, Siemens, Erlangen, Germany) 6–8 weeks after the prostate biopsy procedure (26). The MRI protocol was derived from the European Society of Uro-radiology (ESUR) guidelines 2012 (27) for PCa detection; PIRADS v2.0 was applied in this study, and only PIRAD ≥3 lesions on MRI were marked and PIRADS 1 and 2 were taken as negative findings. All MR images were analysed and scored by two experienced uro-radiologists (MS-B and JS); both the radiologists were blinded to patients' clinicopathology data.

## Statistical Analyses

A two-stage logistic regression process was used to investigate the explanatory factors that could predict upgrading of GS 6 or 7 (3 + 4) and downgrading of GS 7(4 + 3) and above on biopsy. First, univariate logistic regression was applied to examine associations between single explanatory factor and the outcomes, respectively. Absolute percentages of each variable, univariate odds ratio (OR), 95% CI of univariate OR, and p-value were presented. Multivariate logistic regression was then applied to assess and adjust for significant predictive factors regarding patient characteristics. The predictive factors in the multivariate logistic regression model were a combination of significantly associated factors from the bi-variate logistic regression.

Age, PSA, PSAD, maximum percentage of cancer in the core, and prostate gland weight, and quantitatively assessed stiffness using USWE were treated as continuous variables; clinical stage, number of positive cores for cancer, PIRADS scores, and GS were treated as ordinal variables. The reference groups of those predictive factors were set if they were considered as a meaningful reference of that variable. Adjusted OR, 95% CI of adjusted OR, and p-value were derived after multivariate logistic regression.

In addition, logistic regression model coefficients were used to derive a nomogram predicating the probability of GS upgrading or downgrading from biopsy. Non-informative or non-significant variables in univariate logistic regression for GS upgrading were removed. The bias-corrected calibrated values were generated from internal validation based on 200 bootstrap resamples. A decision-analysis curve was constructed assessing the impact of the nomogram using different threshold probabilities of upgrading or downgrading of GS. All analyses were performed using SPSS 22 (IBM Corporation, New York, USA) and R software (v 3.5.3). The alpha level was set at 0.05 to determine two-tailed significance.

## RESULTS

### Change in GS in the Cohort From Biopsy to Radical Surgery

A detailed map of biopsy GS and radical prostatectomy specimen GS is shown in **Table 2**. A Sankey diagram in **Figure 2** presents same data in an alternate way. No change in Gleason Score was seen in 47.2% of all the cases (100/212). Out of the 42 cases with

**TABLE 1 |** Patient characteristics.

| | |
|---|---|
| Age (years) | |
| Median (IQR) | 67.0 (63.8–72.0) |
| Mean (SD) | 67.2 (5.7) |
| Range | 44.0–77.0 |
| Prostate-specific antigen (ng/ml) | |
| Median (IQR) | 9.4 (7.1–12.5) |
| Mean (SD) | 11.4 (7.6) |
| Range | 0.1–47.7 |
| Clinical stage (%) | |
| ≤T2a | 148 (69.8%) |
| T2b/c | 44 (20.8%) |
| T3 | 20 (9.4%) |
| Biopsy Gleason Score (%) | |
| ≤6 | 42 (19.8%) |
| 7 (3 + 4) | 75 (35.4%) |
| 7 (4 + 3) | 39 (18.4%) |
| >7 | 56 (26.4%) |
| No. of positive cores | |
| Median (IQR) | 4.0 (2.0–7.0) |
| Mean (SD) | 4.8 (3.3) |
| Range | 1.0–14.0 |
| Maximum percentage of cancer pre-core (%) | |
| Median (IQR) | 50.0 (20.0–80.0) |
| Mean (SD) | 50.4 (30.0) |
| Range | 5.0–100.0 |
| The interval from biopsy to SWE (days) | |
| Median (IQR) | 102.5 (83–118) |
| Mean (SD) | 102.6 (27.3) |
| Range | 46–189 |
| The interval from biopsy to MRI (days) | |
| Median (IQR) | 43 (35–48) |
| Mean (SD) | 43.8 (11.2) |
| Range | 21–78 |
| Radical prostatectomy weight (g) | |
| Median (IQR) | 59.5 (47.5–76.5) |
| Mean (SD) | 66.9 (29.4) |
| Range | 31.0–207.0 |

GS 6 disease on biopsy, the majority (37/42, 88.1%) was upgraded following radical surgery. GS 3 + 4 disease on biopsies remained stable in most of the cases (55/75; 73.3%) and so did the GS 9 (4 + 5 or 5 + 4) disease in most cases (76.2%). GS 4 + 3 disease on biopsies had downgrading in one-third of cases (13/39, 33.3%). Most downgrading (13/35; 37.1%) was seen in biopsy GS 8 (3 + 5, 4 + 4, 5 + 3) category disease.

## Multifactorial Analysis of GS Change at Radical Surgery

**Table 3** shows the preoperative clinical and imaging parameters in men with and without upgraded GS at radical surgery. As seen, the data indicate that upgraded patients had a higher PSA level (p = 0.001) and a greater PSAD (p = 0.002), stiffer cancerous tissue as estimated by USWE (p = 0.009), and higher PIRADS 4/5 score (p = 0.002). The results also showed a trend that upgraded patients were older (p = 0.130), with more positive cores (p = 0.608), maximum percentage of cancer in a given core (0.071), and smaller prostates (p = 0.806), but none of these variables were statistically significant. In multivariate logistic regression analyses (**Table 4**), higher stiffness values at USWE (p = 0.029) and higher PSA level (p = 0.016) predicted upgrading from biopsy GS ≤ 7 (3 + 4) to GS ≥ 7 (4 + 3) after radical surgery.

The PIRADS score at mpMRI failed to maintain the same significance (p < 0.05) in both univariate analysis (p = 0.056) and multivariate analysis (p = 0.068).

## Nomogram Construction, Validation, and Defining Thresholds for Decision Analysis

**Figures 3A1, A2** show a constructed nomogram predicting the upgrading of GS from biopsies to radical surgery with or without USWE data. Longer scales indicate a higher percentage of impact, and larger points suggest probability of upgrading. PSA level had the greatest impact in both nomograms. USWE counted as the second highest impact factor for GS upgrading. The nomograms were then internally validated using 200 bootstrap samples, and internal calibration curves were highlighted (**Figures 3B1, B2**). The calibration curves based on internal validation results are set for the probability of prediction at different levels. As seen, the curves demonstrated excellent agreement between the prediction according to the nomogram and actual observation. Decision analysis assumed that the threshold probability of a change in GS at which the clinician or patient would make an informed decision weighing the relative harms of a false-positive and a false-negative prediction using USWE information. A range of threshold probabilities was shown at which the magnitude of benefits of USWE was compared with no USWE information (**Figure 4**). The net benefit for the model using USWE was slightly higher but not quantitively proved at various thresholds compared with the model without USWE (blue *vs.* red line). The mean size of lesions from USWE was 16.1 ± 7.2 mm (range from 7.4 to 44.8 mm).

## DISCUSSION

This was the first study to assess the role of both USWE and mpMRI in predicting change in biopsy-assigned GS following radical surgery in men presenting with clinically localised PCa. A review of the literature showed only limited reports of mpMRI parameters used in predicting GS upgrading. Lai et al. (28) found that mpMRI findings could predict upgrading GS 3 + 3 disease on first biopsies in men on active surveillance. Abd-Alazeez et al. (29) concluded that a patient with higher PIRADS score on mpMRI predicted a high likelihood of high GS disease at radical surgery in men with low-risk PCa in biopsy. No 3D fabricated moulds were used to orient imaging to histopathology in any of the reported studies, a clear contrast to the present study. Similar observations were made by our group in the past (22). In contrast and interestingly, Klotz et al. (30) observed in a randomised multicentre prospective trial that adding MRI to clinicopathological factors did not boost the prediction ability of biopsy-assigned GS. In our study, we observed a high number of patients with PIRADS 4 and 5 in upgraded than not-upgraded patients [28.1% (27/96) *vs.* 5.0% (1/20), p = 0.002], but it was not a significant predictor in either univariate logistic regression model (p = 0.056) or multivariate logistic model (p = 0.068). There are no reports in the literature of USWE imaging being used in predicting change in GS. Previous studies only reported

**TABLE 2** | Radical prostatectomy grades stratified by biopsy Gleason Scores.

| LRP GS | | Biopsy GS | | | | | Total |
|---|---|---|---|---|---|---|---|
| | | <7 | 7 (3 + 4) | 7 (4 + 3) | 8 | 9–10 | |
| 6 | Count | 5 | 0 | 0 | 0 | 0 | 5 |
| | % within LRP GS | 100% | 0.0% | 0.0% | 0.0% | 0.0% | 100% |
| | % within Biopsy GS | 11.9% | 0.0% | 0.0% | 0.0% | 0.0% | 2.4% |
| | % of total | 2.4% | 0.0% | 0.0% | 0.0% | 0.0% | 2.4% |
| 7 (3 + 4) | Count | 29 | 55 | 13 | 5 | 0 | 102 |
| | % within LRP GS | 28.4% | 53.9% | 12.7% | 4.9% | 0.0% | 100% |
| | % within biopsy GS | 69.0% | 73.3% | 33.3% | 14.3% | 0.0% | 48.1% |
| | % of total | 13.7% | 25.9% | 6.1% | 2.4% | 0.0% | 48.1% |
| 7 (4 + 3) | Count | 3 | 5 | 16 | 9 | 2 | 35 |
| | % within LRP GS | 8.6% | 14.3% | 45.7% | 25.7% | 5.7% | 100% |
| | % within biopsy GS | 7.1% | 6.7% | 41.0% | 25.7% | 5.7% | 16.5% |
| | % of total | 1.4% | 2.4% | 7.5% | 4.2% | 0.9% | 16.5% |
| 8 | Count | 1 | 6 | 4 | 8 | 3 | 22 |
| | % within LRP GS | 4.5% | 27.3% | 18.2% | 36.4% | 13.6% | 100% |
| | % within biopsy GS | 2.4% | 8.0% | 10.3% | 22.9% | 14.3% | 10.4% |
| | % of total | 0.5% | 2.8% | 1.9% | 3.8% | 1.4% | 10.4% |
| 9–10 | Count | 4 | 9 | 6 | 13 | 16 | 48 |
| | % within LRP GS | 8.3% | 18.8% | 12.5% | 27.1% | 33.3% | 100% |
| | % within biopsy GS | 9.5% | 12.0% | 15.4% | 37.1% | 76.2% | 22.6% |
| | % of total | 1.9% | 4.2% | 2.8% | 6.1% | 7.5% | 22.6% |
| Total | Count | 42 | 75 | 39 | 35 | 21 | 212 |
| | % within LRP GS | 19.8% | 35.4% | 18.4% | 16.5% | 9.9% | 100% |
| | % within biopsy GS | 100% | 100% | 100% | 100% | 100% | 100% |
| | % of total | 19.8% | 35.4% | 18.4% | 16.5% | 9.9% | 100% |

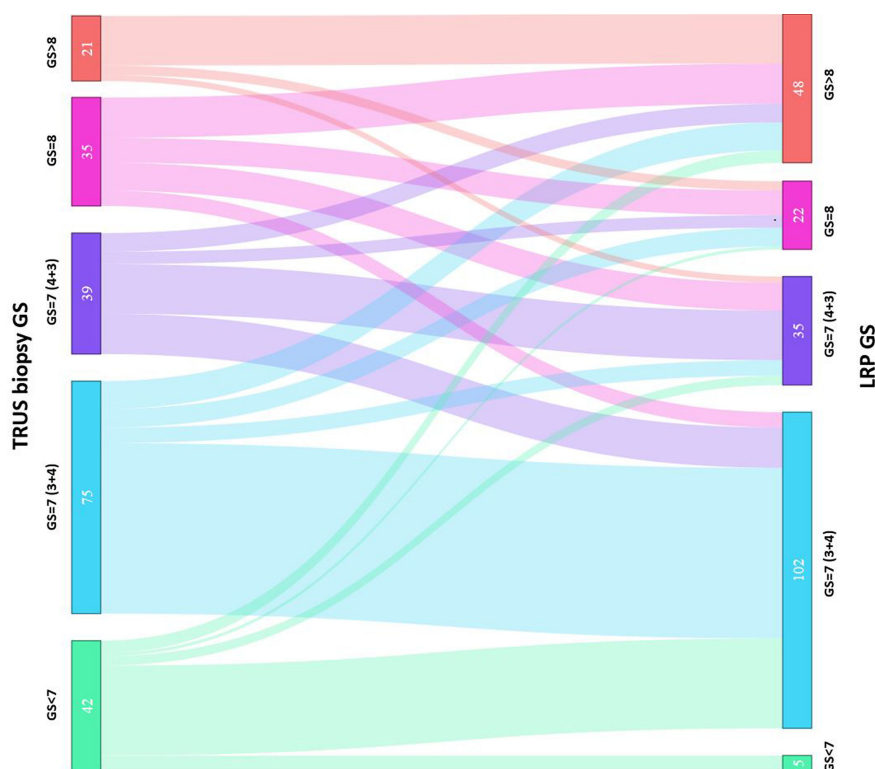*LRP GS: laparoscopic radical prostatectomy Gleason Score; Biopsy GS: biopsy Gleason Score.*



**FIGURE 2** | Sankey diagram of comparison between biopsy Gleason Score and prostatectomy Gleason Score.

**TABLE 3** | Association of clinical and pathologic parameters with Gleason Score (GS) group: upgrading from biopsy GS ≤7(3 + 4) to GS ≥7(4 + 3) at radical prostatectomy.

| variables | Upgrade (n = 28) | No-upgrade (n = 89) | t value (95%CI) | p-value |
|---|---|---|---|---|
| **Age, year** | | | | |
| Median (IQR) | 70.0 (65.0–72.0) | 67.0 (63.0–71.0) | 1.49 (–0.61, 4.14) | 0.130 |
| Mean (SD) | 68.8 (5.5) | 67.0 (5.3) | | |
| **PSA, ng/ml** | | | | |
| Median (IQR) | 11.5 (7.3–16.2) | 8.8 (6.9–10.5) | 3.34 (1.98, 7.78) | 0.001 |
| Mean (SD) | 14.5 (10.1) | 9.6 (5.4) | | |
| **No. of positive cores** | | | | |
| Median (IQR) | 3.0 (1.0–5.0) | 3.0 (2.0–5.0) | 0.47 (–1.07, 1.72) | 0.641 |
| Mean (SD) | 4.0 (3.3) | 3.6 (2.8) | | |
| **Maximum % cancer/core** | | | | |
| Median (IQR) | 45.0(20.0–76.3) | 30.0 (20.0–50.0) | 1.82 (–0.98, 23.02) | 0.071 |
| Mean (SD) | 48.2(32.0) | 37.2 (26.6) | | |
| **Pathology weight (continuous), gram** | | | | |
| Median (IQR) | 62.0 (51.0–76.5) | 63.0 (47.6–86.8) | 1.24 (–5.43, 23.73) | 0.216 |
| Mean (SD) | 64.7 (18.3) | 73.9 (36.8) | | |
| **PSAD (continuous), ng/ml$^2$** | | | | |
| Median (IQR) | 0.2 (0.1–0.2) | 0.1 (0.1–0.2) | 3.21 (0.04, 0.16) | 0.002 |
| Mean (SD) | 0.3 (0.2) | 0.2 (0.10) | | |
| **Clinical stage (%)** | | | | |
| ≤T2a | 16 | 69 | 1.83 (–0.02, 0.51) | 0.070 |
| T2b/c | 9 | 14 | | |
| T3 | 3 | 6 | | |
| **USWE (continuous), kPa** | | | | |
| Median (IQR) | 145.1 (128.8–168.5) | 128.7 (115.3–147.6) | 2.64 (4.98, 34.84) | 0.009 |
| Mean (SD) | 154.2 (42.3) | 134.3 (31.4) | | |
| **PI-RADS** | | | | |
| ≤3 | 1 | 19 | 2.23 (0.02, 0.34) | 0.028 |
| 4 and 5 | 27 | 69 | | |
| Not reported | 0 | 1 | | |

USWE as a promising diagnostic modality in the detection of clinically significant PCa (9–11, 31, 32).

Preoperative PSA levels or PSAD are the most frequently analysed factors as predictor of GS change in the reported literature and were included in this study as well. From the reported publications (19, 21, 28, 33), it appears that PSA or PSAD performed consistently well, although in other studies, the significance was not as strong in comparison to other predictors (34, 35), but all the studies had used preoperative PSA in multivariate logistic regression models. In this study, PSA level

was found to be one of the two significant parameters in multivariate logistic regression analysis.

Smaller prostate size was not a statistically significant predicting factor for GS upgrading in this study (p = 0.086). This is similar to observations by other studies (16, 18–20), although Freedland et al. (36) showed that decreased prostate size was associated with higher Gleason grade, more aggressive behaviour, and higher biochemical recurrence rates.

In studies by Epstein et al. (19) and Gondo et al. (33), age, PSA level, prostate weight, and maximum cancer core

**TABLE 4** | Univariate and multivariate logistic regression models to predict upgrading from biopsy GS ≤7 (3 + 4) to GS ≥7 (4 + 3) at radical prostatectomy.

| | Univariate | | Multivariate | |
|---|---|---|---|---|
| | OR (95% CI) | p-value | OR (95% CI) | p-value |
| Weight | 0.991 (0.976–1.006) | 0.218 | – | – |
| SWE | 1.015 (1.003–1.027) | 0.014 | 1.015 (1.002–1.028) | 0.029 |
| PSA level (ng/ml) | 1.098 (1.026–1.169) | 0.007 | 1.087 (1.016–1.163) | 0.016 |
| PI-RADS | | | | |
| ≤3 | 1 (referent) | – | 1 (referent) | – |
| >3 | 7.435 (0.948–58.305) | 0.056 | 7.317 (0.862–62.097) | 0.068 |
| Positive Core | 1.038 (0.900–1.198) | 0.605 | – | – |
| Percentage | 1.014 (0.999–1.029) | 0.074 | – | – |
| Clinical stage | | | | |
| T3 | 1 (referent) | – | – | – |
| T2b/c | 2.156 (0.487–9.556) | 0.312 | – | – |
| ≤T2a | 0.778 (0.154–3.927) | 0.761 | – | – |

*OR: odd ratio; Univariate and Multivariate analysis are two statistical analyses. Univariate involves the analysis of a single variable while multivariate analysis examines two or more variables. Most multivariate analysis involves a dependent variable and multiple independent variables.*
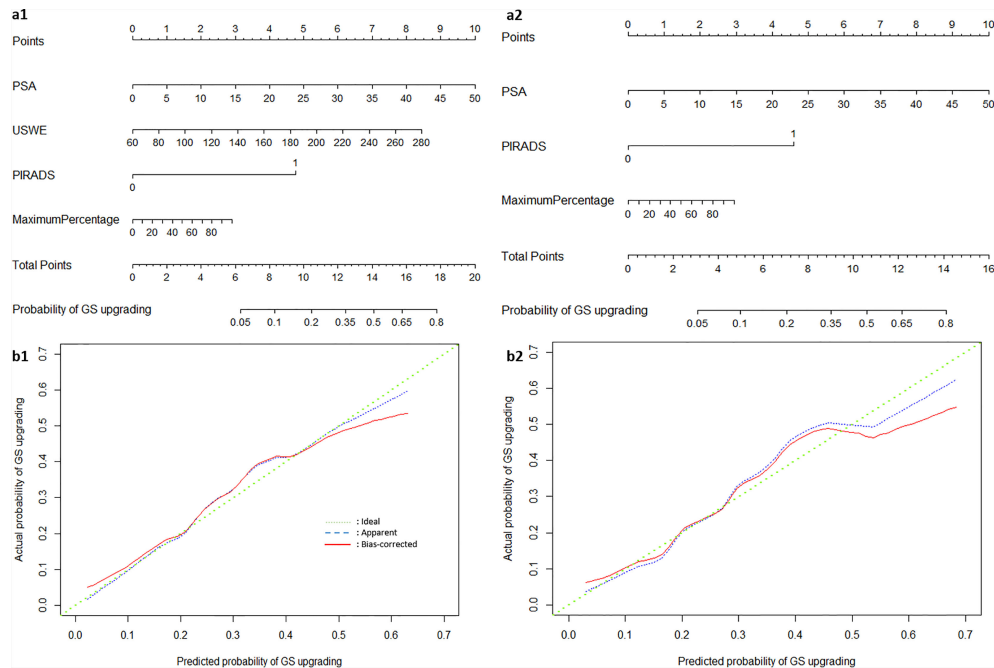
**FIGURE 3** | The nomograms of Gleason Score upgrading prediction with **(A1)** and without USWE score **(A2)**. Calibration plots of observed and predicted probability of GS upgrading with **(B1)** and without USWE score **(B2)**.
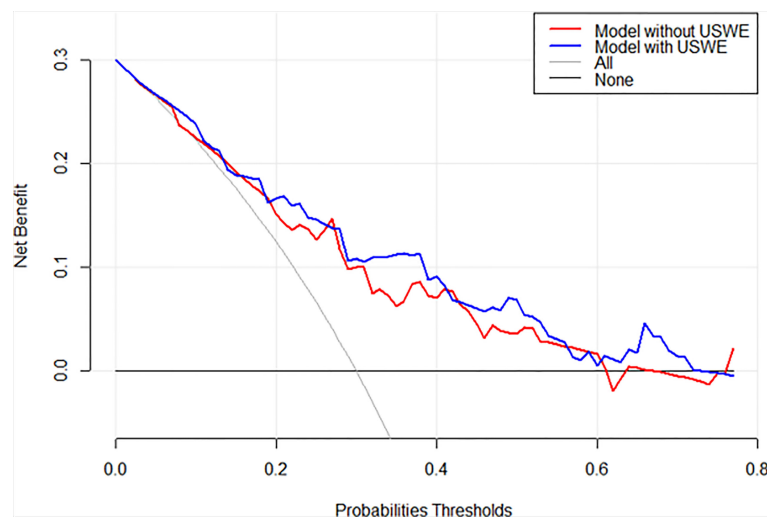


**FIGURE 4** | Decision analysis demonstrated a high net benefit of USWE score model across a wide range of threshold probabilities. Prediction model without USWE score (red line); prediction model with USWE score (blue line).

involvement were all statistically significant predictors of downgrading. In our study, downgraded patients were more likely to have a lower PSA level (8.1 *vs.* 10.6 ng/ml), but this was not statistically significant (p = 0.075). Three studies summarised downgrading from biopsy GS 3 + 4 to biopsy GS <7 at varying rates of 7.3%, 9.0%, and 12.0% (19, 33, 37), respectively. No

patients' postoperative GS was downgraded to GS 3 + 3 in our study (see first raw of **Table 2**).

There were limitations worth mentioning in this study. First, this study recruited men with histologically confirmed PCa and only those opting for radical surgery. The focus of the study was to obtain a robust reference standard of histology from radical

prostatectomy obtained specimens. We used both USWE and MRI imaging modalities in a preoperative setting. This was considered as the standard of care approach at the time of study; however, this had the potential of introducing a selection bias in the study. Second, the biopsy technique used in this study was 12-cores TRUS biopsy without targeting, and this has potentially created sampling error in the patients recruited to this study. Third, MRI scans were obtained after biopsies confirmed PCa, and this might have introduced a detection bias in estimating PIRADS score. Finally, this was a single institutional study, and the findings require external validation and the reproducibility of the USWE technique (38). The use of USWE is not the standard of care, although guidelines and evidence are emerging in this area (39). The study was single centred with only operator performed USWE. Further reproducibility in multi-operator setting needs to be tested. We did not calculate the learning curve for this technology. Future studies could focus on the role of both USWE and/or MRI-targeted biopsy in patients suspected of PCa and in predicting change in GS from biopsies to radical surgery.

Measurement of tissue stiffness using USWE in clinically localised PCa can predict upgrading of GS and better guide patient management options. This information may help in counselling patients opting for PCa therapy for localised disease.

## DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by East of Scotland Research Ethics Service (EoSRES) REC 1. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

GN and CW contributed to the conception and design of the study. CW and YZ organised the database. YZ and XZ performed the statistical analysis. CW wrote the first draft of the manuscript. CW, WA, MS-B, JS, JW, and CL corrected and rewrote sections of the manuscript. All authors contributed to the article and approved the submitted version.

## REFERENCES

1. Kvåle R, Møller B, Wahlqvist R, Fosså SD, Berner A, Busch C, et al. Concordance Between Gleason Scores of Needle Biopsies and Radical Prostatectomy Specimens: A Population-Based Study. *BJU Int* (2009) 103 (12):1647–54. doi: 10.1111/j.1464-410X.2008.08255.x

2. Fernandes ET, Sund Aram CP, Long R, Soltani M, Ercole CJ. Biopsy Gleason Score: How Does it Correlate With the Final Pathological Diagnosis in Prostate Cancer? *Br J Urol* (1997) 79(4):615–7. doi: 10.1046/j.1464-410X.1997.00126.x

3. Calio BP, Sidana A, Sugano D, Gaur S, Maruf M, Jain AL, et al. Risk of Upgrading From Prostate Biopsy to Radical Prostatectomy Pathology—Does Saturation Biopsy of Index Lesion During Multiparametric Magnetic Resonance Imaging-Transrectal Ultrasound Fusion Biopsy Help? *J Urol* (2018) 199(4):976–82. doi: 10.1016/j.juro.2017.10.048

4. Pinthus JH, Witkos M, Fleshner NE, Sweet J, Evans A, Jewett MA, et al. Prostate Cancers Scored as Gleason 6 on Prostate Biopsy are Frequently Gleason 7 Tumors at Radical Prostatectomy: Implication on Outcome. *J Urol* (2006) 176(3):979–84. doi: 10.1016/j.juro.2006.04.102

5. Anastasiadis AG, Lichy MP, Nagele U, Kuczyk MA, Merseburger AS, Hennenlotter J, et al. MRI-Guided Biopsy of the Prostate Increases Diagnostic Performance in Men With Elevated or Increasing PSA Levels After Previous Negative TRUS Biopsies. *Eur Urol* (2006) 50(4):738–49. doi: 10.1016/j.eururo.2006.03.007

6. Ehdaie B, Shariat SF. Magnetic Resonance Imaging-Targeted Prostate Biopsy: Back to the Future. *Eur Urol* (2013) 63(1):141–2; discussion 143-4. doi: 10.1016/j.eururo.2012.06.049

7. Moore CM, Robertson NL, Arsanious N, Middleton T, Villers A, Klotz L, et al. Image-Guided Prostate Biopsy Using Magnetic Resonance Imaging-Derived Targets: A Systematic Review. *Eur Urol* (2013) 63(1):125–40. doi: 10.1016/j.eururo.2012.06.004

8. 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

9. Wei C, Li C, Szewczyk-Bieda M, Upreti D, Lang S, Huang Z, et al. Performance Characteristics of Transrectal Shear Wave Elastography Imaging in the Evaluation of Clinically Localized Prostate Cancer: A Prospective Study. *J Urol* (2018) 200(3):549–58. doi: 10.1016/j.juro.2018.03.116

10. Sigrist RMS, Liau J, Kaffas AE, Chammas MC, Willmann JK. Ultrasound Elastography: Review of Techniques and Clinical Applications. *Theranostics* (2017) 7(5):1303–29. doi: 10.7150/thno.18650

11. Ahmad S, Cao R, Varghese T, Bidaut L, Nabi G. Transrectal Quantitative Shear Wave Elastography in the Detection and Characterisation of Prostate Cancer. *Surg Endosc* (2013) 27(9):3280–7. doi: 10.1007/s00464-013-2906-7

12. Bercoff J, Tanter M, Fink M. Supersonic Shear Imaging: A New Technique for Soft Tissue Elasticity Mapping. *IEEE Trans Ultrasonics Ferroelectrics Frequency Control* (2004) 51(4):396–409. doi: 10.1109/TUFFC.2004.1295425

13. Bercoff J, Chaffai S, Tanter M, Sandrin L, Catheline S, Fink M, et al. *In Vivo* Breast Tumor Detection Using Transient Elastography. *Ultrasound Med Biol* (2003) 29(10):1387–96. doi: 10.1016/S0301-5629(03)00978-5

14. Wei C, Zhang Y, Malik H, Zhang X, Alqahtani S, Upreti D, et al. Prediction of Postprostatectomy Biochemical Recurrence Using Quantitative Ultrasound Shear Wave Elastography Imaging. *Front Oncol* (2019) 9(572). doi: 10.3389/fonc.2019.00572

15. Imamoto T, Suzuki H, Utsumi T, Takano M, Suyama T, Kawamura K, et al. External Validation of a Nomogram Predicting the Probability of Prostate Cancer Gleason Sum Upgrading Between Biopsy and Radical Prostatectomy Pathology Among Japanese Patients. *Urology* (2010) 76(2):404–10. doi: 10.1016/j.urology.2009.05.084

16. Davies JD, Aghazadeh MA, Phillips S, Salem S, Chang SS, Clark PE, et al. Prostate Size as a Predictor of Gleason Score Upgrading in Patients With Low Risk Prostate Cancer. *J Urol* (2011) 186(6):2221–7. doi: 10.1016/j.juro.2011.07.104

17. Corcoran NM, Hong MKH, Casey RG, Hurtado-Coll A, Peters J, Harewood L, et al. Upgrade in Gleason Score Between Prostate Biopsies and Pathology Following Radical Prostatectomy Significantly Impacts Upon the Risk of Biochemical Recurrence. *BJU Int* (2011) 108(8b):E202–10. doi: 10.1111/j.1464-410X.2011.10119.x

18. Tilki D, Schlenker B, John M, Buchner A, Stanislaus P, Gratzke C, et al. Clinical and Pathologic Predictors of Gleason Sum Upgrading in Patients

After Radical Prostatectomy: Results From a Single Institution Series. *Urologic Oncol: Semin Original Investigations* (2011) 29(5):508–14. doi: 10.1016/j.urolonc.2009.07.003

19. Epstein JI, Feng Z, Trock BJ, Pierorazio PM. Upgrading and Downgrading of Prostate Cancer From Biopsy to Radical Prostatectomy: Incidence and Predictive Factors Using the Modified Gleason Grading System and Factoring in Tertiary Grades. *Eur Urol* (2012) 61(5):1019–24. doi: 10.1016/j.eururo.2012.01.050

20. Gershman B, Dahl DM, Olumi AF, Young RH, McDougal WS, Wu C-L, et al. Smaller Prostate Gland Size and Older Age Predict Gleason Score Upgrading. *Urologic Oncol: Semin Original Investigations* (2013) 31(7):1033–7. doi: 10.1016/j.urolonc.2011.11.032

21. Xu N, Wu Y-P, Li X-D, Lin M-Y, Zheng Q-S, Chen S-H, et al. Risk of Upgrading From Prostate Biopsy to Radical Prostatectomy Pathology: Is Magnetic Resonance Imaging-Guided Biopsy More Accurate? *J Cancer* (2018) 9(19):3634–9. doi: 10.7150/jca.26791

22. Alqahtani S, Wei C, Zhang Y, Szewczyk-Bieda M, Wilson J, Huang Z, et al. Prediction of Prostate Cancer Gleason Score Upgrading From Biopsy to Radical Prostatectomy Using Pre-Biopsy Multiparametric MRI PIRADS Scoring System. *Sci Rep* (2020) 10(1):7722. doi: 10.1038/s41598-020-64693-y

23. Sheikh N, Wei C, Szewczyk-Bieda M, Campbell A, Memon S, Lang S, et al. Combined T2 and Diffusion-Weighted MR Imaging With Template Prostate Biopsies in Men Suspected With Prostate Cancer But Negative Transrectal Ultrasound-Guided Biopsies. *World J Urol* (2017) 35(2):213–20. doi: 10.1007/s00345-016-1855-x

24. Wei C, Lang S, Bidaut L, Doull R, Huang Z, Nabi G, et al. Computer Aided Image Analysis and Rapid Prototyping Molds Using Patient-Specific MRI Data for Reliable Comparison Between Imaging and Histopathology of Radical Prostatectomy Specimens. *Br J Surg* (2014) 101:67–7.

25. Barr RG, Cosgrove D, Brock M, Cantisani V, Correas JM, Postema AW, et al. WFUMB Guidelines and Recommendations on the Clinical Use of Ultrasound Elastography: Part 5. Prostate. *Ultrasound Med Biol* (2017) 43(1):27–48.

26. Choi MH, Jung SE, Park YH, Lee JY, Choi Y-J. Multiparametric MRI of Prostate Cancer After Biopsy: Little Impact of Hemorrhage on Tumor Staging. *Investig Magn Reson Imaging* (2017) 21(3):139–47. doi: 10.13104/imri.2017.21.3.139

27. Barentsz J, Richenberg J, Clements R, Choyke P, Verma S, Villeirs G, et al. ESUR Prostate MR Guidelines 2012. *Eur Radiol* (2012) 22(4):746–57. doi: 10.1007/s00330-011-2377-y

28. Lai WS, Gordetsky JB, Thomas JV, Nix JW, Rais-Bahrami S. Factors Predicting Prostate Cancer Upgrading on Magnetic Resonance Imaging–Targeted Biopsy in an Active Surveillance Population. *Cancer* (2017) 123(11):1941–8. doi: 10.1002/cncr.30548

29. Abd-Alazeez M, Ahmed HU, Arya M, Allen C, Dikaios N, Freeman A, et al. Can Multiparametric Magnetic Resonance Imaging Predict Upgrading of Transrectal Ultrasound Biopsy Results at More Definitive Histology? *Urologic Oncol: Semin Original Investigations* (2014) 32(6):741–7.

30. Klotz L, Loblaw A, Sugar L, Moussa M, Berman DM, Van der Kwast T, et al. Active Surveillance Magnetic Resonance Imaging Study (ASIST): Results of a Randomized Multicenter Prospective Trial. *Eur Urol* (2019) 75(2):300–9. doi: 10.1016/j.eururo.2018.06.025

31. Woo S, Kim SY, Cho JY, Kim SH. Shear Wave Elastography for Detection of Prostate Cancer: A Preliminary Study. *Korean J Radiol* (2014) 15(3):346–55. doi: 10.3348/kjr.2014.15.3.346

32. Boehm K, Salomon G, Beyer B, Schiffmann J, Simonis K, Graefen M, et al. Shear Wave Elastography for Localization of Prostate Cancer Lesions and Assessment of Elasticity Thresholds: Implications for Targeted Biopsies and Active Surveillance Protocols. *J Urol* (2015) 193(3):794–800. doi: 10.1016/j.juro.2014.09.100

33. Gondo T, Poon BY, Matsumoto K, Bernstein M, Sjoberg DD, Eastham JA. Clinical Role of Pathological Downgrading After Radical Prostatectomy in Patients With Biopsy Confirmed Gleason Score 3 + 4 Prostate Cancer. *BJU Int* (2015) 115(1):81–6. doi: 10.1111/bju.12769

34. Magheli A, Hinz S, Hege C, Stephan C, Jung K, Miller K, et al. Prostate Specific Antigen Density to Predict Prostate Cancer Upgrading in a Contemporary Radical Prostatectomy Series: A Single Center Experience. *J Urol* (2010) 183(1):126–32. doi: 10.1016/j.juro.2009.08.139

35. Visapää H, Hotakainen K, Lundin J, Ala-Opas M, Stenman UH. The Proportion of Free PSA and Upgrading of Biopsy Gleason Score After Radical Prostatectomy. *Urologia Internationalis* (2010) 84(4):378–81. doi: 10.1159/000296302

36. Freedland SJ, Isaacs WB, Platz EA, Terris MK, Aronson WJ, Amling CL, et al. Prostate Size and Risk of High-Grade, Advanced Prostate Cancer and Biochemical Progression After Radical Prostatectomy: A Search Database Study. *J Clin Oncol* (2005) 23(30):7546–54. doi: 10.1200/JCO.2005.05.525

37. Moussa AS, Kattan MW, Berglund R, Yu C, Fareed AK, Jones JS. A Nomogram for Predicting Upgrading in Patients With Low- and Intermediate-Grade Prostate Cancer in the Era of Extended Prostate Sampling. *BJU Int* (2010) 105(3):352–8. doi: 10.1111/j.1464-410X.2009.08778.x

38. Harvey H, Morgan V, Fromageau J, O'Shea T, Bamber J, deSouza NM. Ultrasound Shear Wave Elastography of the Normal Prostate: Interobserver Reproducibility and Comparison With Functional Magnetic Resonance Tissue Characteristics. *Ultrason Imaging* (2018) 40(3):158–70. doi: 10.1177/0161734618754487

39. Barr RG. Just Because There is a Button Doesn't Mean it is Appropriate to Use or That it Will Give Accurate Information. *J Ultrasound Med* (2019) 0(0):150–70. doi: 10.1002/jum.14990

# Twenty Years On: RECIST as a Biomarker of Response in Solid Tumours an EORTC Imaging Group – ESOI Joint Paper

Laure Fournier[1,2,3*], Lioe-Fee de Geus-Oei[1,4,5], Daniele Regge[2,6,7], Daniela-Elena Oprea-Lager[1,8], Melvin D'Anastasi[2,9], Luc Bidaut[1,10], Tobias Bäuerle[2,11], Egesta Lopci[1,12], Giovanni Cappello[6,7], Frederic Lecouvet[1,13], Marius Mayerhoefer[2,14,15], Wolfgang G. Kunz[1,2,16], Joost J. C. Verhoeff[1,17], Damiano Caruso[2,18], Marion Smits[1,19,20], Ralf-Thorsten Hoffmann[2,21], Sofia Gourtsoyianni[2,22], Regina Beets-Tan[2,23,24], Emanuele Neri[2,25], Nandita M. deSouza[1,26,27,28], Christophe M. Deroose[1,29,30] and Caroline Caramella[1,31]

[1] Imaging Group, European Organisation of Research and Treatment in Cancer (EORTC), Brussels, Belgium, [2] European Society of Oncologic Imaging (ESOI), European Society of Radiology, Vienna, Austria, [3] Université de Paris, Assistance Publique–Hôpitaux de Paris (AP-HP), Hopital europeen Georges Pompidou, Department of Radiology, Paris Cardiovascular Research Center (PARCC) Unité Mixte de Recherche (UMRS) 970, Institut national de la santé et de la recherche médicale (INSERM), Paris, France, [4] Department of Radiology, Leiden University Medical Center, Leiden, Netherlands, [5] Biomedical Photonic Imaging Group, University of Twente, Enschede, Netherlands, [6] Department of Surgical Sciences, University of Turin, Turin, Italy, [7] Radiology Unit, Candiolo Cancer Institute, Fondazione del Piemonte per l'Oncologia-Istituto Di Ricovero e Cura a Carattere Scientifico (FPO-IRCCS), Turin, Italy, [8] Department of Radiology & Nuclear Medicine, Cancer Centre Amsterdam, Amsterdam University Medical Centers [Vrije Universiteit (VU) University], Amsterdam, Netherlands, [9] Medical Imaging Department, Mater Dei Hospital, University of Malta, Msida, Malta, [10] College of Science, University of Lincoln, Lincoln, United Kingdom, [11] Institute of Radiology, University Hospital Erlangen, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Erlangen, Germany, [12] Nuclear Medicine Unit, Istituto Di Ricovero e Cura a Carattere Scientifico (IRCCS) – Humanitas Research Hospital, Milan, Italy, [13] Department of Radiology, Institut de Recherche Expérimentale et Clinique (IREC), Cliniques Universitaires Saint Luc, Université Catholique de Louvain (UCLouvain), Brussels, Belgium, [14] Department of Radiology, Memorial Sloan Kettering Cancer Center, New York, NY, United States, [15] Department of Biomedical Imaging and Image-guided Therapy, Medical University of Vienna, Vienna, Austria, [16] Department of Radiology, University Hospital, Ludwig Maximilian University (LMU) Munich, Munich, Germany, [17] Department of Radiotherapy, University Medical Center Utrecht, Utrecht University, Utrecht, Netherlands, [18] Department of Medical-Surgical Sciences and Translational Medicine, Sapienza University of Rome, Rome, Italy, [19] Department of Radiology & Nuclear Medicine, Erasmus MC, University Medical Centre Rotterdam, Rotterdam, Netherlands, [20] Brain Tumour Centre, Erasmus Medical Centre (MC) Cancer Institute, Rotterdam, Netherlands, [21] Institute and Policlinic for Diagnostic and Interventional Radiology, University Hospital, Carl-Gustav-Carus Technical University Dresden, Dresden, Germany, [22] Department of Radiology, School of Medicine, National and Kapodistrian University of Athens, Areteion Hospital, Athens, Greece, [23] Department of Radiology, The Netherlands Cancer Institute, Amsterdam, Netherlands, [24] School For Oncology and Developmental Biology (GROW) School for Oncology and Developmental Biology, Maastricht University, Maastricht, Netherlands, [25] Diagnostic and Interventional Radiology, Department of Translational Research and of New Surgical and Medical Technologies, University of Pisa, Pisa, Italy, [26] Division of Radiotherapy and Imaging, The Institute of Cancer Research and Royal Marsden National Health Service (NHS) Foundation Trust, London, United Kingdom, [27] European Imaging Biomarkers Alliance (EIBALL), European Society of Radiology, Vienna, Austria, [28] Quantitative Imaging Biomarkers Alliance, Radiological Society of North America, Oak Brook, IL, United States, [29] Nuclear Medicine, University Hospitals Leuven, Leuven, Belgium, [30] Nuclear Medicine & Molecular Imaging, Department of Imaging and Pathology, Katholieke Universiteit (KU) Leuven, Leuven, Belgium, [31] Radiology Department, Hôpital Marie Lannelongue, Groupe Hospitalier Paris Saint Joseph Centre International des Cancers Thoraciques, Université Paris-Saclay, Le Plessis-Robinson, France

Response evaluation criteria in solid tumours (RECIST) v1.1 are currently the reference standard for evaluating efficacy of therapies in patients with solid tumours who are included in clinical trials, and they are widely used and accepted by regulatory agencies. This expert statement discusses the principles underlying RECIST, as well as their reproducibility and limitations. While the RECIST framework may not be perfect, the scientific bases for the anticancer drugs that have been approved using a RECIST-based surrogate endpoint remain valid. Importantly, changes in measurement have to meet thresholds defined by RECIST for response classification within thus partly circumventing the problems of measurement variability. The RECIST framework also applies to clinical patients in individual settings even though the relationship between tumour size changes and outcome from cohort studies is not necessarily translatable to individual cases. As reproducibility of RECIST measurements is impacted by reader experience, choice of target lesions and detection/interpretation of new lesions, it can result in patients changing response categories when measurements are near threshold values or if new lesions are missed or incorrectly interpreted. There are several situations where RECIST will fail to evaluate treatment-induced changes correctly; knowledge and understanding of these is crucial for correct interpretation. Also, some patterns of response/progression cannot be correctly documented by RECIST, particularly in relation to organ-site (e.g. bone without associated soft-tissue lesion) and treatment type (e.g. focal therapies). These require specialist reader experience and communication with oncologists to determine the actual impact of the therapy and best evaluation strategy. In such situations, alternative imaging markers for tumour response may be used but the sources of variability of individual imaging techniques need to be known and accounted for. Communication between imaging experts and oncologists regarding the level of confidence in a biomarker is essential for the correct interpretation of a biomarker and its application to clinical decision-making. Though measurement automation is desirable and potentially reduces the variability of results, associated technical difficulties must be overcome, and human adjudications may be required.

Keywords: tumour, biomarker, imaging, response, RECIST

# INTRODUCTION

Imaging plays a major role in the evaluation of tumour response to cancer treatments. It provides an objective *in-vivo* measurement of tumour burden, and helps oncologists determine whether a treatment should be pursued, interrupted or adapted.

Response evaluation criteria in solid tumours (RECIST) v1.1 currently is the reference standard for evaluating efficacy of therapies in patients with solid tumours who are included in clinical trials, and it is widely used and accepted by regulatory agencies (1). However, many publications question both the reproducibility and the clinical relevance of RECIST. This paper is an expert statement aiming to answer some of the questions regarding the principles underlying RECIST and its reproducibility compared to other biomarkers, as well as the limitations to its application and continued role in an era where other biomarkers exist that are more explicitly geared towards tumour-specific properties.

# HOW WERE RECIST THRESHOLDS ESTABLISHED?

RECIST has instituted several overarching principles underpinning its approach to tumour response evaluation. Primarily, RECIST defines which lesions are measurable in a reliable manner. Among these, it defines a maximal number of lesions ('target lesions') to be measured to yield a quantitative value representative of tumour burden. The remainder are considered 'non-target lesions' and are evaluated qualitatively. On follow-up scans, new lesions indicate progression (**Table 1**). The threshold for response is defined as a decrease of at least 30% of sum of diameters (SOD) of target lesions compared to baseline, AND no progression of non-target lesions AND no new lesions. The threshold for progressive disease (PD) is defined as an increase of at least 20% of SOD of target lesions compared to nadir AND/OR unequivocal progression of non-target lesions AND/OR appearance of new lesions.

The first publication addressing thresholds for determining treatment efficacy was published by Moertel and Hanley in 1976

**TABLE 1 |** RECIST categories of response.

| Overall Response | Target Lesions | Non Target Lesions | New Lesions |
|---|---|---|---|
| **Definition** | •Lesions with longestdiameter≥10 mm and limits that are sufficiently well defined for their measurement to be considered reliable<br>•Lymph nodes: measurement of short axis, target lesion if short-axis measures≥15 mm<br>• Maximum number of selected target lesions 5/patient and 2/organ | •Lesions that are too small (< 10 mm)<br>•Lesions for which measurement is considered unreliable as their limits are difficult to define (bone or leptomeningeal lesions, ascites, pleural or pericardial effusion, lymphangitic carcinomatosis etc.)<br>•Measurable lesions not selected as target lesions<br>•Lymph nodes: measurement of short axis, non-target lesion if 10 mm ≤ short-axis diameter < 15 mm<br>•Levels of tumour markers > normal (if relevant and predefined) | |
| **Complete response (CR)** | • Disappearance of all target lesions and all nodes have short axis < 10 mm | • Disappearance of all non-target lesions and normalisation of tumour marker levels | • No |
| **Partial response (PR)** | •≥ 30% decrease in the sum of target lesions taking as reference the baseline sum | •No progression | • No |
| **Stable disease (SD)** | •Neither response nor progression | • Persistence of one or more non-target lesions and/or tumour marker levels > normal | • No |
| **Progressive disease (PD):** response is PD if at least one category of lesions meets progression criteria | •≥ 20% increase in the sum of target lesions taking as reference the smallest sum measured during follow-up (nadir) and ≥ 5 mm in absolute value | • 'Unequivocal' progression (assessed qualitatively) in lesion size (an increase in size of a single lesion is not sufficient) | • Yes [appearance of new unequivocally metastatic lesion(s)] |

(2). In this study, 16 observers were asked to measure by clinical examination using a calliper the diameters of solid spheres of variable sizes arranged randomly underneath a mattress. Authors suggested the product of two diameters should be used, as this would be more reliable if lesions were not spherical. For this 'clinical' estimate, a 50% reduction in the product of two diameters was shown to have an acceptable measurement error estimated between 7-8%. Interestingly, the authors specifically stated that "the purpose was not to predict long-term efficacy but to determine what change in bidimensional size could be confidently considered a change". Progression, on the other hand, was defined as an increase in the product of diameters ≥ 25%, but the authors could not justify this cut-off, other than by specifying it "should not necessarily be regarded as influencing the management of the patient".

In 1979, the World Health Organization (WHO) provided recommendations for the evaluation of cancer treatments in clinical trials on imaging. Criteria were based not only on the bidimensional measurement of lesions on clinical examination, but also CT or standard radiography (3), transposing results of Moertel and Hanley's study and setting cut-offs for definition of

response to -50% and of progression to +25%. However, many technical aspects were not detailed, such as the number of lesions to be measured or what constituted a measurable lesion.

In 2000, a working group of European, American and Canadian cancer research organizations (EORTC, NCI, NCIC) defined the Response Evaluation Criteria In Solid Tumours – RECIST (4). They used data from over 4,600 patients enrolled in 14 clinical trials to formulate criteria based on imaging. RECIST used unidimensional measurement of lesions, justified by an extensive comparison of methods of measurement (1D *vs.* 2D) (5). Moreover, this working group specified conditions of measurement, number of lesions, and detailed how to document progression. Regarding cut-off values for response and progression, the -50% value for response for bidimensional measurement was altered to -30% for unidimensional measurements, and the +25% value for progression for bidimensional measurement was altered to +20% for unidimensional measurements (**Table 2**).

RECIST was then revised in 2009 (version 1.1) (1), introducing specific rules for measurement of small axis of lymph nodes and reducing the number of target lesions to five

**TABLE 2 |** Relationship between diameter and corresponding volume.

| Diameter ("long axis") | Percentage of variation | Corresponding volume | Percentage of variation |
|---|---|---|---|
| 20 mm | | 4.2 cm3 | |
| 26 mm | +30% | 9.2 cm3 | +120% |
| 34 mm | +30% | 20.6 cm3 | + 120% |
| 27 mm | -20% | 10.3 cm3 | -50% |

*Repeated measurements are given for a theoretical lesion including diameter measured in a single dimension (long axis), percent changes between measurements, and the corresponding volume assuming the lesion is a sphere and percent changes in volume.*

per patient. This new version was also based on data analysis, including a literature review and a simulation using a database of over 6,500 patients and 18,000 lesions. The number of target lesions for example, was chosen by determining the minimum number for which response rates and time to progression were not altered from RECIST 1.0 results (6, 7).

## Statement #1

*RECIST thresholds were chosen to produce a comparable classification of patients in a given category of response when comparing trials or even when comparing patients, taking into account tumour measurement variability.*

## DO RECIST CATEGORIES PREDICT OUTCOME?

RECIST criteria were originally tested and validated to provide an objective and reproducible assessment of treatment effect in cancer patients, without any references to patient outcome (8). Yet it seems intuitive that when a tumour decreases in size, a patient will have a better outcome, and vice versa. There is evidence to support this, including some large studies, which pool data from various trials. In over 500 patients with metastatic colorectal cancer treated with combination chemotherapy, a decrease in size resulted in a decreased hazard ratio for overall survival (OS) (9). In a meta-analysis of 24 phase I trials, a linear relationship was shown between change in tumour size and survival (10). In a pooled analysis of over 2,700 patients with metastatic renal cell carcinoma treated with anti-angiogenic agents, tumour shrinkage of ≥ 30% resulted in improved OS and progression-free survival (PFS) (11). In addition, the authors demonstrated that tumour shrinkage between 60% and 100% at 6-month follow-up represented an independent prognostic factor for OS. Litière et al. also demonstrated in an even larger pooled analysis of over 23,000 patients treated with targeted agents, chemotherapy or a combination thereof (12), that a decrease in tumour size was consistently associated with a lower hazard ratio, while an increase in size was associated with a higher hazard ratio.

Tumour response according to RECIST can only be quantified by a decrease in size or number of target lesions, as non-target lesions are not taken into consideration for partial response (PR). Regarding progression however, it is important to consider non-target lesions, as unequivocal progression of non-target lesions or emergence of new lesions defines tumour progression. In over 3,700 patients from 13 trials in the RECIST trial database, the presence of new lesions and progression of non-target lesions were most strongly associated with worse OS (hazard ratios range 1.5–2.3) regardless of tumour type, whereas percentage tumour growth in target lesions contributed less in a multivariate model of OS (13).

Finally, in two separate studies (14, 15), An et al. compared the predictive ability of RECIST categories vs. longitudinal tumour measurement–based continuous metrics and alternative categorical response metrics such as slope (absolute change in tumour size) and percent change (relative change in tumour size) to predict OS. The databases consisted respectively of almost 2,100 patients from 13 trials and over 1,500 patients from 3 trials with breast cancer, non–small cell lung cancer (NSCLC) or colorectal cancer. Although there seemed to be a slightly better performance for continuous variables, it was not statistically significant, which led the authors to conclude there was no evidence that growth rate or a continuous evaluation of percent change would improve prediction of outcome. However, it may be noted that timing of evaluations, particularly when considering non-continuous variables, may have an impact on their performance and results.

## Statement #2

*Tumour size changes correlate to outcome at a statistical (cohort) level.*

## HOW REPRODUCIBLE IS RECIST?

When considering whether RECIST evaluates tumour response correctly, metrology principles guide us to consider two aspects (16): is the measurement "true" (when compared to a "real" value, which defines its accuracy), and is the measurement "precise" (i.e. repeatable and reproducible)?

Assessing accuracy of change in size measurements would require obtaining "true" values of change in size. As it is not possible to surgically excise all tumours for comparison with imaging, and often inaccurate to compare *ex vivo* with *in vivo* measurements, the true value of an imaging biomarker must be derived from data obtained through a combination of primary tumour excision and phantom studies.

Precision refers to the variability of the measurement process and can be evaluated by repeatability (when measurement conditions do not change) and reproducibility (when measurement conditions vary). The precision of RECIST and of response categories has been studied extensively. **Table 3** lists the documented reproducibility of RECIST and factors that may impact it. Overall, SOD reproducibility is in the order of +/-20% in multi-observer studies, and +/-10% in single observer studies (17). Important factors associated with RECIST measurement reproducibility are the choice and number of target lesions (**Figure 1**) and the experience of the reader(s). Where multiple target lesions are used, their selection affects variability: agreement ranges from 0.58 when different targets are chosen to 0.97 when the same targets are used (23). Variability also increases with the number of target lesions selected. For this reason, it has been recommended that a central review in clinical trials should include two readers and one adjudicator (29). Finally, reader experience has major impact on variability, from the selection of the correct reference examination (baseline vs. previous CT) to the detection and proper interpretation of new lesions (20, 21, 25, 26). Measurements of well-demarcated lesions and bigger lesions are also more reproducible (17–19), which vindicates RECIST recommendations for the choice of target lesions.

**TABLE 3 |** RECIST reproducibility and factors impacting it.

| Biomarker | Reproducibility | | | Factors impacting reproducibility |
|---|---|---|---|---|
| | 95% limits of agreement | Kappa | Other | |
| RECIST (measurement) CT (size) | Per lesion<br>- Intra-obs: -18% to 16%<br>- Inter-obs: -22% to 25% (1 (17)<br>Per sum of diameters<br>- Intra-obs: -10% to 13%<br>- Inter-obs: -20% to 20%<br>Interval change in tumour burden (% change between time points)<br>- -31% to 30%<br>Repeatability (same image on repeat CT taken within 15 minutes)<br>- -4% to +4% (18) | With target lesion selection<br>- Intra-obs: 0.957 (19)<br>- Inter-obs 0.954 (19)<br>Target response classification<br>- Inter-obs: 0.48 (20) to 0.66 (21)<br>Non-target response classification<br>- Inter-obs: 0.58 (20) | Lesion size ICC (22)<br>-Pre-treatment: 0.72<br>-Post-treatment: 0.85<br>-Interval change: 0.70 | -Selection of target lesions differs in 21 to 33% (17, 23, 24)<br>-Practical training (ref 40)/ expertise (21)<br>-Same observer (17, 20)<br>-Well delineated lesions (17, 19)<br>-Lesions size (greater variability for smaller lesions) (18, 19)<br>-Adjudication could reduce easily avoidable inconsistencies (20, 25) |
| RECIST (overall response) | | With target lesion selection<br>- Inter-obs: 0.97 (23)<br>Without target lesion selection<br>- Inter-obs: 0.51 (20), 0.53 (24, 26) to 0.58 (23) | -30% of patients classified differently in a cohort of 39 pts with 2 readers (26) | -Arbitrary nature of CR/PR/SD/PD categories (10)<br>-Inconsistencies mainly due to interpretation of new lesions (20, 26)<br>-Choice of target lesions |
| 3D measurement | - Intra-obs: 0.4 to 33% according to automated volume measurement method (27) | Whole body volumetry<br>- Inter-obs: 0.95 30) | -Discordant classification in overall response in 10 to 21% of patients according to automated volume measurement method (27) | -Time consuming (28)<br>-Do not resolve the discrepancies linked to the choice of target lesions (24) |

*95% limits of agreement are derived from the Bland-Altman method comparing two measurements of the same variable. Kappa coefficients measure agreement between qualitative observations. ICC measures the reliability of measurements by comparing the variability of different ratings of the same subject to the total variation across all ratings and all subjects. Intra-obs, intra-observer; Inter-obs, inter-observer; ICC, Intra-class coefficient; CR, complete response; PR, partial response; SD, stable disease; PD, progressive disease.*

## Statement #3

*RECIST reproducibility is impacted by reader experience, choice of target lesions, lesion characteristics, and detection/interpretation of new lesions. At an individual level, this can result in patients being categorised incorrectly when values of SOD are near thresholds or when new lesions are either missed or incorrectly interpreted.*

## HOW REPRODUCIBLE ARE OTHER BIOMARKERS?

**Table 4** summarises repeatability and reproducibility of some of the other biomarkers suggested or used as alternatives to RECIST for evaluating response. With the abundance of suggested candidate biomarkers in the published literature, the purpose here is not to be comprehensive, but to give a general overview of some of the most frequently explored options for providing a level of comparison with RECIST.

A first alternative to measuring a single size dimension as a response biomarker, would be to measure volume of a single or several lesions as an indicator of tumour bulk. This seems particularly important when lesions are irregular in shape, or

when they change orientation and are therefore not identically represented on standard axial follow-up scans. Volumetric response on first follow-up CT has been shown to better predict OS than RECIST response (65). Tumour volume response has been utilised in lung (66), cervical (67), and other solid malignancies (68). Despite a trend towards better intra- and inter-observer reproducibility, the routine use of volume has been hampered by the need for manual segmentation, which is user-dependent and time-consuming and does not improve the discrepancies linked to the choice of target lesions (24, 28). Aside from tumour bulk, metabolic activity of tumours through functional imaging (e.g. positron emission tomography - PET)) is highly predictive of response in lymphoma (69), lung cancer (70), and metastatic melanoma (71). Other radioligands are utilised for response or recurrence detection, e.g. ^{18}F-fluoroestradiol (FES) in hormone-dependent breast cancer (72) and [18]F- or [68]Ga Prostate-Specific Membrane Antigen (PSMA) ligands in prostate cancer (73). Additionally, radiolabelled ligands of various metabolites and biologically active molecules can assess proliferation, hypoxia, angiogenesis, apoptosis and gene transfection (74).While parameters used for the quantification and measurement of tumour metabolism by PET are generally based on semi-quantitative assessments, these can be made relatively reproducible and harmonised
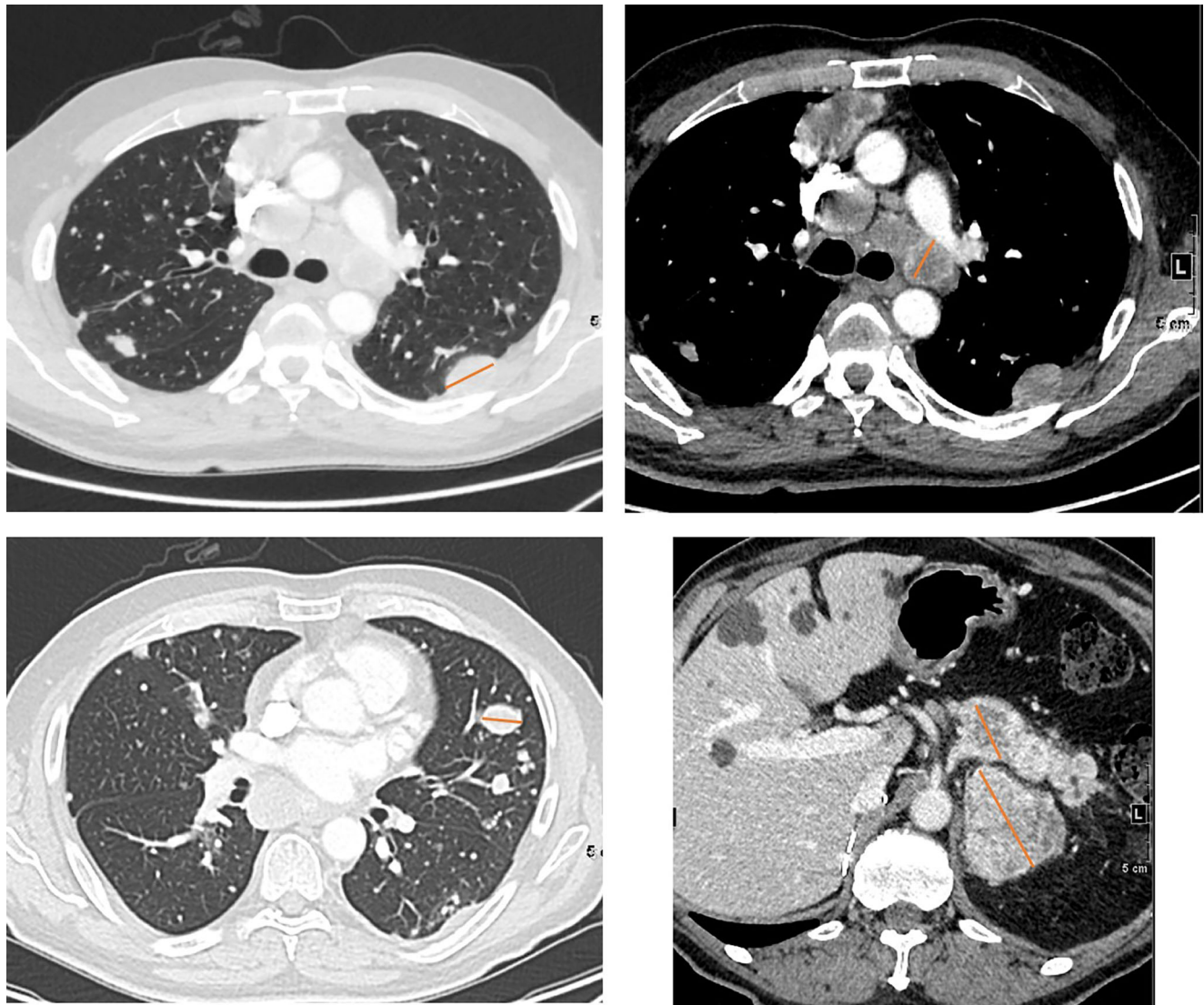
**FIGURE 1** | Selecting target lesions in a 58 yo patient with metastatic renal cell carcinoma. Multiple lung, lymph node, pancreatic and adrenal metastases are present. Lymph nodes should be sampled from different locations where possible. Selection of target lesions at baseline from multiple organ sites is important for response evaluation at a patient level.

throughout the world through standardised imaging protocols and dedicated initiatives promoted by the international scientific societies (75, 76), such as the accreditation program developed by the EANM Research Ltd. (EARL) (34, 77).

Other alternate imaging biomarkers include perfusion and diffusion imaging. As tumours are commonly characterised by neo-angiogenesis, perfusion and permeability derived from dynamic-contrast enhanced studies (e.g. with MR or CT) have been contenders for measuring early response (78), and vascularity can be quantified using most imaging techniques, such as MRI, CT, ultrasound and PET. The utility of biomarkers of vascularity has been demonstrated particularly where anti-angiogenic agents such as bevacizumab have been part of the therapeutic strategy (79). However, their quantitation, which

depends on measuring or estimating an arterial input function, is susceptible to large potential variations (80), and the reproducibility of such data is often low, thus limiting their clinical utility (81). Another biomarker reflecting tissue cellularity, the apparent diffusion coefficient (ADC) from DW-MRI, has proven a robust quantitative measure with good repeatability and reproducibility across vendor platforms (82), and has the potential to detect therapeutic response earlier than size measurements. It is increasingly being introduced routinely into scanning protocols, as it does not require injection of an extrinsic contrast agent and is simple and fast to acquire and analyse. Increasing automation with artificial intelligence (AI) systems may aid the translation of biomarkers indicative of tumour characteristics other than bulk into routine clinical

**TABLE 4** | Reproducibility and factors impacting it of other imaging biomarkers.

| Biomarker | Reproducibility | | | Factors impacting reproducibility |
|---|---|---|---|---|
| | **ICC** | **Coefficient of Variation** | **Other** | |
| Metabolic activity ($^{18}$-FDG PET) *Semiquantitative*: SUV (SUV$_{max}$, SUV$_{mean}$; SUV$_{peak}$), SUL (SUL$_{max}$, SUL$_{mean}$, SUL$_{peak}$); MTV, TLG *Response criteria*: PERCIST (30)/EORTC (31) | SUV$_{max}$ (4 observers) (22)<br>- Pre-treatment: 0.93<br>- Post-treatment: 0.91<br>- Interval change: 0.94<br>- SUV$_{mean}$ repeatability (32)<br>- 0.91 (meta-analysis)<br>SUV$_{peak}$<br>- -31% to 30% | SUVmax (4 observers) (22)<br>- Pre-treatment: 6.3%<br>- Post-treatment: 18.4%<br>- Interval change: 16.7% | Repeatability standard deviation (33)<br>- SUV$_{max}$: 1.01<br>- SUV$_{mean}$: 0.28 | *Technical factors*:<br>Scanner calibration/injected activity calibration (34, 35)<br>Incorrect decay correction (36)<br>Tracer extravasation (37, 38)<br>Residual activity in syringe (34)<br>Synchronization of clocks (34)<br>*Biological factors*:<br>Blood glucose levels (38)<br>Inflammation (34)<br>Patient preparation (38)<br>Injection-acquisition interval (39, 40)<br>BMI/metabolic syndrome (41)<br>Drug interaction/corticosteroids (38)<br>*Physical factors*:<br>Acquisition parameters/matrix size (34, 36)<br>Reconstruction algorithm (39, 42, 43)<br>Partial volume effect (44)<br>Normalization factor for SUV (34, 45)<br>Use of contrast agents (34)<br>ROI/VOI definition (39, 42)<br>Semiautomated/manual contouring (46)<br>Movement artifacts/respiratory movements (34)<br>Recovery effect/motion blur (47)<br>Image noise (44, 48)<br>Background activity/visual assessment (42, 49)<br>Lesion size/location (50) |
| Vascularity (DCE MRI) | DCE-MRI ktrans<br>- Intra-obs: 0.98 (51)<br>DCE CT (arterial flow, blood volume, permeability) - Intra-obs: 0.72-0.89<br>- Inter-obs: 0.70-0.91 (52)<br>DCE and DSC-MRI intersoftware reproducibility ICC 0.31 to 0.58 (53) | DCE MRI<br>- model-free parameters (ex: AUC$_{60}$, peak…): 12-24%<br>- modelled parameters (ex: distribution volume, blood flow, mean transit time): 21-29% (54)<br>DSC MRI normalised rCBVmax<br>- repeatability: 50%,<br>- reproducibility: 6% (55)<br>DCE-CT (blood flow, blood volume, mean transit time, permeability)<br>- within subject: 18% to 25%;<br>DCE-MRI (Ktrans, k(ep), v(e))<br>- within subject 16% to 23% (56), | | - Parameter extraction model (54)<br>- Segmentation: 3D vs 2D regions of interest (52)<br>- Software (53) |
| Cellularity (MR) ADC | ADC mean value<br>- Intra-obs: 0.91 (57) – 0.99 (51)<br>- Inter-obs: 0.92 (57)<br>ROI segmentation method (Inter-obs)<br>- Manual method: 0.69<br>- Semi-automated volumetric method: 0.96 (58) | Repeatability<br>- ADC total = 4.8% (57), 7.1% (59) to 13.3% (60)<br>Different post-processing platforms<br>- 2.8% (59)<br>Different sites<br>- multicentric: 9% (61)<br>- ice-water phantom: 1.6% (61)<br>- breast fibroglandular tissue: 7.0% (61) | Repeatability (single centre)<br>- ≤ ± 0.1x10$^{-3}$ mm$^2$/s (62) | - Field homogeneity gradient linearity (63)<br>- QA procedure by trained operators assessing artifacts, fat suppression, and signal-to-noise ratio (57)<br>- Segmentation: 2D vs. 3D, manual vs. semi-automatic (58)<br>- Choice of measurement: mean/min/max/percentiles of ADC (64)<br>- Lesion size (59) |

*SUV$_{max}$ is measured as the maximum single voxel value of SUV, SUV$_{mean}$ is the average value of SUV in all voxels above a threshold, SUV$_{peak}$ (is the average value of SUV in a region of interest positioned so as to maximize the enclosed average.*

*SUV, standardized uptake value; SUL, lean body mass corrected SUV; MTV, metabolic tumour volume; TLG, total lesion glycolysis; PERCIST, PET Response Criteria in Solid Tumours; EORTC, European Organization for Research and Treatment of Cancer; wCV, within-subject coefficient of variation; BMI, body mass index; ROI, region of interest; VOI, volume of interest; ICC inter correlation coefficient; DCE dynamic contrast enhanced; DSC-MRI dynamic susceptibility contrast magnetic resonance imaging; ADC, apparent diffusion coefficient; QA quality assurance; 3D, three-dimensional; 2D, bi-dimensional; AUC$_{60}$, area under the curve at 60s; rCBV, relative cerebral blood volume; Ktrans, transfer constant; k(ep), wash-out transfer constant; v(e), extracellular volume.*

workflows. Unfortunately, tightened legal rules are slowing down the process of their adoption (83).

Although historically dependent on imaging, response assessment for malignancies may now also include liquid biopsies [quantification of circulating tumour cells or DNA (CTC, ctDNA)], as well as histological sampling. ctDNA shedding is influenced by the overall tumour burden (cells) and may thus inform the use of imaging in relation to likely tumour size (84), because ctDNA estimations require less workflow and infrastructure than repeated monitoring with imaging. Initial clinical evaluations showed that ctDNA detected response earlier than imaging-based assessment (85). The simplest clinical implementation of ctDNA may be in postoperative monitoring of disease recurrence (86) but even here reproducibility and standardisation issues remain limiting. In one study, ctDNA quantities based on measurement of some target genes (e.g. TERT) were, on average, more than two-fold higher than those of other assays (e.g. ERV3) (87). In another, quantities of cell-free DNA for the different isolation methods for detection of EGFR variants in NSCLC varied between medians of 1.6 ng/mL and 28.1 ng/mL (88). Moreover, concordance between tissue and plasma variant detection for leading platforms has been shown to range from 70% to 90% (89). Thus, ctDNA extraction/isolation methods (87, 88) may need to be standardised before routine clinical use.

Finally, histopathology may also be a method for tumour response evaluation. However, serial histological sampling is not routinely used for response assessment and has thus far shown agreement with imaging-based responses only in a few studies (90). Histopathological evaluation of response is performed usually after neoadjuvant therapy, when the organ is surgically resected. Qualitative or semi-quantitative histopathological evaluation also presents variable reproducibility according to organs, methods and published studies (91–94). Agreement between pathologists yielded kappa values ranging from 0.21 for extent in prostate cancer (92), to 0.49 for multiple well-trained observers in cervical cancer (93), 0.64 for a 5-point tumour regression grade in rectal cancer (90) and 0.83 for a central review in bladder cancer (91). As with macroscopic imaging, reader experience (94), and central review (92) improve reproducibility.

## Statement #4:

*Alternative biomarkers for tumour response yield reproducibility generally comparable to RECIST. Each technique has its sources of variability, and it is important to understand inherent variability and limitations of individual biomarkers. It is critical that imaging experts communicate their level of confidence in any chosen biomarker.*

## WHAT ARE COMMON RECIST LIMITATIONS?

## Challenging Organs: Bone

Bone metastases were considered unmeasurable in the initial RECIST initiative, because of the lack of sensitivity of existing techniques to bone marrow infiltration (4). On CT it is the bone's osteolytic or osteosclerotic reaction to the presence of tumour, or its response to therapy (flare lesions) that is visualised rather than the tumour itself (95, 96). With the updated RECIST 1.1. version, bone metastases with soft tissue masses ≥10 mm are recognized as measurable target lesions (1). Nevertheless, bone lesions without soft tissue involvement, whether lytic, mixed or sclerotic, remain unmeasurable by RECIST. Since the early 1990s, bone marrow MRI has been shown to be superior to bone scintigraphy and CT for the assessment of bone metastatic disease. Bone marrow replacement by neoplastic foci is detected and quantified on T1-weighted and fat-suppressed T2-weighted MRI sequences (97, 98), more recently complemented with diffusion-weighted imaging (DWI) sequences (99, 100). However, to date, RECIST 1.1 has not validated quantitative bone MRI for tumour response assessment. Positron Emission Tomography Response Criteria in Solid Tumours (PERCIST), introduced in 2009 (30, 101), enables response to be measured in $^{18}$F-fluorodeoxyglucose ($^{18}$F-FDG) avid bone metastatic lesions based on their metabolic activity in the absence of any obvious anatomic changes. Finally, PSMA-PET appears promising for identifying bone marrow invasion due to prostate cancer, regardless of the impact on the bone mineral content (102, 103).

## Challenging Diseases: GIST and mCRC

As RECIST is not organ-specific, it might not capture the key parameters that are associated with survival outcomes in certain cancer types, and under certain types of treatment. In gastrointestinal malignancies, the hepatic tumour burden and its response commonly outperform other sites of metastatic disease for survival prediction. A study in metastatic colorectal cancer (mCRC) showed that the depth and uniformity of response in liver metastases represented a highly useful and clinically relevant indicator for therapy monitoring (104). Organ-specific response patterns may also occur under immunotherapy possibly due to varying immune microenvironments in organs or the lymphatic system (105–107). Thus, choice of target lesions would largely impact the response observed according to the organ, as well as the predictive ability of RECIST. In this case also, reader experience and knowledge of the disease is crucial for proper target lesion selection.

Response to therapy in patients with advanced GIST was drastically improved by the introduction of imatinib, a tyrosine-kinase inhibitor. Imatinib treatment has been shown to induce necrosis with a marked decrease in vascularity of GIST lesions, resulting in a decrease in CT density often before any significant decrease in size is seen, thus leading to underestimation of the initial tumour response (108, 109) (**Figure 2**). A paradoxical increase in volume is occasionally observed, simulating progression (110). Choi et al. therefore proposed adapted criteria for GIST, combining changes in tumour density on contrast-enhanced CT expressed in Hounsfield units (HU) and/or size to determine tumour response (109): PR is defined as a decrease of ≥10% in the SOD or a decrease of ≥ 15% in tumour density of target lesions, whereas PD is defined as a ≥ 10% increase in size and not meeting the PR criteria by tumour density. PD may also occur if new intra-tumoural nodules are present or existing intra-tumoural nodules show an increase in

size, factors which are not catered for in RECIST. In patients treated with imatinib, Choi criteria showed a significantly better correlation with survival rates than RECIST (111).

## Challenging Treatments: Focal Therapies

Treatment of tumour lesions with ablative therapies, such as radiofrequency ablation, microwave ablation or cryoablation, results in a larger defect than the original lesion and such treated lesions are not considered measurable unless there is progression at this site (1), such as the development of a new measurable nodule within the ablation defect. Distinguishing normal post-ablation changes from residual disease and recurrence can be challenging (112).

Intravascular therapies are also a challenge for the use of RECIST. Trans-arterial radioembolization (TARE) induces inflammatory changes with a generally delayed morphologic response (112). A reduction of $^{18}$F-FDG uptake on early PET-CT has been found to be helpful in predicting further outcome of these patients (113). As a consequence, both TARE and intra-arterial therapies such as trans-arterial chemoembolization (TACE) in hepatocarcinoma require modified RECIST (mRECIST) criteria derived from arterial and portal venous enhancement phases of CT or MRI (114), and which take into account both lesion size and vascularity.

High-intensity focused ultrasound (HIFU), under the guidance of ultrasound or MRI, has also been used as a non-invasive technique for tissue ablation in prostate cancer and more recently in recurrent gynecological malignancy (115). The use of HIFU for hepatic tumour lesions is still in the exploratory stage. As for other ablative therapies and for similar reasons (116), RECIST 1.1 appears to be unsuitable for local response evaluation following HIFU applied to liver lesions.

Finally, tumour lesions in a previously irradiated area (via CyberKnife, stereotactic radiotherapy or traditional fractionated radiation therapy) are not considered measurable (1) and must be excluded from RECIST evaluation due to the inflammatory or fibrotic changes that may be observed, thus making evaluation of size unreliable.

### Statement #5

*There are several scenarios in which RECIST criteria fail to evaluate treatment-induced changes correctly. Informed appreciation that RECIST criteria are not applicable to all tumour sites and situations is thus crucial for proper interpretation and again dependent on reader experience.*

## WHEN IS RECIST RESPONSE ASSESSMENT MISLEADING?

### Pseudo-Progression

During immunotherapy, RECIST may describe progression that can be misleading and is thus classified as "pseudo-progression". In fact, in around 5 to 10% of patients with metastatic disease treated with check-point inhibitors, an initial increase of tumour burden has been observed, followed by actual response or long-term stabilisation of disease (117–119). This phenomenon relates to the mechanism of action of immunotherapy, which stimulates the immune response and initially induces inflammation and tumour swelling, thus delaying visible tumour shrinkage. For this reason, adaptations of RECIST criteria for assessing treatment response to immunotherapy (iRECIST) have been developed. The first ascertainment of progression by iRECIST is considered



**FIGURE 2** | Response unrelated to tumour size in a 66 yo patient treated with imatinib for a gastrointestinal stromal tumour (GIST). Compared to the baseline image (left), after treatment (right) the tumour shows a dramatic decrease in density rather than in size.

as "immune unconfirmed progressive disease"(or iUPD), and requires, if possible, a subsequent evaluation 4 to 8 weeks later in order to confirm true progression (120) (**Figure 3**).

## Mixed Response/Progression

In some patients, the tumour bulk does not respond homogeneously, with some lesions increasing and others decreasing. Mixed or heterogeneous response is defined as an increase in size of some tumour lesions and decrease of others in the same patient during treatment. This lesion-specific response has been attributed to the emergence of drug-resistant clones and indicates that tumour heterogeneity is likely causing treatment failure (121, 122). Mixed response has the same incidence in patients treated with targeted cancer agents and those undergoing chemotherapy alone or even combined with targeted agents (12, 28).

Since RECIST records overall patient response rather than individual lesion response, the choice of target lesions critically affects the objective assessment of overall patient response in patients with mixed response in individual lesions (**Figure 4**) (12). As lesions escaping treatment control will weigh negatively on patient prognosis (123), their presence should be annotated in order to offer the best alternative treatment for the patient.

Lesion cavitation, necrosis and residual non-viable masses represent other forms of response than decrease in size and may complicate RECIST assessment (124). Tumour necrosis with cavitation is present in approximately 14-24% of NSCLC patients undergoing anti-angiogenic drug therapy (125–127). When cavitation is present, lesion size may not change significantly and RECIST may therefore under-estimate the effect of therapy. Conversely, cavitation also risks missing progression if there is tumour regrowth inside the cavity. While alternative criteria have been proposed in such cases, e.g. subtracting the longest cavitation diameter from the largest lesion diameter (such as Crabb criteria) (126), these are not commonly used.

When residual tissue is present after therapy, evaluation with RECIST criteria is subject to pitfalls. First, an asymmetric shrinkage of the tumour may result in a similar longest diameter and consequent stable disease (SD) rating not reflecting the real response to treatment (**Figure 2**). Second, it may be difficult to distinguish between viable tumour and fibrosis. In such cases, best response assessment, an important endpoint in phase 2 studies (partial vs. complete response; PR vs. CR) may be affected (126). According to RECIST guidelines, in equivocal cases, residual lesions should be evaluated by either biopsy or PET(-CT) (**Figure 2**). This may well then allow upgrading PR to CR. However, false positive PET findings are not uncommon (128). Alternatively, other advanced imaging tests, such as DWI-MRI or perfusion imaging (e.g. from MR or CT) could be used.

## Statement #6

*Some patterns of response/progression cannot be correctly documented by RECIST. These require specialist reader experience and communication with oncologists to determine appropriate evaluation approaches and/or therapeutic options.*
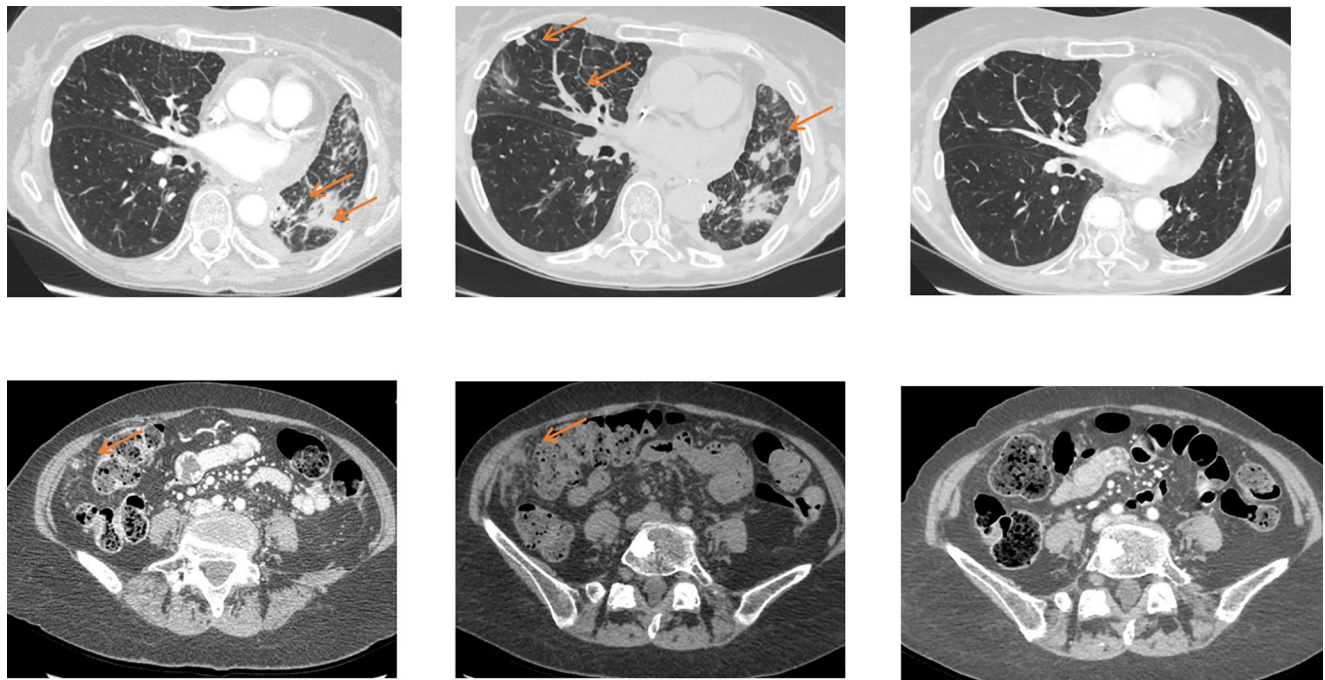


**FIGURE 3** | Pseudoprogression on immunotherapy in a 56 yo patient with metastatic non-small cell lung cancer. The baseline image (left) shows lung and peritoneal nodules (arrows). After 4 wks of antiPDL1 therapy (middle), CT shows an increase in previous lesions and the appearance of new lung nodules. Disease was considered immune unconfirmed progressive disease. Six weeks later (right) a dramatic response in all previous lesions was seen classifying the patients as a complete responder and endorsing an earlier diagnosis of pseudoprogression.
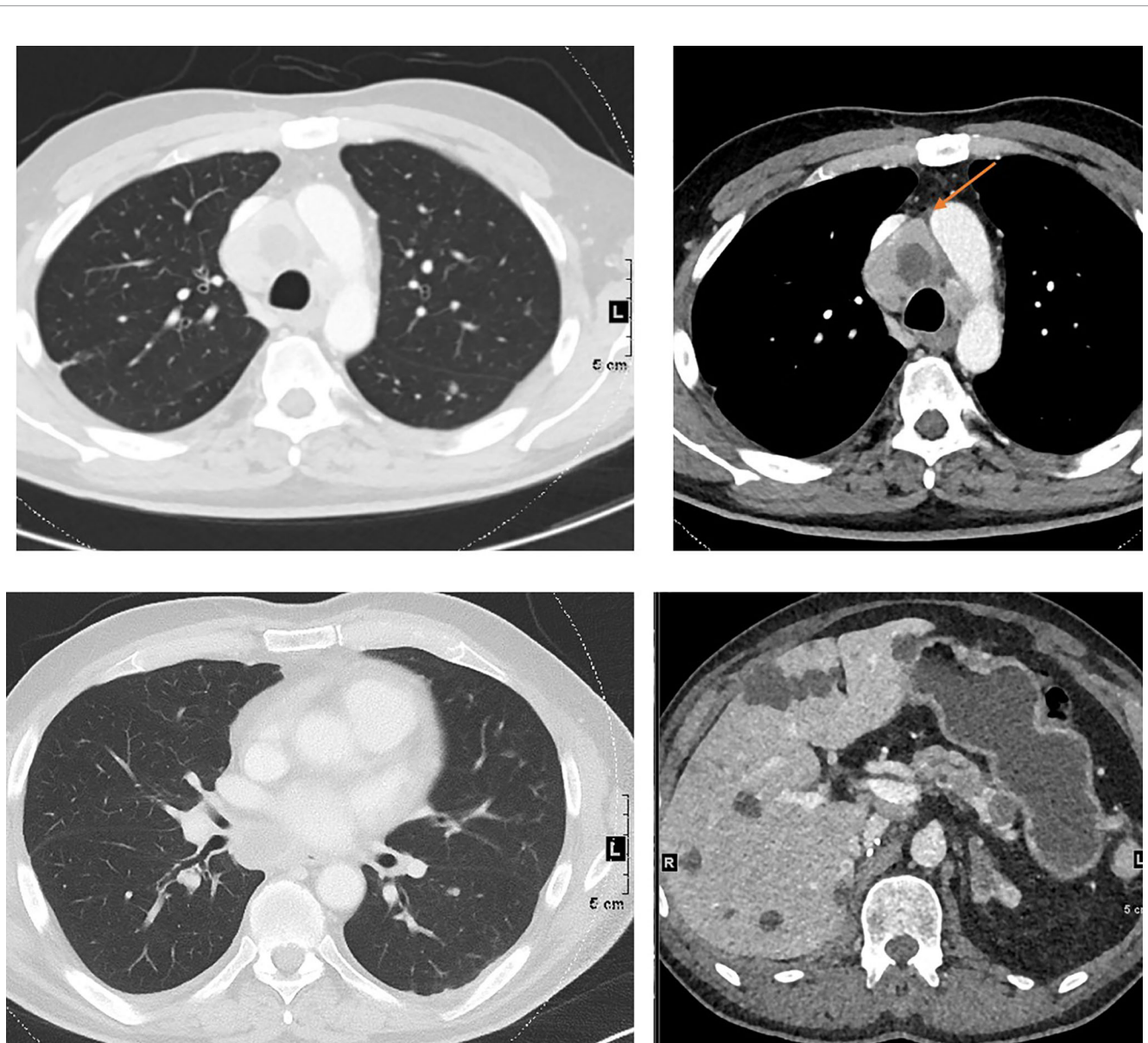
**FIGURE 4** | Mixed response to treatment in the same patient illustrated in **Figure 1**. Eight weeks after targeted therapy lung, adrenal and pancreatic metastases decreased, whereas one mediastinal lymph node (top right, arrow) increased.

## SHOULD (COULD) RECIST BE AUTOMATED?

The core assumption of RECIST is that a single diameter on the cross-sectional imaging slice presenting the largest cross-section of a given lesion (or sum thereof) is a surrogate for tumour burden. This assumes that lesions are grossly spherical and that their size represents their overall activity. To streamline the determination of this single diameter and make it less subject to possible human-induced variability, semi- or fully-automated 2D or even 3D segmentation techniques can be applied to target lesions, which can also be semi- or fully-automatically tracked between scans

acquired at different time points (129–134). The 2D or 3D mask resulting from the segmentation process then readily permits the automated and accurate extraction of the largest diameter from the segmented lesion. With 3D segmentation, the full volume of a target lesion can be provided alongside an automatically extracted largest diameter, which may not be oriented in the 2D plane of the source images in a broader RECIST interpretation, together with any other geometric metric of relevance. Using the largest 3D diameter would allow RECIST to be used beyond 2D constraints, and can account for non-orthogonal motion of target lesions between scans at different time-points. While segmentation and tracking can now plausibly be fully automated, especially with

newer approaches using machine learning, and such capabilities are already implemented in several commercially available clinical systems, some challenges remain with key RECIST operations, such as the proper selection of target lesions and dealing with new or disappeared lesions. These are currently still best addressed or verified with a human (e.g. a radiologist) in the loop (20, 135).

## Statement #7

*Though automation is desirable to streamline the process and potentially reduce the variability of results within the RECIST paradigm, remaining technical challenges must be overcome to ensure proper repeatability, and human adjudication is still required.*

## RECIST IN NOVEL DRUG DEVELOPMENT

RECIST measurements play a pivotal role in the development of novel oncological drugs (136). In most registered randomised controlled trials (RCTs), studies are powered to meet primary endpoints such as OS/PFS, which determines the number of patients recruited. A study of RCTs between 2006 and 2015 looking for evidence of clinical efficacy of novel oncology drugs in order to gain US Food and Drug Administration (FDA) approvals had PFS as primary endpoint in 28 out of 42 RCTs (66%), and OS in 14 (33%). In 2012, 12 novel anticancer drugs were approved by the FDA; only three drugs showed improvement of overall survival (137). Similarly, a study of drugs approved by the European Medicine Agency (EMA) between 2009 and 2013 also showed that only 18 of 68 (26%) novel drug uses were supported by OS data, whereas PFS was used in 31 (46%) (138). In the vast majority of trials, PFS is determined using the RECIST1.1 framework, or iRECIST for immune-oncology trials. It is acknowledged however, that in some disease types other criteria are used: e.g. Lugano criteria for $^{18}$F-FDG PET/CT or RECIL in lymphoma (139, 140) and RANO criteria for brain tumours (141, 142). The fact that PFS can predict OS outcome in large patients cohorts with commonly occurring cancers, reinforces the use of RECIST criteria in clinical trials (143). Moreover, rapid progress in drug development will make the reliance on OS as endpoint for novel drugs in oncology increasingly challenging because treatment options on progression on trial, including in-trial cross-over, are increasing.

## Statement #8

*Although the RECIST framework might not be perfect, the scientific basis for the anticancer drugs that have been approved using a RECIST-based surrogate endpoint remains valid.*

## REFERENCES

1. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New Response Evaluation Criteria in Solid Tumours: Revised RECIST Guideline (Version 1.1). *Eur J Cancer* (2009) 45(2):228–47. doi: 10.1016/j.ejca.2008.10.026
2. Moertel CG, Hanley JA. The Effect of Measuring Error on the Results of Therapeutic Trials in Advanced Cancer. *Cancer* (1976) 38(1):388–94. doi: 10.1002/1097-0142(197607)38:1<388::AID-CNCR2820380156>3.0.CO;2-A

## RECIST: ONLY AS GOOD AS ITS USERS?

RECIST criteria were developed for clinical trials and thresholds chosen to produce a comparable classification of patients, taking into account tumour measurement variability. These criteria are widely used in clinical trials and accepted by regulatory agencies. Despite some limitations, the scientific basis for the anticancer drugs that have been approved using a RECIST-based surrogate endpoint remains valid. Reader experience, choice of target lesions and detection of new lesions impact RECIST reproducibility, which necessitates adequate training of radiologists using these criteria. Automation is not currently sufficiently reliable to replace human experience. Unfortunately, some organ-, disease- or drug-specific patterns of response/progression cannot be correctly documented by RECIST.

This expert statement includes that RECIST remains a tool for radiologists that needs to be used with discrimination and good understanding of its purpose and limitations. Training of radiologists is essential to improve its application and reproducibility. RECIST conclusions should not go against common (or informed) sense. Furthermore, RECIST criteria have the advantage of simplicity, availability, cost-effectiveness, and intuitiveness. Overall, therefore, RECIST provides a common language between oncologists and imaging experts (e.g. radiologists), provided there is full understanding of how measurements are made, what they represent, and their inherent limitations.

## 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

All authors contributed to conception and design. CC and LF wrote the first draft of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

## ACKNOWLEDGMENTS

3. Miller AB, Hoogstraten B, Staquet M, Winkler A. Reporting Results of Cancer Treatment. *Cancer* (1981) 47(1):207–14. doi: 10.1002/1097-0142 (19810101)47:1<207::AID-CNCR2820470134>3.0.CO;2-6
4. Therasse P, Arbuck SG, Eisenhauer EA, Wanders J, Kaplan RS, Rubinstein L, et al. New Guidelines to Evaluate the Response to Treatment in Solid Tumors. *JNCI: J Natl Cancer Inst* (2000) 92(3):205–16. doi: 10.1093/jnci/92.3.205
5. James K, Eisenhauer E, Christian M, Terenziani M, Vena D, Muldal A, et al. Measuring Response in Solid Tumors: Unidimensional Versus

Bidimensional Measurement. *JNCI: J Natl Cancer Inst* (1999) 91(6):523–8. doi: 10.1093/jnci/91.6.523

6. Bogaerts J, Ford R, Sargent D, Schwartz LH, Rubinstein L, Lacombe D, et al. Individual Patient Data Analysis to Assess Modifications to the RECIST Criteria. *Eur J Cancer* (2009) 45(2):248–60. doi: 10.1016/j.ejca.2008.10.027

7. Moskowitz CS, Jia X, Schwartz LH, Gönen M. A Simulation Study to Evaluate the Impact of the Number of Lesions Measured on Response Assessment. *Eur J Cancer* (2009) 45(2):300–10. doi: 10.1016/j.ejca.2008.11.010

8. Booth CM, Eisenhauer EA. Progression-Free Survival: Meaningful or Simply Measurable? *JCO* (2012) 30(10):1030–3. doi: 10.1200/JCO.2011.38.7571

9. Suzuki C, Blomqvist L, Sundin A, Jacobsson H, Byström P, Berglund Å, et al. The Initial Change in Tumor Size Predicts Response and Survival in Patients With Metastatic Colorectal Cancer Treated With Combination Chemotherapy. *Ann Oncol* (2012) 23(4):948–54. doi: 10.1093/annonc/mdr350

10. Jain RK, Lee JJ, Ng C, Hong D, Gong J, Naing A, et al. Change in Tumor Size by RECIST Correlates Linearly With Overall Survival in Phase I Oncology Studies. *J Clin Oncol* (2012) 30(21):2684–90. doi: 10.1200/JCO.2011.36.4752

11. Grünwald V, McKay RR, Krajewski KM, Kalanovic D, Lin X, Perkins JJ, et al. Depth of Remission is a Prognostic Factor for Survival in Patients With Metastatic Renal Cell Carcinoma. *Eur Urol* (2015) 67(5):952–8. doi: 10.1016/j.eururo.2014.12.036

12. Litière S, Isaac G, De Vries EGE, Bogaerts J, Chen A, Dancey J, et al. RECIST 1.1 for Response Evaluation Apply Not Only to Chemotherapy-Treated Patients But Also to Targeted Cancer Agents: A Pooled Database Analysis. *J Clin Oncol* (2019) 37(13):1102–10. doi: 10.1200/JCO.18.01100

13. Litière S, de Vries EGE, Seymour L, Sargent D, Shankar L, Bogaerts J. The Components of Progression as Explanatory Variables for Overall Survival in the Response Evaluation Criteria in Solid Tumours 1.1 Database. *Eur J Cancer* (2014) 50(10):1847–53. doi: 10.1016/j.ejca.2014.03.014

14. An M-W, Mandrekar SJ, Branda ME, Hillman SL, Adjei AA, Pitot HC, et al. Comparison of Continuous Versus Categorical Tumor Measurement-Based Metrics to Predict Overall Survival in Cancer Treatment Trials. *Clin Cancer Res* (2011) 17(20):6592–9. doi: 10.1158/1078-0432.CCR-11-0822

15. An M-W, Dong X, Meyers J, Han Y, Grothey A, Bogaerts J, et al. Evaluating Continuous Tumor Measurement-Based Metrics as Phase II Endpoints for Predicting Overall Survival. *J Natl Cancer Inst* (2015) 107(11):djv239. doi: 10.1093/jnci/djv239

16. Sullivan DC, Obuchowski NA, Kessler LG, Raunig DL, Gatsonis C, Huang EP, et al. Metrology Standards for Quantitative Imaging Biomarkers. *Radiol* (2015) 277(3):813–25. doi: 10.1148/radiol.2015142202

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. Oxnard GR, Zhao B, Sima CS, Ginsberg MS, James LP, Lefkowitz RA, et al. Variability of Lung Tumor Measurements on Repeat Computed Tomography Scans Taken Within 15 Minutes. *J Clin Oncol* (2011) 29(23):3114–9. doi: 10.1200/JCO.2010.33.7071

19. McErlean A, Panicek DM, Zabor EC, Moskowitz CS, Bitar R, Motzer RJ, et al. Intra- and Interobserver Variability in CT Measurements in Oncology. *Radiol* (2013) 269(2):451–9. doi: 10.1148/radiol.13122665

20. 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(1):50. doi: 10.1186/s40644-018-0186-0

21. Bellomi M, De Piano F, Ancona E, Lodigiani AF, Curigliano G, Raimondi S, et al. Evaluation of Inter-Observer Variability According to RECIST 1.1 and Its Influence on Response Classification in CT Measurement of Liver Metastases. *Eur J Radiol* (2017) 95:96–101. doi: 10.1016/j.ejrad.2017.08.001

22. Jacene HA, Leboulleux S, Baba S, Chatzifotiadis D, Goudarzi B, Teytelbaum O, et al. Assessment of Interobserver Reproducibility in Quantitative 18F-FDG PET and CT Measurements of Tumor Response to Therapy. *J Nucl Med* (2009) 50(11):1760–9. doi: 10.2967/jnumed.109.063321

23. 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. *Radiol* (2019) 290(2):349–56. doi: 10.1148/radiol.2018180648

24. Keil S, Barabasch A, Dirrichs T, Bruners P, Hansen NL, Bieling HB, et al. Target Lesion Selection: An Important Factor Causing Variability of Response Classification in the Response Evaluation Criteria for Solid Tumors 1.1. *Invest Radiol* (2014) 49(8):509–17. doi: 10.1097/RLI.0000000000000048

25. Karmakar A, Kumtakar A, Sehgal H, Kumar S, Kalyanpur A. Interobserver Variation in Response Evaluation Criteria in Solid Tumors 1.1. *Acad Radiol* (2019) 26(4):489–501. doi: 10.1016/j.acra.2018.05.017

26. Suzuki C, Torkzad MR, Jacobsson H, Aström G, Sundin A, Hatschek T, et al. Interobserver and Intraobserver Variability in the Response Evaluation of Cancer Therapy According to RECIST and WHO-Criteria. *Acta Oncol* (2010) 49(4):509–14. doi: 10.3109/02841861003705794

27. Rothe JH, Grieser C, Lehmkuhl L, Schnapauff D, Fernandez CP, Maurer MH, et al. Size Determination and Response Assessment of Liver Metastases With Computed Tomography–Comparison of RECIST and Volumetric Algorithms. *Eur J Radiol* (2013) 82(11):1831–9. doi: 10.1016/j.ejrad.2012.05.018

28. Zimmermann M, Kuhl CK, Engelke H, Bettermann G, Keil S. Factors That Drive Heterogeneity of Response-To-Treatment of Different Metastatic Deposits Within the Same Patients as Measured by RECIST 1.1 Analyses. *Acad Radiol* (2020) 28(8):e235–9. doi: 10.1016/j.acra.2020.05.029

29. Ford R, Schwartz L, Dancey J, Dodd LE, Eisenhauer EA, Gwyther S, et al. Lessons Learned From Independent Central Review. *Eur J Cancer* (2009) 45(2):268–74. doi: 10.1016/j.ejca.2008.10.031

30. Wahl RL, Jacene H, Kasamon Y, Lodge MA. From RECIST to PERCIST: Evolving Considerations for PET Response Criteria in Solid Tumors. *J Nucl Med* (2009) 50 Suppl 1:122S–50S. doi: 10.2967/jnumed.108.057307

31. Young H, Baum R, Cremerius U, Herholz K, Hoekstra O, Lammertsma AA, et al. Measurement of Clinical and Subclinical Tumour Response Using [18F]-Fluorodeoxyglucose and Positron Emission Tomography: Review and 1999 EORTC Recommendations. European Organization for Research and Treatment of Cancer (EORTC) PET Study Group. *Eur J Cancer* (1999) 35(13):1773–82. doi: 10.1016/S0959-8049(99)00229-4

32. de Langen AJ, Vincent A, Velasquez LM, van Tinteren H, Boellaard R, Shankar LK, et al. Repeatability of 18F-FDG Uptake Measurements in Tumors: A Metaanalysis. *J Nucl Med* (2012) 53(5):701–8. doi: 10.2967/jnumed.111.095299

33. Burger IA, Huser DM, Burger C, von Schulthess GK, Buck A. Repeatability of FDG Quantification in Tumor Imaging: Averaged SUVs Are Superior to SUVmax. *Nucl Med Biol* (2012) 39(5):666–70. doi: 10.1016/j.nucmedbio.2011.11.002

34. Aide N, Lasnon C, Veit-Haibach P, Sera T, Sattler B, Boellaard R. EANM/EARL Harmonization Strategies in PET Quantification: From Daily Practice to Multicentre Oncological Studies. *Eur J Nucl Med Mol Imaging* (2017) 44(Suppl 1):17–31. doi: 10.1007/s00259-017-3740-2

35. Kurland BF, Peterson LM, Shields AT, Lee JH, Byrd DW, Novakova-Jiresova A, et al. Test-Retest Reproducibility of 18F-FDG PET/CT Uptake in Cancer Patients Within a Qualified and Calibrated Local Network. *J Nucl Med* (2019) 60(5):608–14. doi: 10.2967/jnumed.118.209544

36. Brendle C, Kupferschläger J, Nikolaou K, la Fougère C, Gatidis S, Pfannenberg C. Is the Standard Uptake Value (SUV) Appropriate for Quantification in Clinical PET Imaging? - Variability Induced by Different SUV Measurements and Varying Reconstruction Methods. *Eur J Radiol* (2015) 84(1):158–62. doi: 10.1016/j.ejrad.2014.10.018

37. Miyashita K, Takahashi N, Oka T, Asakawa S, Lee J, Shizukuishi K, et al. SUV Correction for Injection Errors in FDG-PET Examination. *Ann Nucl Med* (2007) D21(10):607–13. doi: 10.1007/s12149-007-0068-1

38. Boellaard R, Oyen WJG, Hoekstra CJ, Hoekstra OS, Visser EP, Willemsen AT, et al. The Netherlands Protocol for Standardisation and Quantification of FDG Whole Body PET Studies in Multi-Centre Trials. *Eur J Nucl Med Mol Imaging* (2008) 35(12):2320–33. doi: 10.1007/s00259-008-0874-2

39. Zhuang M, García DV, Kramer GM, Frings V, Smit EF, Dierckx R, et al. Variability and Repeatability of Quantitative Uptake Metrics in 18F-FDG PET/CT of Non-Small Cell Lung Cancer: Impact of Segmentation Method, Uptake Interval, and Reconstruction Protocol. *J Nucl Med* (2019) 60(5):600–7. doi: 10.2967/jnumed.118.216028

40. Laffon E, de Clermont H, Marthan R. A Method of Adjusting SUV for Injection-Acquisition Time Differences in (18)F-FDG PET Imaging. *Eur Radiol* (2011) 21(11):2417–24. doi: 10.1007/s00330-011-2204-5

41. Kamimura K, Nagamachi S, Wakamatsu H, Higashi R, Ogita M, Ueno S, et al. Associations Between Liver (18)F Fluoro-2-Deoxy-D-Glucose Accumulation and Various Clinical Parameters in a Japanese Population: Influence of the Metabolic Syndrome. *Ann Nucl Med* (2010) 24(3):157–61. doi: 10.1007/s12149-009-0338-1

42. Krak NC, Boellaard R, Hoekstra OS, Twisk JWR, Hoekstra CJ, Lammertsma AA. Effects of ROI Definition and Reconstruction Method on Quantitative Outcome and Applicability in a Response Monitoring Trial. *Eur J Nucl Med Mol Imaging* (2005) 32(3):294–301. doi: 10.1007/s00259-004-1566-1

43. Lasnon C, Quak E, Le Roux P-Y, Robin P, Hofman MS, Bourhis D, et al. EORTC PET Response Criteria Are More Influenced by Reconstruction Inconsistencies Than PERCIST But Both Benefit From the EARL Harmonization Program. *EJNMMI Phys* (2017) 4(1):17. doi: 10.1186/s40658-017-0185-4

44. Akamatsu G, Ikari Y, Nishida H, Nishio T, Ohnishi A, Maebatake A, et al. Influence of Statistical Fluctuation on Reproducibility and Accuracy of SUVmax and SUVpeak: A Phantom Study. *J Nucl Med Technol* (2015) 43 (3):222–6. doi: 10.2967/jnmt.115.161745

45. Paquet N, Albert A, Foidart J, Hustinx R. Within-Patient Variability of (18) F-FDG: Standardized Uptake Values in Normal Tissues. *J Nucl Med* (2004) 45(5):784–8.

46. JH O, Lim SJ, Wang H, Leal JP, Shu H-KG, Wahl RL, et al. Quantitation of Cancer Treatment Response by 2-[18F]FDG PET/CT: Multi-Center Assessment of Measurement Variability Using AUTO-PERCIST™. *EJNMMI Res* (2021) 11(1):15. doi: 10.1186/s13550-021-00754-1

47. Apostolova I, Wiemker R, Paulus T, Kabus S, Dreilich T, van den Hoff J, et al. Combined Correction of Recovery Effect and Motion Blur for SUV Quantification of Solitary Pulmonary Nodules in FDG PET/Ct. *Eur Radiol* (2010) 20(8):1868–77. doi: 10.1007/s00330-010-1747-1

48. Lodge MA, Chaudhry MA, Wahl RL. Noise Considerations for PET Quantification Using Maximum and Peak Standardized Uptake Value. *J Nucl Med* (2012) 53(7):1041–7. doi: 10.2967/jnumed.111.101733

49. Adams MC, Turkington TG, Wilson JM, Wong TZ. A Systematic Review of the Factors Affecting Accuracy of SUV Measurements. *AJR Am J Roentgenol* (2010) 195(2):310–20. doi: 10.2214/AJR.10.4923

50. Quak E, Le Roux P-Y, Hofman MS, Robin P, Bourhis D, Callahan J, et al. Harmonizing FDG PET Quantification While Maintaining Optimal Lesion Detection: Prospective Multicentre Validation in 517 Oncology Patients. *Eur J Nucl Med Mol Imaging* (2015) 42(13):2072–82. doi: 10.1007/s00259-015-3128-0

51. Gürses B, Altınmakas E, Böge M, Aygün MS, Bayram O, Balık E. Multiparametric MRI of Rectal Cancer-Repeatability of Quantitative Data: A Feasibility Study. *Diagn Interv Radiol* (2020) 26(2):87–94. doi: 10.5152/dir.2019.19127

52. Lundsgaard Hansen M, Fallentin E, Axelsen T, Lauridsen C, Norling R, Svendsen LB, et al. Interobserver and Intraobserver Reproducibility With Volume Dynamic Contrast Enhanced Computed Tomography (DCE-CT) in Gastroesophageal Junction Cancer. *Diagn (Basel)* (2016) 6(1):E8. doi: 10.3390/diagnostics6010008

53. Conte GM, Castellano A, Altabella L, Iadanza A, Cadioli M, Falini A, et al. Reproducibility of Dynamic Contrast-Enhanced MRI and Dynamic Susceptibility Contrast MRI in the Study of Brain Gliomas: A Comparison of Data Obtained Using Different Commercial Software. *Radiol Med* (2017) 122(4):294–302. doi: 10.1007/s11547-016-0720-8

54. Aronhime S, Calcagno C, Jajamovich GH, Dyvorne HA, Robson P, Dieterich D, et al. DCE-MRI of the Liver: Effect of Linear and Nonlinear Conversions on Hepatic Perfusion Quantification and Reproducibility. *J Magn Reson Imaging* (2014) 40(1):90–8. doi: 10.1002/jmri.24341

55. Smits M, Bendszus M, Collette S, Postma LA, Dhermain F, Hagenbeek RE, et al. Repeatability and Reproducibility of Relative Cerebral Blood Volume Measurement of Recurrent Glioma in a Multicentre Trial Setting. *Eur J Cancer* (2019) 114:89–96. doi: 10.1016/j.ejca.2019.03.007

56. Ng CS, Waterton JC, Kundra V, Brammer D, Ravoori M, Han L, et al. Reproducibility and Comparison of DCE-MRI and DCE-CT Perfusion Parameters in a Rat Tumor Model. *Technol Cancer Res Treat* (2012) 11 (3):279–88. doi: 10.7785/tcrt.2012.500296

57. Newitt DC, Zhang Z, Gibbs JE, Partridge SC, Chenevert TL, Rosen MA, et al. Test-Retest Repeatability and Reproducibility of ADC Measures by Breast

58. DWI: Results From the ACRIN 6698 Trial. *J Magn Reson Imaging* (2019) 49 (6):1617–28. doi: 10.1002/jmri.26539

58. Kwee RM, Dik AK, Sosef MN, Berendsen RCM, Sassen S, Lammering G, et al. Interobserver Reproducibility of Diffusion-Weighted MRI in Monitoring Tumor Response to Neoadjuvant Therapy in Esophageal Cancer. *PLoS One* (2014) 9(4):e92211. doi: 10.1371/journal.pone.0092211

59. Weller A, Papoutsaki MV, Waterton JC, Chiti A, Stroobants S, Kuijer J, et al. Diffusion-Weighted (DW) MRI in Lung Cancers: ADC Test-Retest Repeatability. *Eur Radiol* (2017) 27(11):4552–62. doi: 10.1007/s00330-017-4828-6

60. Koh D-M, Blackledge M, Collins DJ, Padhani AR, Wallace T, Wilton B, et al. Reproducibility and Changes in the Apparent Diffusion Coefficients of Solid Tumours Treated With Combretastatin A4 Phosphate and Bevacizumab in a Two-Centre Phase I Clinical Trial. *Eur Radiol* (2009) 19(11):2728–38. doi: 10.1007/s00330-009-1469-4

61. Sorace AG, Elkassem AA, Galgano SJ, Lapi SE, Larimer BM, Partridge SC, et al. Imaging for Response Assessment in Cancer Clinical Trials. *Semin Nucl Med* (2020) 50(6):488–504. doi: 10.1053/j.semnuclmed.2020.05.001

62. 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

63. Newitt DC, Tan ET, Wilmes LJ, Chenevert TL, Kornak J, Marinelli L, et al. Gradient Nonlinearity Correction to Improve Apparent Diffusion Coefficient Accuracy and Standardization in the American College of Radiology Imaging Network 6698 Breast Cancer Trial. *J Magn Reson Imaging* (2015) 42(4):908–19. doi: 10.1002/jmri.24883

64. Newitt DC, Amouzandeh G, Partridge SC, Marques HS, Herman BA, Ross BD, et al. Repeatability and Reproducibility of ADC Histogram Metrics From the ACRIN 6698 Breast Cancer Therapy Response Trial. *Tomogr* (2020) 6(2):177–85. doi: 10.18383/j.tom.2020.00008

65. Hayes SA, Pietanza MC, O'Driscoll D, Zheng J, Moskowitz CS, Kris MG, et al. Comparison of CT Volumetric Measurement With RECIST Response in Patients With Lung Cancer. *Eur J Radiol* (2016) 85(3):524–33. doi: 10.1016/j.ejrad.2015.12.019

66. Lee JH, Lee HY, Ahn M-J, Park K, Ahn JS, Sun J-M, et al. Volume-Based Growth Tumor Kinetics as a Prognostic Biomarker for Patients With EGFR Mutant Lung Adenocarcinoma Undergoing EGFR Tyrosine Kinase Inhibitor Therapy: A Case Control Study. *Cancer Imaging* (2016) 1616:5. doi: 10.1186/s40644-016-0063-7

67. deSouza NM, Soutter WP, Rustin G, Mahon MM, Jones B, Dina R, et al. Use of Neoadjuvant Chemotherapy Prior to Radical Hysterectomy in Cervical Cancer: Monitoring Tumour Shrinkage and Molecular Profile on Magnetic Resonance and Assessment of 3-Year Outcome. *Br J Cancer* (2004) 90 (12):2326–31. doi: 10.1038/sj.bjc.6601870

68. Fenerty KE, Folio LR, Patronas NJ, Marté JL, Gulley JL, Heery CR. Predicting Clinical Outcomes in Chordoma Patients Receiving Immunotherapy: A Comparison Between Volumetric Segmentation and RECIST. *BMC Cancer* (2016) 16(1):672. doi: 10.1186/s12885-016-2699-x

69. Schmitz C, Hüttmann A, Müller SP, Hanoun M, Boellaard R, Brinkmann M, et al. Dynamic Risk Assessment Based on Positron Emission Tomography Scanning in Diffuse Large B-Cell Lymphoma: *Post-Hoc* Analysis From the PETAL Trial. *Eur J Cancer* (2020) 124:25–36. doi: 10.1016/j.ejca.2019.09.027

70. Mac Manus MP, Hicks RJ, Matthews JP, Wirth A, Rischin D, Ball DL. Metabolic (FDG-PET) Response After Radical Radiotherapy/ Chemoradiotherapy for Non-Small Cell Lung Cancer Correlates With Patterns of Failure. *Lung Cancer* (2005) 49(1):95–108. doi: 10.1016/j.lungcan.2004.11.024

71. Sachpekidis C, Anwar H, Winkler J, Kopp-Schneider A, Larribere L, Haberkorn U, et al. The Role of Interim 18F-FDG PET/CT in Prediction of Response to Ipilimumab Treatment in Metastatic Melanoma. *Eur J Nucl Med Mol Imaging* (2018) 45(8):1289–96. doi: 10.1007/s00259-018-3972-9

72. Evangelista L, Guarneri V, Conte PF. 18f-Fluoroestradiol Positron Emission Tomography in Breast Cancer Patients: Systematic Review of the Literature & Meta-Analysis. *Curr Radiopharm* (2016) 9(3):244–57. doi: 10.2174/1874471009666161019144950

73. Tan N, Oyoyo U, Bavadian N, Ferguson N, Mukkamala A, Calais J, et al. PSMA-Targeted Radiotracers Versus 18F Fluciclovine for the Detection of Prostate Cancer Biochemical Recurrence After Definitive Therapy: A

Systematic Review and Meta-Analysis. *Radiol* (2020) 296(1):44–55. doi: 10.1148/radiol.2020191689

74. Wester H-J. Nuclear Imaging Probes: From Bench to Bedside. *Clin Cancer Res* (2007) 13(12):3470–81. doi: 10.1158/1078-0432.CCR-07-0264

75. Boellaard R, O'Doherty MJ, Weber WA, Mottaghy FM, Lonsdale MN, Stroobants SG, et al. FDG PET and PET/CT: EANM Procedure Guidelines for Tumour PET Imaging: Version 1.0. *Eur J Nucl Med Mol Imaging* (2010) 37(1):181–200. doi: 10.1007/s00259-009-1297-4

76. Graham MM, Wahl RL, Hoffman JM, Yap JT, Sunderland JJ, Boellaard R, et al. Summary of the UPICT Protocol for 18F-FDG PET/CT Imaging in Oncology Clinical Trials. *J Nucl Med* (2015) 56(6):955–61. doi: 10.2967/jnumed.115.158402

77. Makris NE, Huisman MC, Kinahan PE, Lammertsma AA, Boellaard R. Evaluation of Strategies Towards Harmonization of FDG PET/CT Studies in Multicentre Trials: Comparison of Scanner Validation Phantoms and Data Analysis Procedures. *Eur J Nucl Med Mol Imaging* (2013) 40(10):1507–15. doi: 10.1007/s00259-013-2465-0

78. Jun W, Cong W, Xianxin X, Daqing J. Meta-Analysis of Quantitative Dynamic Contrast-Enhanced MRI for the Assessment of Neoadjuvant Chemotherapy in Breast Cancer. *Am Surg* (2019) 85(6):645–53. doi: 10.1177/000313481908500630

79. Hirashima Y, Yamada Y, Tateishi U, Kato K, Miyake M, Horita Y, et al. Pharmacokinetic Parameters From 3-Tesla DCE-MRI as Surrogate Biomarkers of Antitumor Effects of Bevacizumab Plus FOLFIRI in Colorectal Cancer With Liver Metastasis. *Int J Cancer* (2012) 130 (10):2359–65. doi: 10.1002/ijc.26282

80. Rata M, Collins DJ, Darcy J, Messiou C, Tunariu N, Desouza N, et al. Assessment of Repeatability and Treatment Response in Early Phase Clinical Trials Using DCE-MRI: Comparison of Parametric Analysis Using MR- and CT-Derived Arterial Input Functions. *Eur Radiol* (2016) 26(7):1991–8. doi: 10.1007/s00330-015-4012-9

81. 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):e101–21. doi: 10.1002/jmri.26518

82. Winfield JM, Tunariu N, Rata M, Miyazaki K, Jerome NP, Germuska M, et al. Extracranial Soft-Tissue Tumors: Repeatability of Apparent Diffusion Coefficient Estimates From Diffusion-Weighted MR Imaging. *Radiol* (2017) 284(1):88–99. doi: 10.1148/radiol.2017161965

83. FDA: Food and Drug Agency. *Health C for D and R. Artificial Intelligence and Machine Learning in Software as a Medical Device*. FDA (2021). Available at: https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device.

84. Avanzini S, Kurtz DM, Chabon JJ, Moding EJ, Hori SS, Gambhir SS, et al. A Mathematical Model of ctDNA Shedding Predicts Tumor Detection Size. *Sci Adv* (2020) 6(50):eabc4308. doi: 10.1126/sciadv.abc4308

85. Nabet BY, Esfahani MS, Moding EJ, Hamilton EG, Chabon JJ, Rizvi H, et al. Noninvasive Early Identification of Therapeutic Benefit From Immune Checkpoint Inhibition. *Cell* (2020) 183(2):363–76. doi: 10.1016/j.cell.2020.09.001

86. Dasari A, Morris VK, Allegra CJ, Atreya C, Benson AB, Boland P, et al. ctDNA Applications and Integration in Colorectal Cancer: An NCI Colon and Rectal-Anal Task Forces Whitepaper. *Nat Rev Clin Oncol* (2020) 17 (12):757–70. doi: 10.1038/s41571-020-0392-0

87. Devonshire AS, Whale AS, Gutteridge A, Jones G, Cowen S, Foy CA, et al. Towards Standardisation of Cell-Free DNA Measurement in Plasma: Controls for Extraction Efficiency, Fragment Size Bias and Quantification. *Anal Bioanal Chem* (2010) 406(26):6499–512. doi: 10.1007/s00216-014-7835-3

88. Fleischhacker M, Schmidt B, Weickmann S, Fersching DM, Leszinski GS, Siegele B, et al. Methods for Isolation of Cell-Free Plasma DNA Strongly Affect DNA Yield. *Clin Chim Acta* (2011) 412(23-24):2085–8. doi: 10.1016/j.cca.2011.07.011

89. Merker JD, Oxnard GR, Compton C, Diehn M, Hurley P, Lazar AJ, et al. Circulating Tumor DNA Analysis in Patients With Cancer: American Society of Clinical Oncology and College of American Pathologists Joint Review. *J Clin Oncol* (2018) 36(16):1631–41. doi: 10.1200/JCO.2017.76.8671

90. Tapia C, Aung PP, Roy-Chowdhuri S, Xu M, Ouyang F, Alshawa A, et al. Decrease in Tumor Content Assessed in Biopsies is Associated With Improved Treatment Outcome Response to Pembrolizumab in Patients With Rare Tumors. *J Immunother Cancer* (2020) 8(1). doi: 10.1136/jitc-2020-000665

91. Ryan R, Gibbons D, Hyland JMP, Treanor D, White A, Mulcahy HE, et al. Pathological Response Following Long-Course Neoadjuvant Chemoradiotherapy for Locally Advanced Rectal Cancer. *Histopathol* (2005) 47(2):141–6. doi: 10.1111/j.1365-2559.2005.02176.x

92. Voskuilen CS, Oo HZ, Genitsch V, Smit LA, Vidal A, Meneses M, et al. Multicenter Validation of Histopathologic Tumor Regression Grade After Neoadjuvant Chemotherapy in Muscle-Invasive Bladder Carcinoma. *Am J Surg Pathol* (2019) 43(12):1600–10. doi: 10.1097/PAS.0000000000001371

93. Jaraj SJ, Camparo P, Boyle H, Germain F, Nilsson B, Petersson F, et al. Intra- and Interobserver Reproducibility of Interpretation of Immunohistochemical Stains of Prostate Cancer. *Virchows Arch* (2009) 455(4):375–81. doi: 10.1007/s00428-009-0833-8

94. Stoler MH, Schiffman M. Atypical Squamous Cells of Undetermined Significance-Low-Grade Squamous Intraepithelial Lesion Triage Study (ALTS) Group. Interobserver Reproducibility of Cervical Cytologic and Histologic Interpretations: Realistic Estimates From the ASCUS-LSIL Triage Study. *JAMA* (2001) 285(11):1500–5. doi: 10.1001/jama.285.11.1500

95. Stattaus J, Hahn S, Gauler T, Eberhardt W, Mueller SP, Forsting M, et al. Osteoblastic Response as a Healing Reaction to Chemotherapy Mimicking Progressive Disease in Patients With Small Cell Lung Cancer. *Eur Radiol* (2009) 19(1):193–200. doi: 10.1007/s00330-008-1115-6

96. Messiou C, Cook G, Reid AHM, Attard G, Dearnaley D, de Bono JS, et al. The CT Flare Response of Metastatic Bone Disease in Prostate Cancer. *Acta Radiol* (2011) 52(5):557–61. doi: 10.1258/ar.2011.100342

97. Ciray I, Lindman H, Aström KG, Bergh J, Ahlström KH. Early Response of Breast Cancer Bone Metastases to Chemotherapy Evaluated With MR Imaging. *Acta Radiol* (2001) 42(2):198–206. doi: 10.1080/028418501127346503

98. Tombal B, Rezazadeh A, Therasse P, Van Cangh PJ, Vande Berg B, Lecouvet FE. Magnetic Resonance Imaging of the Axial Skeleton Enables Objective Measurement of Tumor Response on Prostate Cancer Bone Metastases. *Prostate* (2005) 65(2):178–87. doi: 10.1002/pros.20280

99. Padhani AR, Lecouvet FE, Tunariu N, Koh D-M, De Keyzer F, Collins DJ, et al. METastasis Reporting and Data System for Prostate Cancer: Practical Guidelines for Acquisition, Interpretation, and Reporting of Whole-Body Magnetic Resonance Imaging-Based Evaluations of Multiorgan Involvement in Advanced Prostate Cancer. *Eur Urol* (2017) 71(1):81–92. doi: 10.1016/j.eururo.2016.05.033

100. Van Nieuwenhove S, Van Damme J, Padhani AR, Vandecaveye V, Tombal B, Wuts J, et al. Whole-Body Magnetic Resonance Imaging for Prostate Cancer Assessment: Current Status and Future Directions. *J Magn Reson Imaging* (2020). doi: 10.1002/jmri.27485

101. JH O, Lodge MA, Wahl RL. Practical PERCIST: A Simplified Guide to PET Response Criteria in Solid Tumors 1.0. *Radiol* (2016) 280(2):576–84. doi: 10.1148/radiol.2016142043

102. Schmidkonz C, Cordes M, Schmidt D, Bäuerle T, Goetz TI, Beck M, et al. 68Ga-PSMA-11 PET/CT-Derived Metabolic Parameters for Determination of Whole-Body Tumor Burden and Treatment Response in Prostate Cancer. *Eur J Nucl Med Mol Imaging* (2018) 45(11):1862–72. doi: 10.1007/s00259-018-4042-z

103. Schmidkonz C, Cordes M, Goetz TI, Prante O, Kuwert T, Ritt P, et al. 68Ga-PSMA-11 PET/CT Derived Quantitative Volumetric Tumor Parameters for Classification and Evaluation of Therapeutic Response of Bone Metastases in Prostate Cancer Patients. *Ann Nucl Med* (2019) 33(10):766–75. doi: 10.1007/s12149-019-01387-0

104. Zhou J, Li Q, Cao Y. Spatiotemporal Heterogeneity Across Metastases and Organ-Specific Response Informs Drug Efficacy and Patient Survival in Colorectal Cancer. *Cancer Res* (2021) 81(9):2522–33. doi: 10.1158/0008-5472.CAN-20-3665

105. Pires da Silva I, Lo S, Quek C, Gonzalez M, Carlino MS, Long GV, et al. Site-Specific Response Patterns, Pseudoprogression, and Acquired Resistance in Patients With Melanoma Treated With Ipilimumab Combined With Anti-PD-1 Therapy. *Cancer* (2020) 126(1):86–97. doi: 10.1002/cncr.32522

106. Osorio JC, Arbour KC, Le DT, Durham JN, Plodkowski AJ, Halpenny DF, et al. Lesion-Level Response Dynamics to Programmed Cell Death Protein (PD-1) Blockade. *J Clin Oncol* (2019) 37(36):3546–55. doi: 10.1200/JCO.19.00709

107. Schmid S, Diem S, Li Q, Krapf M, Flatz L, Leschka S, et al. Organ-Specific Response to Nivolumab in Patients With Non-Small Cell Lung Cancer (NSCLC). *Cancer Immunol Immunother* (2018) 67(12):1825–32. doi: 10.1007/s00262-018-2239-4

108. Choi H, Charnsangavej C, de Castro Faria S, Tamm EP, Benjamin RS, Johnson MM, et al. CT Evaluation of the Response of Gastrointestinal Stromal Tumors After Imatinib Mesylate Treatment: A Quantitative Analysis Correlated With FDG PET Findings. *AJR Am J Roentgenol* (2004) 183(6):1619–28. doi: 10.2214/ajr.183.6.01831619

109. Choi H, Charnsangavej C, Faria SC, Macapinlac HA, Burgess MA, Patel SR, et al. Correlation of Computed Tomography and Positron Emission Tomography in Patients With Metastatic Gastrointestinal Stromal Tumor Treated at a Single Institution With Imatinib Mesylate: Proposal of New Computed Tomography Response Criteria. *J Clin Oncol* (2007) 25(13):1753–9. doi: 10.1200/JCO.2006.07.3049

110. Fournier L, Ammari S, Thiam R, Cuénod C-A. Imaging Criteria for Assessing Tumour Response: RECIST, mRECIST, Cheson. *Diagn Interv Imaging* (2014) 95(7–8):689–703. doi: 10.1016/j.diii.2014.05.002

111. Benjamin RS, Choi H, Macapinlac HA, Burgess MA, Patel SR, Chen LL, et al. We Should Desist Using RECIST, at Least in GIST. *J Clin Oncol* (2007) 25(13):1760–4. doi: 10.1200/JCO.2006.07.3411

112. Maas M, Beets-Tan R, Gaubert J-Y, Gomez Munoz F, Habert P, Klompenhouwer LG, et al. Follow-Up After Radiological Intervention in Oncology: ECIO-ESOI Evidence and Consensus-Based Recommendations for Clinical Practice. *Insights Imaging* (2020) 11(1):83. doi: 10.1186/s13244-020-00884-5

113. Sabet A, Meyer C, Aouf A, Sabet A, Ghamari S, Pieper CC, et al. Early Post-Treatment FDG PET Predicts Survival After 90Y Microsphere Radioembolization in Liver-Dominant Metastatic Colorectal Cancer. *Eur J Nucl Med Mol Imaging* (2015) 42(3):370–6. doi: 10.1007/s00259-014-2935-z

114. 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

115. Imseeh G, Giles SL, Taylor A, Brown MRD, Rivens I, Gordon-Williams R, et al. Feasibility of Palliating Recurrent Gynecological Tumors With MRGHIFU: Comparison of Symptom, Quality-of-Life, and Imaging Response in Intra and Extra-Pelvic Disease. *Int J Hyperthermia* (2021) 38(1):623–32. doi: 10.1080/02656736.2021.1904154

116. Ji Y, Zhu J, Zhu L, Zhu Y, Zhao H. High-Intensity Focused Ultrasound Ablation for Unresectable Primary and Metastatic Liver Cancer: Real-World Research in a Chinese Tertiary Center With 275 Cases. *Front Oncol* (2020) 10:519164. doi: 10.3389/fonc.2020.519164

117. Wolchok JD, Hoos A, O'Day S, Weber JS, Hamid O, Lebbe C, et al. Guidelines for the Evaluation of Immune Therapy Activity in Solid Tumors: Immune-Related Response Criteria. *Clin Cancer Res* (2009) 15(23):7412–20. doi: 10.1158/1078-0432.CCR-09-1624

118. Tazdait M, Mezquita L, Lahmar J, Ferrara R, Bidault F, Ammari S, et al. Patterns of Responses in Metastatic NSCLC During PD-1 or PDL-1 Inhibitor Therapy: Comparison of RECIST 1.1, irRECIST and iRECIST Criteria. *Eur J Cancer* (2018) 88:38–47. doi: 10.1016/j.ejca.2017.10.017

119. Park HJ, Kim KW, Pyo J, Suh CH, Yoon S, Hatabu H, et al. Incidence of Pseudoprogression During Immune Checkpoint Inhibitor Therapy for Solid Tumors: A Systematic Review and Meta-Analysis. *Radiol* (2020) 297(1):87–96. doi: 10.1148/radiol.2020200443

120. Seymour L, Bogaerts J, Perrone A, Ford R, Schwartz LH, Mandrekar S, et al. iRECIST: Guidelines for Response Criteria for Use in Trials Testing Immunotherapeutics. *Lancet Oncol* (2017) 18(3):e143–52. doi: 10.1016/S1470-2045(17)30074-8

121. Russo M, Crisafulli G, Sogari A, Reilly NM, Arena S, Lamba S, et al. Adaptive Mutability of Colorectal Cancers in Response to Targeted Therapies. *Sci* (2019) 366(6472):1473–80. doi: 10.1126/science.aav4474

122. Siravegna G, Lazzari L, Crisafulli G, Sartore-Bianchi A, Mussolin B, Cassingena A, et al. Radiologic and Genomic Evolution of Individual Metastases During HER2 Blockade in Colorectal Cancer. *Cancer Cell* (2018) 34(1):148–62.e7. doi: 10.1016/j.ccell.2018.06.004

123. Dong Z-Y, Zhai H-R, Hou Q-Y, Su J, Liu S-Y, Yan H-H, et al. Mixed Responses to Systemic Therapy Revealed Potential Genetic Heterogeneity and Poor Survival in Patients With Non-Small Cell Lung Cancer. *Oncologist* (2017) 22(1):61–9. doi: 10.1634/theoncologist.2016-0150

124. de Castro J, Cobo M, Isla D, Puente J, Reguart N, Cabeza B, et al. Recommendations for Radiological Diagnosis and Assessment of Treatment Response in Lung Cancer: A National Consensus Statement by the Spanish Society of Medical Radiology and the Spanish Society of Medical Oncology. *Clin Transl Oncol* (2015) 17(1):11–23. doi: 10.1007/s12094-014-1231-5

125. Ferretti GR, Reymond E, Delouche A, Sakhri L, Jankowski A, Moro-Sibilot D, et al. Personalized Chemotherapy of Lung Cancer: What the Radiologist Should Know. *Diagn Interv Imaging* (2016) 97(3):287–96. doi: 10.1016/j.diii.2015.11.013

126. Crabb SJ, Patsios D, Sauerbrei E, Ellis PM, Arnold A, Goss G, et al. Tumor Cavitation: Impact on Objective Response Evaluation in Trials of Angiogenesis Inhibitors in Non-Small-Cell Lung Cancer. *J Clin Oncol* (2009) 27(3):404–10. doi: 10.1200/JCO.2008.16.2545

127. Marom EM, Martinez CH, Truong MT, Lei X, Sabloff BS, Munden RF, et al. Tumor Cavitation During Therapy With Antiangiogenesis Agents in Patients With Lung Cancer. *J Thorac Oncol* (2008) 3(4):351–7. doi: 10.1097/JTO.0b013e318168c7e9

128. Kong BY, Menzies AM, Saunders CAB, Liniker E, Ramanujam S, Guminski A, et al. Residual FDG-PET Metabolic Activity in Metastatic Melanoma Patients With Prolonged Response to Anti-PD-1 Therapy. *Pigment Cell Melanoma Res* (2016) 29(5):572–7. doi: 10.1111/pcmr.12503

129. Rubin DL, Willrett D, O'Connor MJ, Hage C, Kurtz C, Moreira DA. Automated Tracking of Quantitative Assessments of Tumor Burden in Clinical Trials. *Transl Oncol* (2014) 7(1):23–35. doi: 10.1593/tlo.13796

130. Barash Y, Klang E. Automated Quantitative Assessment of Oncological Disease Progression Using Deep Learning. *Ann Transl Med* (2019) 7(Suppl 8):S379. doi: 10.21037/atm.2019.12.101

131. Kickingereder P, Isensee F, Tursunova I, Petersen J, Neuberger U, Bonekamp D, et al. Automated Quantitative Tumour Response Assessment of MRI in Neuro-Oncology With Artificial Neural Networks: A Multicentre, Retrospective Study. *Lancet Oncol* (2019) 20(5):728–40. doi: 10.1016/S1470-2045(19)30098-1

132. Baidya Kayal E, Kandasamy D, Yadav R, Bakhshi S, Sharma R, Mehndiratta A. Automatic Segmentation and RECIST Score Evaluation in Osteosarcoma Using Diffusion MRI: A Computer Aided System Process. *Eur J Radiol* (2020) 133:109359. doi: 10.1016/j.ejrad.2020.109359

133. Tang Y, Yan K, Xiao J, Summers RM. "One Click Lesion RECIST Measurement and Segmentation on CT Scans". In: AL Martel, P Abolmaesumi, D Stoyanov, D Mateus, MA Zuluaga, SK Zhou, et al. editors. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*, vol. p . Cham: Springer International Publishing (2020). p. 573–83. (Lecture Notes in Computer Science).

134. Moawad AW, Fuentes D, Khalaf AM, Blair KJ, Szklaruk J, Qayyum A, et al. Feasibility of Automated Volumetric Assessment of Large Hepatocellular Carcinomas' Responses to Transarterial Chemoembolization. *Front Oncol* (2020) 10:572. doi: 10.3389/fonc.2020.00572

135. Iannessi A, Beaumont H, Liu Y, Bertrand A-S. RECIST 1.1 and Lesion Selection: How to Deal With Ambiguity at Baseline? *Insights Imaging* (2021) 12(1):36. doi: 10.1186/s13244-021-00976-w

136. Ruchalski K, Braschi-Amirfarzan M, Douek M, Sai V, Gutierrez A, Dewan R, et al. A Primer on RECIST 1.1 for Oncologic Imaging in Clinical Drug Trials. *Radiol Imaging Cancer* (2021) 3(3):e210008. doi: 10.1148/rycan.2021210008

137. Kantarjian HM, Fojo T, Mathisen M, Zwelling LA. Cancer Drugs in the United States: Justum Pretium–the Just Price. *J Clin Oncol* (2013) 31(28):3600–4. doi: 10.1200/JCO.2013.49.1845

138. Davis C, Naci H, Gurpinar E, Poplavska E, Pinto A, Aggarwal A. Availability of Evidence of Benefits on Overall Survival and Quality of Life of Cancer Drugs Approved by European Medicines Agency: Retrospective Cohort Study of Drug Approvals 2009-13. *BMJ* (2017) 359:j4530. doi: 10.1136/bmj.j4530

139. Cheson BD, Fisher RI, Barrington SF, Cavalli F, Schwartz LH, Zucca E, et al. Recommendations for Initial Evaluation, Staging, and Response Assessment of Hodgkin and Non-Hodgkin Lymphoma: The Lugano Classification. *J Clin Oncol* (2014) 32(27):3059–68. doi: 10.1200/JCO.2013.54.8800

140. Younes A, Hilden P, Coiffier B, Hagenbeek A, Salles G, Wilson W, et al. International Working Group Consensus Response Evaluation Criteria in Lymphoma (RECIL 2017). *Ann Oncol* (2017) 28(7):1436–47. doi: 10.1093/annonc/mdx097

141. van den Bent MJ, Wefel JS, Schiff D, Taphoorn MJB, Jaeckle K, Junck L, et al. Response Assessment in Neuro-Oncology (a Report of the RANO Group): Assessment of Outcome in Trials of Diffuse Low-Grade Gliomas. *Lancet Oncol* (2011) 12(6):583–93. doi: 10.1016/S1470-2045(11)70057-2

142. Lin NU, Lee EQ, Aoyama H, Barani IJ, Barboriak DP, Baumert BG, et al. Response Assessment Criteria for Brain Metastases: Proposal From the RANO Group. *Lancet Oncol* (2015) 16(6):e270–278. doi: 10.1016/S1470-2045(15)70057-4

143. Shafrin J, Brookmeyer R, Peneva D, Park J, Zhang J, Figlin RA, et al. The Value of Surrogate Endpoints for Predicting Real-World Survival Across Five Cancer Types. *Curr Med Res Opin* (2016) 32(4):731–9. doi: 10.1185/03007995.2016.1140027

**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.

# Assessing Histology Structures by *Ex Vivo* MR Microscopy and Exploring the Link Between MRM-Derived Radiomic Features and Histopathology in Ovarian Cancer

Marion Tardieu[1], Yulia Lakhman[2], Lakhdar Khellaf[3], Maida Cardoso[4], Olivia Sgarbura[1,5], Pierre-Emmanuel Colombo[5], Mireia Crispin-Ortuzar[6], Evis Sala[6,7], Christophe Goze-Bac[4] and Stephanie Nougaret[1,8]*

[1] Montpellier Cancer Research Institute (IRCM), INSERM U1194, University of Montpellier, Montpellier, France, [2] Department of Radiology, Memorial Sloan Kettering Cancer Center, New York, NY, United States, [3] Department of Pathology, Montpellier Cancer Institute (ICM), Montpellier, France, [4] BNIF Facility, L2C, UMR 5221, CNRS, University of Montpellier, Montpellier, France, [5] Department of Surgery, Montpellier Cancer Institute (ICM), Montpellier, France, [6] Cancer Research UK, Cambridge Institute, University of Cambridge, Cambridge, United Kingdom, [7] Department of Radiology, University of Cambridge, Cambridge, United Kingdom, [8] Department of Radiology, Montpellier Cancer Institute (ICM), Montpellier, France

The value of MR radiomic features at a microscopic scale has not been explored in ovarian cancer. The objective of this study was to probe the associations of MR microscopy (MRM) images and MRM-derived radiomic maps with histopathology in high-grade serous ovarian cancer (HGSOC). Nine peritoneal implants from 9 patients with HGSOC were imaged *ex vivo* with MRM using a 9.4-T MR scanner. All MRM images and computed pixel-wise radiomics maps were correlated with the slice-matched stroma and tumor proportion maps derived from whole histopathologic slide images (WHSI) of corresponding peritoneal implants. Automated MRM-derived segmentation maps of tumor and stroma were constructed using holdout test data and validated against the histopathologic gold standard. Excellent correlation between MRM images and WHSI was observed (Dice index = 0.77). Entropy, correlation, difference entropy, and sum entropy radiomic features were positively associated with high stromal proportion ($r = 0.97, 0.88, 0.81,$ and $0.96$ respectively, $p < 0.05$). MR signal intensity, energy, homogeneity, auto correlation, difference variance, and sum average were negatively associated with low stromal proportion ($r = -0.91, -0.93, -0.94, -0.9, -0.89, -0.89,$ respectively, $p < 0.05$). Using the automated model, MRM predicted stromal proportion with an accuracy ranging from 61.4% to 71.9%. In this hypothesis-generating study, we showed that it is feasible to resolve histologic structures in HGSOC using *ex vivo* MRM at 9.4 T and radiomics.

Keywords: ovarian cancer, MRI, radiomics, machine learning, histology

# 1 INTRODUCTION

High-grade serous ovarian cancer (HGSOC) is the most prevalent histological subtype of ovarian cancer (1). Advanced-stage HGSOC is often approached with neoadjuvant chemotherapy to reduce tumor burden followed by cytoreductive surgery with or without hyperthermic intraperitoneal chemotherapy (HIPEC) (2, 3). Response to neoadjuvant treatment at histopathology manifests as an increase in stromal tissue and decrease in tumor cells, but those changes cannot be assessed on standard cross-sectional imaging (4). Indeed, CT only provides anatomic information, whereas MRI captures both anatomic and functional data. Preliminary results in ovarian cancer suggest that the quantitative parameters derived from diffusion-weighted MR imaging (DWI-MRI) may serve as a biomarkers of cell density (5–7). For example, the increase in ADC values in HGSOC peritoneal implants (decrease in cell density) during neoadjuvant chemotherapy was associated with good treatment response as assessed by RECIST 1.1 criteria and CA125 level (8).

To date, few studies have explored the associations between CT/MR images which interrogate tumor at macroscopic scale and histopathologic images which depict tumor at microscopic scale. In contrast, studies have recently focused on the role of radiomics. Radiomic analysis extracts a large amount of quantitative data and, thus, has the potential to uncover salient features that are imperceptible to human observers yet possibly reflective of microscopic changes in tumor in response to treatment (9–13). Preliminary work in CT found that higher image-based tumor heterogeneity was associated with worse prognosis and greater risk of incomplete surgical resection in HGSOC (14).

However, image-based tumor heterogeneity and radiomics features have not yet been exactly correlated with findings at histopathology. Exploring the associations between image-based tumor heterogeneity and biologic underpinnings at histopathology would offer an important avenue for monitoring response in time and space non-invasively using imaging, as "virtual biopsy". For example, it has been recently reported that a high content of stroma present in HGSOC was associated with a high pathologic stage at diagnosis and displayed a reduced overall survival and poor prognosis independently from the histology type (15, 16). For a long time, the potential role in carcinogenesis of stromal cells has been neglected, as they were regarded just as part of an inflammatory reaction induced by necrotic cancer cells. It is now recognized that the stroma composition and architecture, in terms of vascularization, type of cells, and their secretion, play a role in the establishment and progression of cancer cells. It is also now well established that the stroma contributes to ovarian tumorigenesis and progression (17). Concerning the recent clinical radiomics model, Lu et al. found a radiomics model associated with DFS and genomic pathway (18). This radiomic model was positively correlated with a stroma marker, the fibronectin, and associated with a proportion of tumor-associated stromal cells and patient prognosis (18). Being able to assess the stroma–tumor ratio in a non-invasive way with radiomics may open a new pathway in assessing tumor response in HGSOC.

Beyond the analysis of standard and functional MR imaging, *ex vivo* high-field MRI (i.e., magnetic resonance microscopy, MRM) offers a unique opportunity to probe the tumor at the microscopic scale because MRM can reach a resolution of 40 μm. Although much lower than that of optical microscopy (0.25 μm), this resolution is significantly higher compared to standard MRI (1 mm) allowing the visualization of histological details. Work performed *ex vivo* at 7.0 T was able to visualize distinctive features of both benign and malignant lesions in breast tissue (19). In prostate, Fan et al. have evaluated the feasibility of 9.4-T *ex vivo* MRI to guide pathologists' examination in the evaluation of prostate cancer (20). They demonstrated excellent anatomical detail as well as significant T2 values and ADC differences between cancer and normal prostatic tissues (20). Durand et al. performed *ex vivo* MRI with direct histological correlation of the prostate gland that approached histological spatial resolution enabling the visualization of gland microanatomy with MRM (21). More recently, a study has investigated the ability of *ex vivo* 7.0-T MRI to localize prostate cancer and to predict the margin status in fresh radical prostatectomy specimens using histology as the reference standard (22). The author found that *ex vivo* MRI was able to accurately localize prostate cancer in radical prostatectomy specimens, and the technique provided information on the margin status (22). In brain tumors, Martinez-Bisbal et al. combined MRM with MR spectroscopy to study metabolites of *ex vivo* tumors at high resolution (23). However, to our knowledge the role of MRM in ovarian cancer remains unexplored.

The aim of this proof-of-concept study was to evaluate the correlation between tumor-stroma maps derived from MRM and whole histopathology slide images (WHSI) of histopathologic specimens and to develop an automated visual map of stromal proportion in HGSOC peritoneal implants using quantitative analysis of MRM images.

# 2 MATERIALS AND METHODS

## 2.1 Patient Inclusion

Nine patients (mean age 65.6 ± 8.2 years) with advanced HGSOC referred to surgery were included in the study; among them, 2 had primary debulking surgery and 7 had interval debulking surgery after neoadjuvant chemotherapy. Patient characteristics are summarized in **Table 1**. The Institutional Review Board approved this prospective study, and all patients signed the written informed consent form prior to enrollment. **Figure 1** shows the study experimental workflow.

## 2.2 Biopsy Preparation and Evaluation

For each patient, index peritoneal implant was selected *a priori* based on the review of preoperative MRI in conjunction with the surgeon. Resected implants were transported from the operating room to the pathology laboratory per routine procedure where they were transformed into $3 \times 1.5 \times 1$ cm$^3$ blocks and fixed in formalin solution. Prior to MRM acquisitions, fixed specimens were soaked in saline solution with 1% Gd-BOPTA (MultiHance; Bracco Imaging, Milan, Italy) during 1 h. The implants were then

**TABLE 1 |** Patients characteristics.

| | |
|---|---|
| Patient age (mean + SD) | 65.6 ± 8.2 |
| BRCA-mutant | |
| Yes | 1 |
| No | 8 |
| Neoadjuvant chemotherapy | |
| Yes | 7 |
| No | 2 |

laid on a plastic plate inside a tube filled with a perfluorocarbon solution (Fluorinert FC-40, 3M™ Electronic Liquids, Saint Paul, USA) to reduce susceptibility artifacts induced by air. After imaging, the specimens were transported back to the pathology laboratory for histological evaluation.

## 2.3 MRM Technique and Radiomic Feature Extraction

MRM acquisitions were performed on a 9.4-T MR scanner (Agilent Varian 9.4/160/ASR, CA, USA) associated with the VnmrJ imaging acquisition system (Agilent, Palo Alto, California, USA) and using a dedicated ribbon solenoid coil (24). MRI experiments included two sets of acquisitions: a high-resolution and a $90 \times 90 \times 180$ μm$^3$ T1-weighted fat-suppressed gradient echo images. The later set of images was selected to facilitate radiomic analysis by allowing to interpolate images to an isotropic voxel spacing with a minimum interpolation factor (25, 26). All MRM image acquisition parameters are summarized

in **Table 2**. Prior to radiomic analysis, MRM images were preprocessed using the open-source software 3D slicer (27) (http://www.slicer.org) and by following IBSI guidelines (26). Briefly, noise filtering, bias field correction, image interpolation to isotropic voxel size, and intensity outlier filtering [μ ± 3σ] were performed. Texture feature maps were then extracted on a per-pixel basis with an in-house software implemented in Matlab (The MathWorks, Natick, MA, USA). The gray-level co-occurrence matrix (GLCM) (28) was computed in 2D for each pixel using the 3 neighboring pixels of each direction around it. The dynamic range of signal intensities was reduced to 64 gray levels. The GLCM, $p(i,j)$, represents the spatial relationship of pixels by measuring the occurrence between a pixel $i$ with a certain intensity with a pixel $j$ of another intensity along the 13 directions in 3D. Thirteen Haralick feature maps were extracted from GLCM: energy, contrast, entropy, homogeneity, dissimilarity, correlation, variance, sum average, sum entropy, difference variance, difference entropy, autocorrelation, and cluster tendency. The total time acquisition was 1 h, and images were obtained usually within 3 h of resection.

### 2.3.1 Tissue Scanning

After MRM experiments, each block was sent to the pathology laboratory and cut into 4-μm sections parallel to MR sections, discerned by the plastic plate. Histologic specimens were stained using hematoxylin–eosin–saffron (HES) stain, and whole-slice sections were then scanned into digital data with an automated
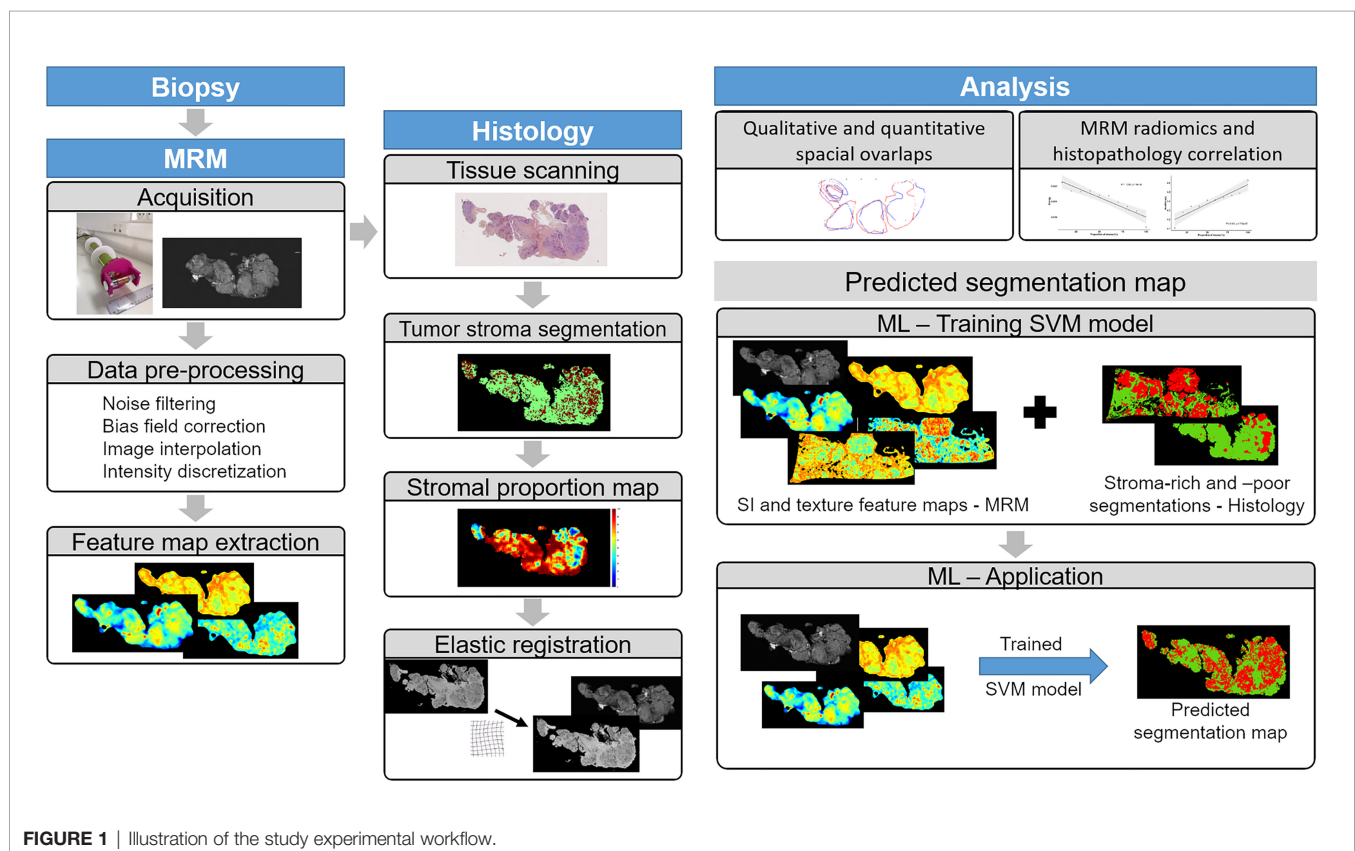


**FIGURE 1 |** Illustration of the study experimental workflow.

**TABLE 2 |** Sequences parameters.

| Acquisition name | High-resolution | For radiomics analysis |
|---|---|---|
| Tr/TE (ms) | 2000/9.14 | |
| Flip angle | 60° | |
| Averages | 4 | |
| Matrix | 512 × 256 | |
| FOV | From 26 × 13 to 32 × 16 mm$^2$ | 46 × 23 mm$^2$ |
| Resolution | From 51 × 51 to 62.5 × 62.5 µm$^2$ | 90 × 90 µm$^2$ |
| Slice/thickness | 30/300 µm | 40/180 µm |
| Scan time | 34 min 12 s | 17 min 8 s |

whole-slide scanner (NanoZoomer-XR scanner C12000, Hamamatsu, Japan), at ×20 magnification and a pixel size of 0.46 µm.

### 2.3.2 Tumor Stroma Segmentation

Stoma and tumor regions on WHSI were identified and segmented with the open-source software QuPath (29). In order to correlate MRM and WHSI, as MR texture features were estimated for each pixel, stromal proportion was also locally calculated. Thus, from the tumor-stroma segmentation WHSI map, stromal proportion was assigned to each pixel by measuring this proportion in a circular neighborhood with a radius of 3 pixels. Stromal proportion calculation was not possible for two of the nine specimens. One of them had mostly fat tissue, causing too few tumor and stroma tissues to be correlated with corresponding MRM slices (specimen iii. on **Supplementary Material Figure S1**). The second tumor was resected after chemotherapy and had a significant necrosis, with the consequence that neither tumor tissue nor stroma remained (specimen iv. on **Supplementary Material Figure S1**).

### 2.3.3 Elastic Registration

Due to histological fixation and sectioning, elastic deformations existed between histology sections and corresponding MRM slices. To remedy this and compare those images at the corresponding pixel, manual 3D non-linear co-registration was performed using the 3D slicer. Three specimens could not be precisely co-registered: deformations were too important, making the precise pixel-wise registration impossible between WHSI and MRM images (specimen ii. on **Figure 2** and specimens ii. and v. on **Supplementary Material Figure S1**).

## 2.4 Statistical Analysis

### 2.4.1 MRM Image and WHSI Comparison

A qualitative visual assessment of high-resolution MR images and texture maps with their corresponding WHSI of the nine *ex vivo* peritoneal implants was evaluated by the pathologist and the radiologist. Spatial overlaps between WHSI and MRM images were measured and quantified using the Dice similarity coefficient using the following formula: $DSC = 2\,(A \cap B)\,/\,(A + B)$.

### 2.4.2 MRM Radiomics and Histopathology Pixel-Wise Correlation

For the MRM and histology correlation, 5 implants had to be excluded: three of them could not be precisely co-registered, two for which the stromal proportion maps could not be extracted.

Clinical characteristics of these four remaining tumors were 0 BRCA mutant and 2 obtained after neoadjuvant chemotherapy. MRM signal intensity (SI) and texture maps were compared pixel by pixel to the stromal proportion map from WHSI. Stromal proportion maps from WHSI were divided into increments of 10 equal percentage points, and mean SI and texture values were computed for these 10 regions. Pearson coefficients were calculated to evaluate the correlation between mean texture feature values and stromal proportion. $p$-values less than 0.05 were considered statistically significant. All statistical analyses were performed with R software version 4.0.0.

### 2.4.3 Predicted MR Segmentation Maps

Pixels of the four remaining tumors were classified as stroma-rich, i.e., stromal proportion >50%, and stroma-poor, i.e., stromal proportion <50%, Of these pixels, 30% were randomly selected from the four tumors, to generate a classification table of 58,945 pixels associated with 14 inputs (13 texture features and the label). Of these selected pixels, 50% were randomly chosen to train the machine learning model with balanced number of stroma-rich and -poor pixels; the 50% remaining pixels were used to test the algorithm (29,473 pixels). The classification model was completed using the support vector machine (SVM) classifier and 20-fold cross-validation, with the classification learner toolbox of Matlab 2020a. SVM is a supervised learning model that classifies by mapping the input data into a higher-dimensional space allowing to find a hyperplane separator. The SVM training model was exported and used to measure the confusion matrix and accuracy value on the test set. Finally, this trained algorithm was applied to the 4 tumors separately to generate predicted segmentation maps and were compared to the stroma-rich and -poor segmentations, extracted from the histopathologic images.

## 3 RESULTS

## 3.1 Comparison Between MRM and Histopathology Reading

### 3.1.1 Visual Assessment, Qualitative Analysis

All nine implants were evaluated for visual MRM assessment and subsequently compared with histopathology. In all 9 cases, the pathologist was able to identify different relevant histological structures (i.e., tumor, stroma and fat) on MRM images. On MRM images, tumor cells appeared as high signal intensity areas separated by lower signal intensity foci representing fibrous
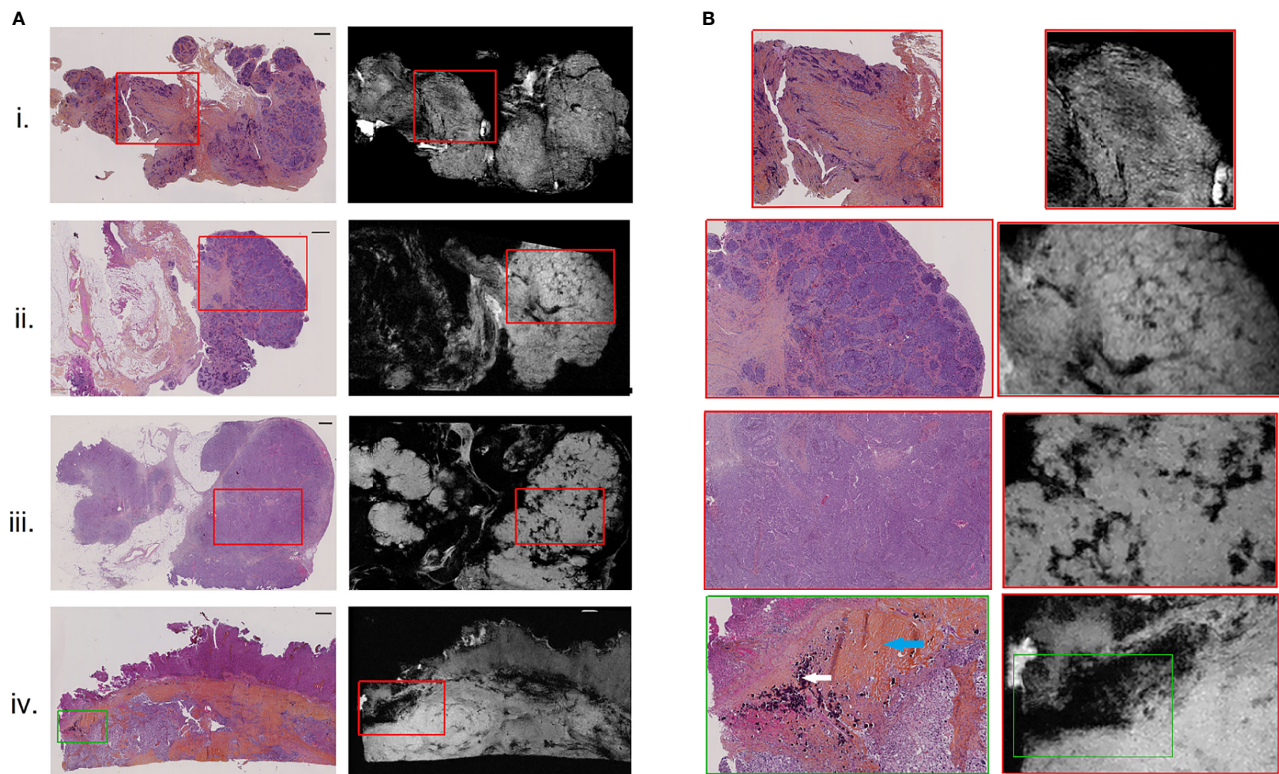
**FIGURE 2** | H&E-stained histological images **(A)**, left) with corresponding high-resolution MR images **(A)**, right), for 4 of the 9 resected peritoneal implants (i–iv). Magnified regions **(B)** of histological (left) and high-resolution MR (right) images, indicating by red and green boxes on **(A)**. Scale bar at top right of histological images **(A)** is 1 mm. White and blue arrows on the histologically magnified region **(B)**. (iv) Indicated respectively psammoma bodies and hyaline stroma.

stroma (**Figure 2**.i). In some cases, massive tumor infiltration appeared clustered in lobules separated by stroma, giving the appearance of a cauliflower (**Figure 2**.ii), well recognized on both MRM and histology images. Interestingly, three histological structures resulted on signal loss on MRM images. Glandular lumens, forming slit-like spaces inside dense tumor regions (**Figure 2**.iii), and psammoma bodies that are round collections of calcium did not produce any MR signal (**Figure 2**.iv, white arrow). Finally, regions of hyaline stroma, composed of hypocellular old collagen, were also associated with a signal loss area in MRM images (**Figure 2**.iv, blue arrow). Histology and MRM images of the implants not shown in **Figure 2** are presented on **Supplementary Material Figure S1**.

### 3.1.2 Quantitative Analysis
A spatial overlap between tumor, stroma, and fat regions on MRM images and at histopathology was measured (**Figure 3**). The Dice similarity coefficient was 77% for the entire dataset.

## 3.2 Association Between Radiomics and Histopathology Features: Quantitative Assessment
*Ex vivo* MRM radiomics features were extracted, and texture maps were generated for each of the four implants. **Figure 4** illustrates

energy, entropy, and homogeneity maps from a representative peritoneal carcinomatosis implant, with corresponding tumor-stroma segmentation and stromal proportion maps. The stromal proportion map was divided into increments of 10 percentage points, and mean texture values were calculated for these 10 regions. Correlation plots between mean texture feature values and stromal proportion were constructed (**Figure 5**), and Pearson's correlation coefficients with corresponding p-values were then determined (**Table 3**). Pearson correlation coefficients ranged from 0.47 to 0.97 with p-values significant for 10 of 13 features. Entropy, correlation, difference entropy, and sum entropy radiomic features were positively associated with high stromal proportion (r = 0.97, 0.88, 0.81, and 0.96, respectively, $p < 0.05$). MR signal intensity, energy, homogeneity, auto correlation, difference variance, and sum average radiomic features were negatively associated with low stromal proportion and as such linked with higher tumor proportion (r = –0.91, –0.93, –0.94, –0.9, –0.89, and –0.89, respectively, $p < 0.05$).

## 3.3 Estimated Segmentation Map
Stroma-rich and -poor histological segmentations (proportion of stroma >50% and <50%) were defined from the stromal proportion map, allowing to train an SVM model with MR texture maps as inputs. Boxplots showing classified features values between >50%
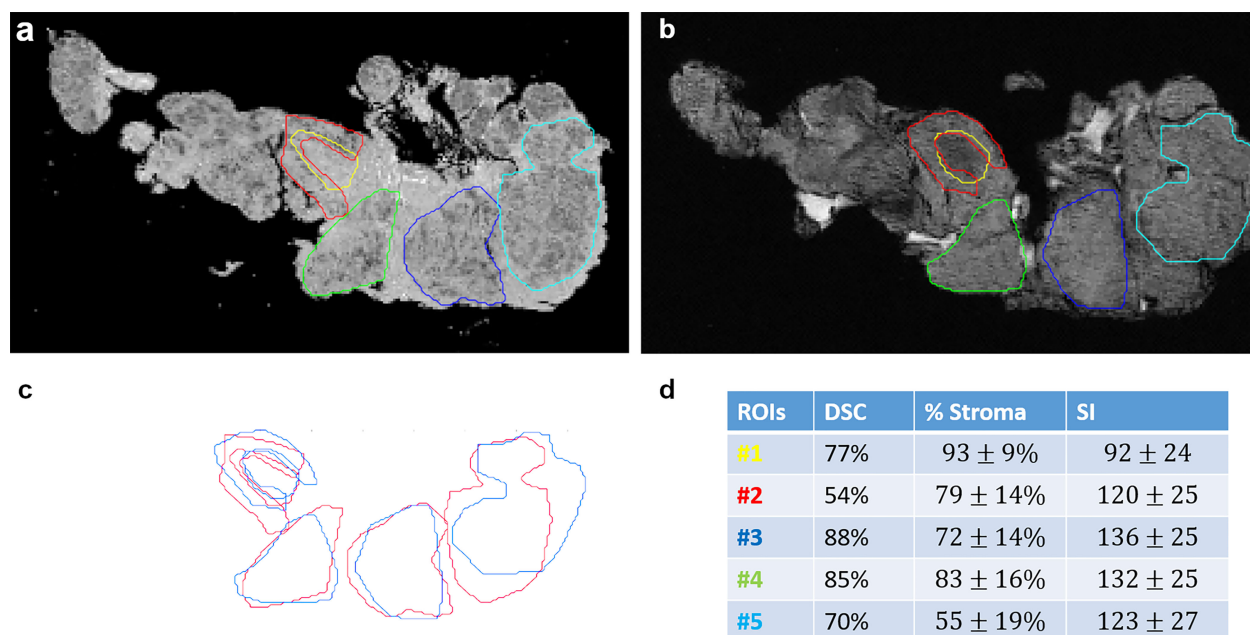
**FIGURE 3** | Spatial overlaps between histological images **(A)** and MR images **(B)** for stroma and tumor portion. ROIs were manually drawn on MRM and whole slide images and then superposed in **(C)**. Dice similarity index (DSI), stroma proportion, and signal intensity are presented in a table **(D)** for each ROI.
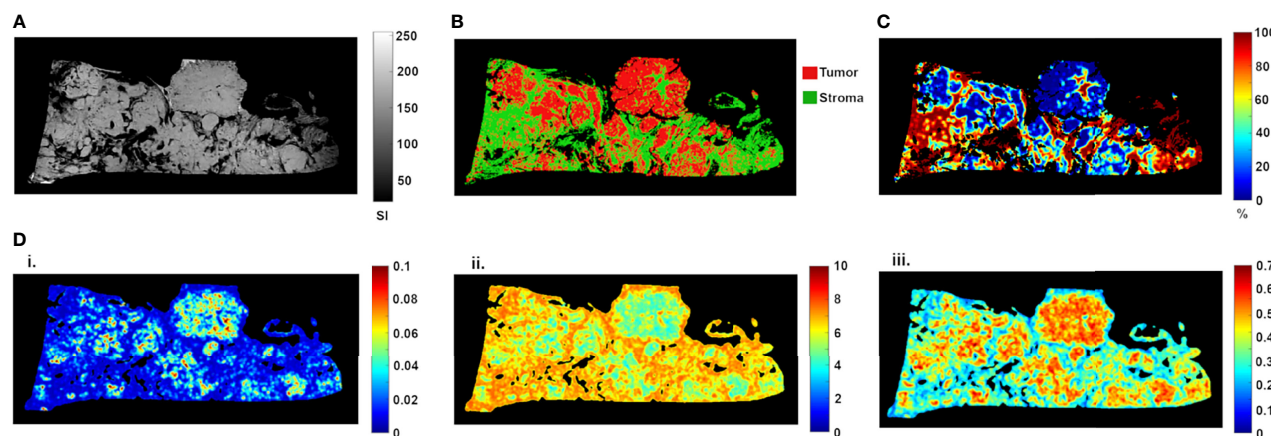


**FIGURE 4** | High-resolution MR image **(A)** of resected peritoneal implant with corresponding tissue segmentation map **(B)** and stromal proportion map **(C)**, in %). Corresponding texture maps **(D)** for energy (i), entropy (ii), and homogeneity (iii) features. For this implant, high tumor proportion (low proportion of stroma) was associated with higher energy (from 0.029 ± 0.020 to 0.019 ± 0.014, t-test p = 0.0002), homogeneity (from 0.44 ± 0.10 to 0.38 ± 0.10, t-test p = 0.0001) and signal intensity (from 141.41 ± 28.30 to 116.09 ± 41.52, t-test p = 0.0002) scores (**Figure 3**). In contrast, high stromal proportion (low tumor proportion) was associated with higher entropy score (from 5.87 ± 1.00 to 6.47 ± 0.83, t-test p = 0.0002).

and <50% are presented in supplementary material **Figure S2**. The estimated segmentation maps were performed for the 4 samples, using the training SVM model, and are presented in **Figure 6** with their corresponding stroma-rich and -poor histological segmentations, allowing confusion matrix calculation (**Table 4**). An accuracy for predicting stromal proportion from MRM images ranged from 61.4% to 71.9% on the holdout test data.

## 4 DISCUSSION

This study demonstrates the feasibility of obtaining *ex vivo* images of peritoneal implants from HGSOC at 9.4 T at a resolution of 40 μm. Slice-matched MRM demonstrated strong structural similarities compared to whole-slide histology specimens. Tumor, stroma, and adipose tissues were all apparent on MRM images,
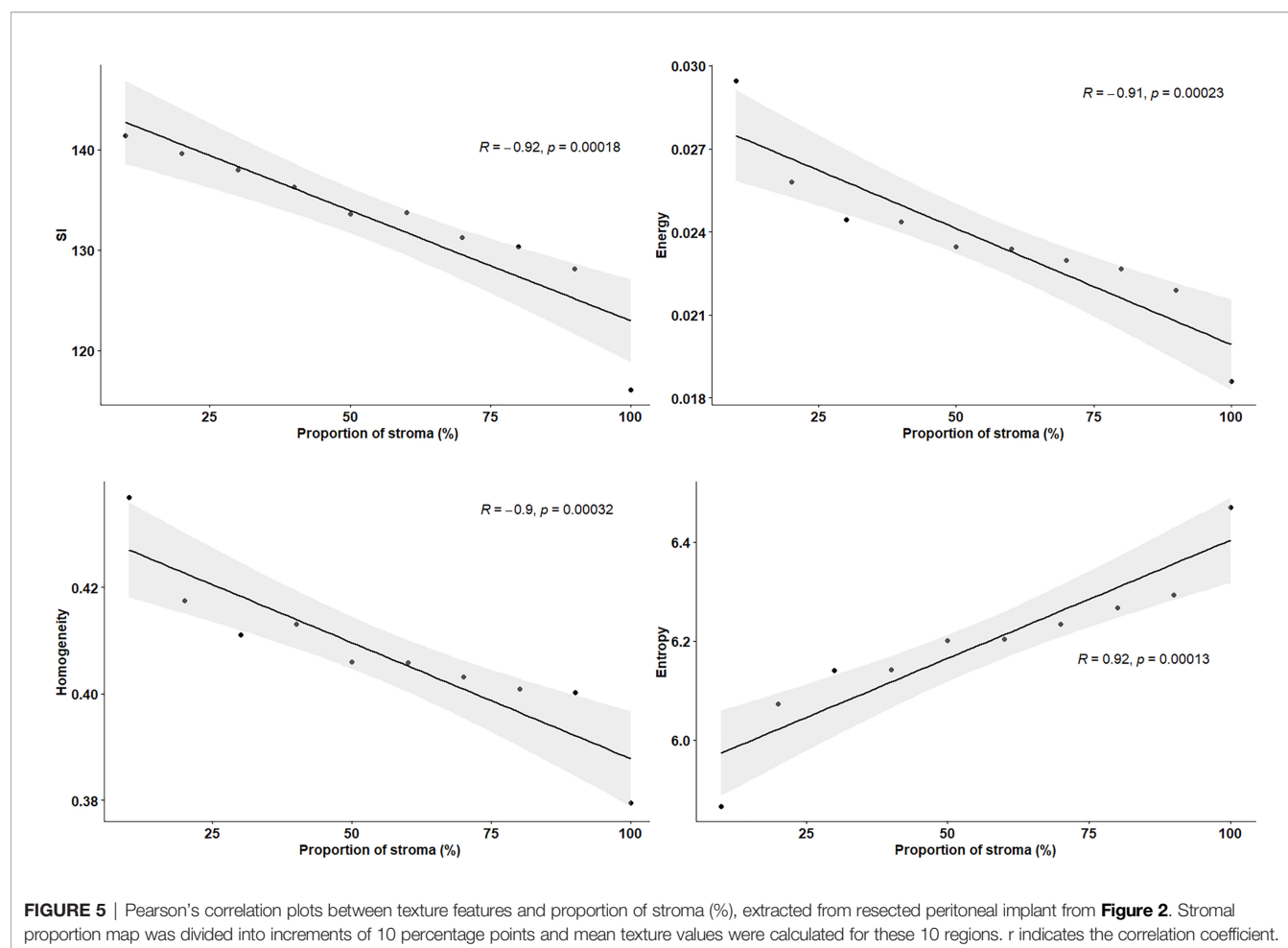
**FIGURE 5** | Pearson's correlation plots between texture features and proportion of stroma (%), extracted from resected peritoneal implant from **Figure 2**. Stromal proportion map was divided into increments of 10 percentage points and mean texture values were calculated for these 10 regions. r indicates the correlation coefficient.

**TABLE 3** | Pearson's correlation (r) between texture features and stromal proportion.

|                 | SI           | Constrat    | Dissimilarity | Energy       | Entropy      |
| --------------- | ------------ | ----------- | ------------- | ------------ | ------------ |
| r               | -0.91        | -0.76       | 0.47          | -0.93        | 0.97         |
| $p$-value       | <0.0001      | 0.011       | 0.17          | <0.0001      | <0.0001      |
|                 | **Homogeneity** | **AutoCorrel** | **ClusterTend** | **Correlation** | **DiffEntropy** |
| r               | -0.94        | -0.9        | 0.68          | 0.88         | 0.81         |
| $p$-value       | <0.0001      | 0.0005      | 0.029         | 0.0007       | 0.0046       |
|                 | **DiffVariance** | **SumAver** | **SumEntropy** | **Variance** |              |
| r               | -0.89        | -0.89       | 0.96          | 0.56         |              |
| $p$-value       | 0.0005       | 0.0006      | <0.0001       | 0.093        |              |

confirmed on their slice-matched histological samples. In addition, we found that MRM-derived radiomic features were able to discriminate tumor tissue, stroma, and adipose structures and could evaluate the stroma–tumor ratio.

To date, few studies have imaged tumor at a microscopic scale. Preliminary studies in breast, prostate and brain cancer demonstrated excellent correlation between MRM images and histopathologic characteristics of malignant tissue (19–23, 30, 31). Similarly, for the first time in HGSOC, we were able to obtain a direct correlation between histology and MRM images at 9.4 T.

In our study, we also correlated radiomics features with the tumor/stromal proportion pixel by pixel, from four tumors. Besides the extensive work performed in MRI on radiomics in all fields of study, the link between data extracted from the image and the pathology results is still lacking as radiomics has never been evaluated at a microscopic scale. For example, Vargas et al. found with CT that inter- and intratumor heterogeneity was linked to poor prognosis in HGSOC (14). Rizzo et al. evaluated whether CT radiomics features extracted from the primary tumor alone or combined with clinical data were associated with residual
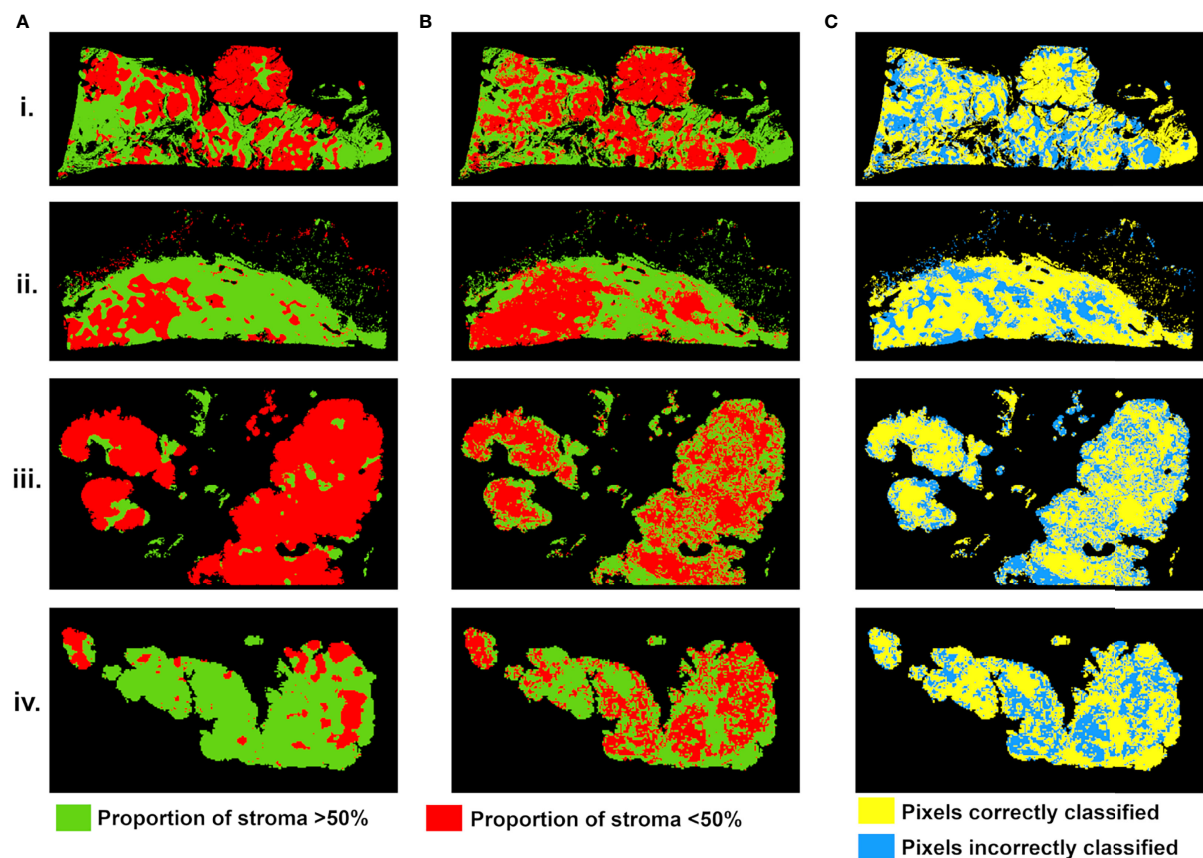
**FIGURE 6** | Stroma-rich (>50%) and -poor (<50%) histological segmentations **(A)** and predicted segmentation map from MR feature maps **(B)**, with pixels in green where stromal proportion >50% (stroma-rich) and in red where stromal proportion <50% (stroma-poor). **(C)** Evaluation map with pixels correctly classified in yellow and incorrectly classified in blue. For four tumors from (i–iv).

**TABLE 4** | Confusion matrix results with TPR, true positive rate; TNR, true negative rate; FPR, false positive rate; FNR, false negative rate.

|          | Accuracy | TPR   | TNR   | FPR   | FNR   |
|----------|----------|-------|-------|-------|-------|
| Test set | 66.6%    | 70.6% | 62.4% | 37.6% | 29.4% |
| i.       | 63.3%    | 65.9% | 60.5% | 39.5% | 34.1% |
| ii.      | 71.9%    | 82.5% | 66.4% | 33.6% | 17.5% |
| iii.     | 62.1%    | 60.9% | 73.4% | 26.6% | 39.1% |
| iv.      | 61.4%    | 71.3% | 59.4% | 40.7% | 28.7% |

tumor at surgery in 101 patients with ovarian cancer. They were able to predict the risk of disease progression within 12 months (11). The authors found that along with other features, homogeneity was associated with residual tumor at surgery. In our study, entropy, correlation, difference entropy, and sum entropy radiomics features were positively associated with stromal proportion while MR signal intensity, energy, homogeneity, auto correlation, difference variance, and sum radiomics features were positively associated with tumor proportion. The energy feature is high when GLCM contains only a few high-intensity pixels, which means that the MR image grayscale is uniform. In our study, the energy feature is high when the tumor region is mostly composed of high density of tumor cells. In contrast, entropy reflects the complexity of the MR image and its value is high when the MR image is disorderly. In this work, high entropy values matched with areas encircling tumor cells, mostly composed of stroma. Thus, we found that areas of tumor cells appeared uniform on MR images, while areas of stroma appeared heterogeneous. Those radiomics features associated with stroma heterogeneity may be explained by stroma neoangiogenesis known to be disorderly and the consistence of the stroma itself. Indeed, it mostly consists of heterogeneous cell types and a mixture of amorphous components. Various cell types are found in the stroma of HGSOC, including immune and inflammatory cells, endothelial cells, adipocytes, and the "cancer-associated fibroblasts".

In our study, we built a model allowing to compute a visualize map of estimated stromal and tumor regions from MR images. This model was trained on 29,473 pixels with SI and 13 texture feature values as inputs and binary stromal proportion, extracted from histopathologic data, as targets. The 4 generated predicted segmentation maps were compared to the actual segmentation maps measured from histology and found a good accuracy (61.4 to 71.9). Relative low accuracy values can be explained by the presence of histological structures not taken into account in the model training, such as glandular lumens resulting in signal loss (**Figures 2**.iii, **Figure 6**.iii). These values can also be explained by the difference of slice thickness between MR and histological images, with a factor of 45%. Future studies will validate the algorithm with new samples prior to and after chemotherapy to assess tumor response and will try to translate it to clinical MRI.

Finally, predicted segmentation maps allowed us to assess the stroma–tumor ratio by *ex vivo* imaging. Multiple studies have demonstrated the importance of stroma in ovarian tumorigenesis, progression, and reduced overall survival (15–17). Our interesting results linking radiomics stroma heterogeneity and pathology may explain some of the radiomics associations found by Lu et al., where their radiomics model was correlated with fibronectin (18). Future ongoing work will try to translate this *ex vivo* findings to *in vivo* real-time evaluation from 1.5- or 3-T scanners used in clinical routine.

Our study has several limitations that can be attributed to its exploratory nature in which the feasibility of a new method was evaluated. The total sample size was rather small, limiting the statistical power. Moreover, from the 9 studied specimens, only four were exploited for the MRM and histology pixel-wise correlation, reducing this sample size. From the five excluded samples, three were removed because of the impossibility to precisely co-register WHSI and MRM images. One of the main elements that could explain this impossibility concerns fatty tissue. Indeed, intracellular fat droplets get dissolved during the preparation of HES slides. While this fatty tissue participates in maintaining the shape of the tumor, its dissolution can lead to important deformation of the shape in 3D. However, performing a pixel-wise correlation allowed us to compensate this weakness by increasing the number of information.

Only one pathologist and radiologist reviewed the image. Regarding radiomics analysis, we limited the number of features extracted to relatively basic texture features giving the small number of samples. Finally, we only evaluate the stroma tumor globally. As the stromal component contains a mixture of cells of different origins, the exact elements in the stroma measured by radiomics remain unclear. A study to associate radiomics features with each component in stroma including fibroblast activation, immune cell infiltration, and extracellular matrix density is necessary to better understand the link between radiomics and stromal proportion.

## 5 CONCLUSION

In conclusion, MRM can be optimized to achieve high-resolution images of HGSOC with images obtained within 3 h of resection. This technique offers the possibility of providing valuable information to surgeons in the intraoperative setting. Furthermore, MRM with radiomics analysis allowed us to associate radiomics features to tumor and stromal proportion.

## 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 local ethics committee. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

Conceptualization, MT, YL, MC, CG-B, and SN. Data curation, MT, LK, CG-B, and SN. Formal analysis, MT, LK, MC, P-EC, CG-B, and SN. Funding acquisition, MT and SN. Investigation, MT, YL, MC, OS, and SN. Methodology, MT, MC, P-EC, MC-O, C-GB, and SN. Project administration, MT and SN. Resources, MT, LK, OS, P-EC, CG-B, and SN. Software, MT, LK, MC, P-EC, and SN. Supervision, MT, MC-O, ES, and SN. Validation, MT, MC-O, ES, and SN. Visualization, MT, MC-O, ES, and SN. Writing—original draft, MT, YL, ES, and SN. Writing—review and editing, MT, YL, LK, OS, P-EC, MC-O, ES, and SN. All authors contributed to the article and approved the submitted version.

## FUNDING

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2021.771848/full#supplementary-material

**Supplementary Figure 1 |** H&E stained histological images (left) with corresponding high-resolution MR images (right), for 5 of the 9 resected peritoneal implants (i. to v.).

**Supplementary Figure 2 |** Box plots showing feature values between pixels labeled as stromal proportion greater than 50% and as less than 50%, measured from the four available HGSOC peritoneal implants (total of 196073 classified pixels). Wilcoxon test, *p*-values <0.0001, for the 4 presented features.

# REFERENCES

1. Torre LA, Trabert B, DeSantis CE, Miller KD, Samimi G, Runowicz CD, et al. Ovarian Cancer Statistics, 2018. *CA: A Cancer J Clin* (2018) 68(4):284–96. doi: 10.3322/caac.21456

2. Lee J-M, Minasian L, Kohn EC. New Strategies in Ovarian Cancer Treatment. *Cancer* (2019) 125(S24):4623–9. doi: 10.1002/cncr.32544

3. van Driel WJ, Koole SN, Sikorska K, Schagen van Leeuwen JH, Schreuder HWR, Hermans RHM, et al. Hyperthermic Intraperitoneal Chemotherapy in Ovarian Cancer. *N Engl J Med* (2018) 378(3):230–40. doi: 10.1056/NEJMoa1708618

4. Jacob F, Marchetti RL, Kind AB, Russell K, Schoetzau A, Heinzelmann-Schwarz VA. High-Grade Serous Peritoneal Cancer Follows a High Stromal Response Signature and Shows Worse Outcome Than Ovarian Cancer. *Mol Oncol* (2021) 91–103. doi: 10.1002/1878-0261.12811

5. Dai G, Liang K, Xiao Z, Yang Q, Yang S-G. A Meta-Analysis on the Diagnostic Value of Diffusion-Weighted Imaging on Ovarian Cancer. *J Buon* (2019) 2333–40.

6. Deen SS, Priest AN, McLean MA, Gill AB, Brodie C, Crawford R, et al. Diffusion Kurtosis MRI as a Predictive Biomarker of Response to Neoadjuvant Chemotherapy in High Grade Serous Ovarian Cancer. *Sci Rep* (2019) 9(1):10742. doi: 10.1038/s41598-019-47195-4

7. Rizzo S, De Piano F, Buscarino V, Pagan E, Bagnardi V, Zanagnolo V, et al. Pre-Operative Evaluation of Epithelial Ovarian Cancer Patients: Role of Whole Body Diffusion Weighted Imaging MR and CT Scans in the Selection of Patients Suitable for Primary Debulking Surgery. A Single-Centre Study. *Eur J Radiol* (2020) 123:108786. doi: 10.1016/j.ejrad.2019.108786

8. Kyriazi S, Collins DJ, Messiou C, Pennert K, Davidson RL, Giles SL, et al. Metastatic Ovarian and Primary Peritoneal Cancer: Assessing Chemotherapy Response With Diffusion-Weighted MR Imaging—Value of Histogram Analysis of Apparent Diffusion Coefficients. *Radiology* (2011) 261(1):182–92. doi: 10.1148/radiol.11110577

9. 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

10. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RGPM, Granton P, et al. Radiomics: Extracting More Information From Medical Images Using Advanced Feature Analysis. *Eur J Cancer* (2012) 48(4):441–6. doi: 10.1016/j.ejca.2011.11.036

11. Rizzo S, Botta F, Raimondi S, Origgi D, Buscarino V, Colarieti A, et al. Radiomics of High-Grade Serous Ovarian Cancer: Association Between Quantitative CT Features, Residual Tumour and Disease Progression Within 12 Months. *Eur Radiol* (2018) 28(11):4849–59. doi: 10.1007/s00330-018-5389-z

12. Nougaret S, Tardieu M, Vargas HA, Reinhold C, Vande Perre S, Bonanno N, et al. Ovarian Cancer: An Update on Imaging in the Era of Radiomics. *Diagn Interventional Imaging* (2019) 100(10):647–55. doi: 10.1016/j.diii.2018.11.007

13. Song X, Ren J-L, 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

14. 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

15. Athavale R, Thomakos N, Godfrey K, Kew F, Cross P, Lopes ADB, et al. The Effect of Epithelial and Stromal Tumor Components on FIGO Stages III and IV Ovarian Carcinosarcomas Treated With Primary Surgery and Chemotherapy. *Int J Gynecologic Cancer* (2007) 17(5). doi: 10.1111/j.1525-1438.2007.00919.x

16. Labiche A, Heutte N, Herlin P, Chasle J, Gauduchon P, Elie N. Stromal Compartment as a Survival Prognostic Factor in Advanced Ovarian Carcinoma. *Int J Gynecologic Cancer* (2010) 20(1). doi: 10.1111/IGC.0b013e3181bda1cb

17. Greenaway J, Moorehead R, Shaw P, Petrik J. Epithelial–stromal Interaction Increases Cell Proliferation, Survival and Tumorigenicity in a Mouse Model of Human Epithelial Ovarian Cancer. *Gynecologic Oncol* (2008) 108(2):385–94. doi: 10.1016/j.ygyno.2007.10.035

18. Lu H, Arshad M, Thornton A, Avesani G, Cunnea P, Curry E, et al. A Mathematical-Descriptor of Tumor-Mesoscopic-Structure From Computed-Tomography Images Annotates Prognostic- and Molecular-Phenotypes of Epithelial Ovarian Cancer. *Nat Commun* (2019) 10(1):764. doi: 10.1038/s41467-019-08718-9

19. Dashevsky BZ, D'Alfonso T, Sutton EJ, Giambrone A, Aronowitz E, Morris EA, et al. The Potential of High Resolution Magnetic Resonance Microscopy in the Pathologic Analysis of Resected Breast and Lymph Tissue. *Sci Rep* (2015) 5:17435. doi: 10.1038/srep17435

20. Fan X, Haney CR, Agrawal G, Pelizzari CA, Antic T, Eggener SE, et al. High-Resolution MRI of Excised Human Prostate Specimens Acquired With 9.4T in Detection and Identification of Cancers: Validation of a Technique. *J Magn Reson Imaging* (2011) 34(4):956–61. doi: 10.1002/jmri.22745

21. Durand M, Jain M, Robinson B, Aronowitz E, Douahy YE, Leung R, et al. Magnetic Resonance Microscopy may Enable Distinction Between Normal Histomorphological Features and Prostate Cancer in the Resected Prostate Gland. *BJU Int* (2017) 119(3):414–23. doi: 10.1111/bju.13523

22. Heidkamp J, Hoogenboom M, Kovacs IE, Veltien A, Maat A, Sedelaar JPM, et al. *Ex Vivo* MRI Evaluation of Prostate Cancer: Localization and Margin Status Prediction of Prostate Cancer in Fresh Radical Prostatectomy Specimens. *J Magn Reson Imaging* (2018) 47(2):439–48. doi: 10.1002/jmri.25785

23. Martínez-Bisbal MC, Martínez-Granados B, Rovira V, Celda B, Esteve V. Magnetic Resonance Spectroscopy and Imaging on Fresh Human Brain Tumor Biopsies at Microscopic Resolution. *Anal Bioanal Chem* (2015) 407 (22):6771–80. doi: 10.1007/s00216-015-8847-3

24. Coillot C, Sidiboulenouar R, Nativel E, Zanca M, Alibert E, Cardoso M, et al. Signal Modeling of an MRI Ribbon Solenoid Coil Dedicated to Spinal Cord Injury Investigations. *J Sens Sens Syst* (2016) 5:137–45. doi: 10.5194/jsss-5-137-2016

25. Mayerhoefer ME, Szomolanyi P, Jirak D, Berg A, Materka A, Dirisamer A, et al. Effects of Magnetic Resonance Image Interpolation on the Results of Texture-Based Pattern Classification: A Phantom Study. *Invest Radiol* (2009) 44(7):405–11. doi: 10.1097/RLI.0b013e3181a50a66

26. Zwanenburg A, Leger S, Vallières M, Löck S. Initiative for the IBS. Image Biomarker Standardisation Initiative. *arXiv* (2016) 161207003. doi: 10.1148/radiol.2020191145

27. Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin J-C, Pujol S, et al. 3d Slicer as an Image Computing Platform for the Quantitative Imaging Network. *Magn Reson Imaging* (2012) 30(9):1323–41. doi: 10.1016/j.mri.2012.05.001

28. Haralick RM, Shanmugam K, Dinstein I. Textural Features for Image Classification. *IEEE Trans Syst Man Cybern* (1973) SMC-3(6):610–21. doi: 10.1109/TSMC.1973.4309314

29. Bankhead P, Loughrey MB, Fernández JA, Dombrowski Y, McArt DG, Dunne PD, et al. QuPath: Open Source Software for Digital Pathology Image Analysis. *Sci Rep* (2017) 7(1):16878. doi: 10.1038/s41598-017-17204-5

30. Gibson E, Crukley C, Gaed M, Gómez JA, Moussa M, Chin JL, et al. Registration of Prostate Histology Images to *Ex Vivo* MR Images *via* Strand-Shaped Fiducials. *J Magn Reson Imaging* (2012) 36(6):1402–12. doi: 10.1002/jmri.23767

31. Lopater J, Colin P, Beuvon F, Sibony M, Dalimier E, Cornud F, et al. Real-Time Cancer Diagnosis During Prostate Biopsy: *Ex Vivo* Evaluation of Full-Field Optical Coherence Tomography (FFOCT) Imaging on Biopsy Cores. *World J Urol* (2016) 34(2):237–43. doi: 10.1007/s00345-015-1620-6

# Exploring the Interobserver Agreement in Computer-Aided Radiologic Tumor Measurement and Evaluation of Tumor Response

Hongsen Li[1†], Jiaying Shen[1†], Jiawei Shou[1], Weidong Han[1], Liu Gong[1], Yiming Xu[2], Peng Chen[3], Kaixin Wang[3], Shuangfeng Zhang[3], Chao Sun[3], Jie Zhang[3], Zhongfeng Niu[4], Hongming Pan[1*], Wenli Cai[5*] and Yong Fang[1*]

[1] Department of Medical Oncology, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou, China, [2] Quantilogic Healthcare Zhejiang Co. Ltd, Hangzhou, China, [3] School of Medical Imaging, Shanghai University of Medicine and Health Sciences, Shanghai, China, [4] Department of Radiology, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou, China, [5] Department of Radiology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, United States

The accurate, objective, and reproducible evaluation of tumor response to therapy is indispensable in clinical trials. This study aimed at investigating the reliability and reproducibility of a computer-aided contouring (CAC) tool in tumor measurements and its impact on evaluation of tumor response in terms of RECIST 1.1 criteria. A total of 200 cancer patients were retrospectively collected in this study, which were randomly divided into two sets of 100 patients for experiential learning and testing. A total of 744 target lesions were identified by a senior radiologist in distinctive body parts, of which 278 lesions were in data set 1 (learning set) and 466 lesions were in data set 2 (testing set). Five image analysts were respectively instructed to measure lesion diameter using manual and CAC tools in data set 1 and subsequently tested in data set 2. The interobserver variability of tumor measurements was validated by using the coefficient of variance (CV), the Pearson correlation coefficient (PCC), and the interobserver correlation coefficient (ICC). We verified that the mean CV of manual measurement remained constant between the learning and testing data sets (0.33 vs. 0.32, $p = 0.490$), whereas it decreased for the CAC measurements after learning (0.24 vs. 0.19, $p < 0.001$). The interobserver measurements with good agreement (CV < 0.20) were 29.9% (manual) vs. 49.0% (CAC) in the learning set ($p < 0.001$) and 30.9% (manual) vs. 64.4% (CAC) in the testing set ($p < 0.001$). The mean PCCs were $0.56 \pm 0.11$ mm (manual) vs. $0.69 \pm 0.10$ mm (CAC) in the learning set ($p = 0.013$) and $0.73 \pm 0.07$ mm (manual) vs. $0.84 \pm 0.03$ mm (CAC) in the testing set ($p < 0.001$). ICCs were 0.633 (manual) vs. 0.698 (CAC) in the learning set ($p < 0.001$) and 0.716 (manual) vs. 0.824 (CAC) in the testing set ($p < 0.001$). The Fleiss' kappa analysis revealed that the overall agreement was 58.7% (manual) vs. 58.9% (CAC) in the learning set and 62.9% (manual) vs. 74.5% (CAC) in the testing set. The 80% agreement of tumor

response evaluation was 55.0% (manual) vs. 66.0% in the learning set and 60.6% (manual) vs. 79.7% (CAC) in the testing set. In conclusion, CAC can reduce the interobserver variability of radiological tumor measurements and thus improve the agreement of imaging evaluation of tumor response.

Keywords: tumor measurements, evaluation agreement, response evaluation criteria in solid tumors (RECIST), measurement variability, treatment assessment

# INTRODUCTION

Radiological imaging examination plays an important role in monitoring of tumor progression or evaluation of tumor response to treatment in oncological clinical trials and clinical care (1–4). Cancer patients may undergo longitudinally radiological imaging examinations, such as CT (computed tomography), MRI (magnetic resonance imaging), and/or PET (positron emission tomography), to quantify tumor burden for assessment of tumor response to treatment (4). The criteria for evaluation of tumor response vary in terms of tumor types and treatment methods. In 1981, the World Health Organization (WHO) published the first criteria for solid tumor response evaluation, which adopted bidimensional measurement as the tumor imaging biomarkers for quantifying tumor burden (5). The Response Evaluation Criteria in Solid Tumors (RECIST) published in 2000 and its revised version (RECIST 1.1) in 2009 adopted the unidimensional instead of bidimensional measurement as the tumor imaging biomarkers to quantifying tumor burden (6). Nowadays, oncology clinical trials increasingly rely on image-based surrogate endpoints. RECIST represents the internationally recognized evaluation criteria for solid tumors (5, 6). With the advent of oncologic therapies (targeted therapy, immunotherapy, etc.), the radiological response assessment criteria are also evolving, for instance, the modified RECIST (mRECIST) for evaluating the response of primary hepatocellular carcinoma (HCC) to targeted therapy (5, 7) and the modified RECIST for assessment of cancer immunotherapy (iRECIST) (8). Moreover, the FDA published guidelines in 1994, with updates in 2004 (9) and 2019 (10) for standardizing the radiological assessment of tumor responses as a primary endpoint in clinical trials.

Reliable evaluation of tumor response depends on two aspects: the correct selection of target lesions and the accurate and reproducible measurement of tumor burden. Although the concept concerning the measurement of the maximal tumor diameter is undoubtedly simple and convenient for physicians, has a long history of clinical applications, and is familiar to management agencies (FDA), the methodology of measuring lesions is poorly defined in RECIST or WHO. The subjective linear measurement has been criticized for its low reproducibility and high inter- and intra-observer variabilities of the tumor response assessment. Several studies observed that the intraobserver variability was among 6% to 14%, and the interobserver variability was approximately 10% to 25% (11). These measurement variabilities may lead to a misinterpretation of tumor response, in particular the large interobserver

variability. Some studies found that the misclassification of tumor responses caused by the interobserver measurement variabilities was as high as 43% (WHO) and 30% (RECIST) (12). A meta-analysis summarized the RECIST-based observer variability of manual measurements on CT images (13): relative measurement differences ranged from 17.8% to 16.1% for the same observers (5 studies, 648 lesions measured), −22.1% to 25.4% between two observers (8 studies, 1,878 lesions measured), and −31.3% to 30.3% among multiple observers (3 studies, 575 lesions measured). It has been reported that even for expert radiologists, there was considerable variability in interpretation of lesion boundaries, in particular for irregular lesions, with the interobserver variability accounting for 40% of a lesion size (14), which may inevitably result in significant difference in tumor size measurements (12). Because this interobserver variability in tumor measurements may lead to a misclassification of tumor growth rate or response even for the same selected target lesions, the methodology for measuring lesions needs to be improved (15).

The aims of this study therefore were (1) to develop a computer-aided contouring (CAC) tool for interactive measurement of maximum tumor diameter required by RECIST or WHO criteria, (2) to validate the interobserver agreement of the CAC tool in tumor diameter measurements, and (3) to assess the consistency of the CAC tool in the evaluation of tumor response in terms of RECIST 1.1 criteria.

# MATERIALS AND METHODS

The institutional ethics committee has approved this retrospective study, in which informed consent was waived, but patient confidentiality was protected. The study design is illustrated in **Figure 1**.

## Patient Cohort

This study was approved by the ethics committee of the Sir Run Run Shaw Hospital, Zhejiang University School of Medicine (Scientific Research No.2020617-33). The clinical information used in this study was previously archived data, files, records, pathological specimens, and diagnostic specimens, and was approved of exempting informed consent by Institutional Review Board (IRB). Two hundred patients with pathological confirmed cancers between January 1, 2015 and December 31, 2019 enrolled in several Phase I or II oncology clinical trials evaluating systemic therapies in lung, liver, and colorectal cancer patients were retrospectively selected in a random sample from

**FIGURE 1** | The flow chart of the study. Data collection and the design of the study.
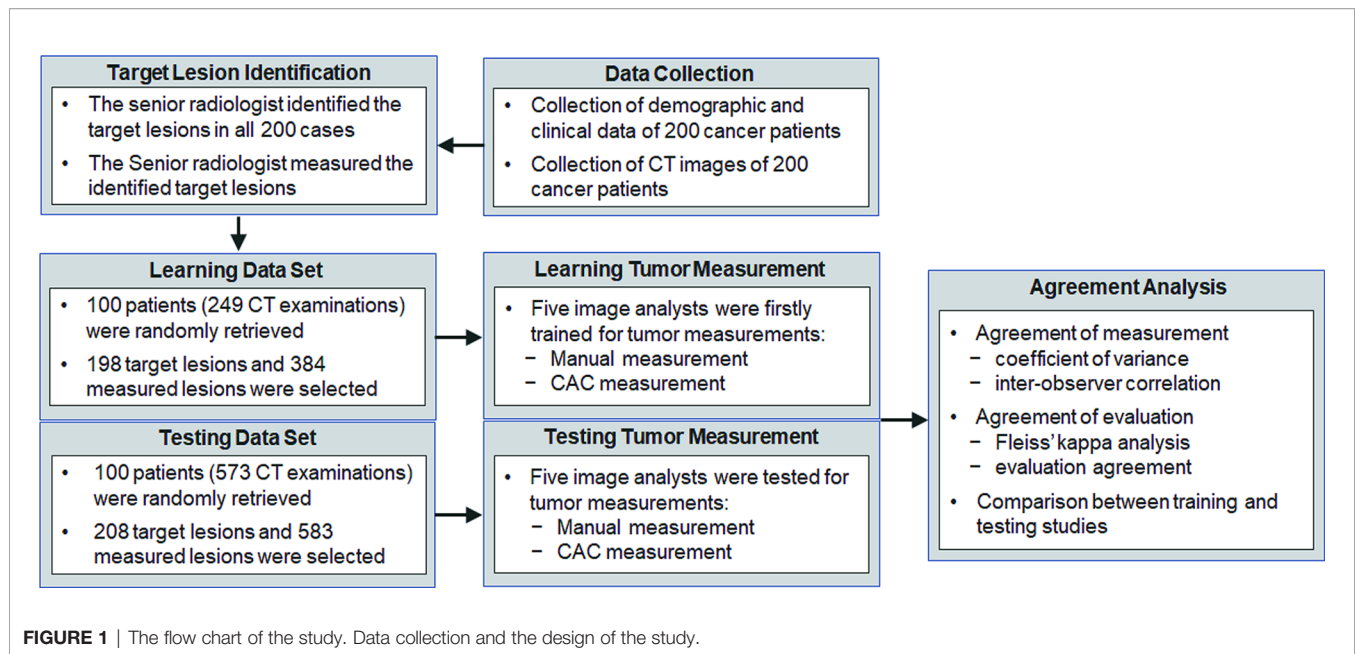
the hospital medical records system. Inclusion criteria for this study were as follows: (1) an adult patient over 18 years old; (2) pathological confirmed malignant solid tumors in head and neck, lung, liver, abdomen, lymph nodes, and other body parts; (3) received anti-cancer treatment such as radiotherapy, chemotherapy, targeted therapy, etc.; (4) performed contrast-enhanced CT examinations before and after treatment; and (5) had at least one measurable lesion ≥10 mm in terms of RECIST 1.1 criteria.

The exclusion criteria were as follows: (1) patients had no measurable lesions defined by RECIST criteria; (2) patients had different image modalities other than CT before or after treatment; (3) the imaging examination did not cover the entire lesion; and (4) incomplete clinical records, for instance, the date of treatment and missing CT images.

## Imaging Examination

Contrast-enhanced CT examinations were acquired on multi-detector CT scanners (GE, Siemens) with a tube voltage of 120 kVp, an automatic tube current, a slice collimation of 0.6 to 1.5 mm, and standard reconstruction kernel and slice thickness ranging from 1.25 to 5 mm defined in the clinical trial protocols.

## Lesion Selection and Lesion Size Measurement

According to the requirements of target lesion selection in RECIST 1.1 (6), a senior radiologist selected up to 5 measurable lesions per patient with a maximum of 2 lesions per organ as the target lesions. Although RECIST 1.1 defines that the target lesions at baseline should be ≥10 mm in the longest diameter for extranodal disease and ≥15 mm in the short-axis diameter for nodal disease, some target lesions may be smaller than 10 mm on follow-up examinations due to treatment effects.

For the purpose of treatment response evaluation, those <10 mm lesions in follow-up examinations were also included in the study to assess the variability. Each selected lesion was independently measured using a manual tool and a CAC tool by five image analysts, respectively. Furthermore, to reduce the variability caused by subjective selection of target lesions, we recruited a senior radiologist (ZN) to identify these target lesions using an arrow to enable the recognition by the image analysts. The image analysts were informed that the arrow was randomly marked on one of the slices of the lesion and thus did not indicate the slice with the longest lesion diameter. The image analysts were required to determine the slice for longest lesion diameter based on their own judgements after examining all the slices of the lesion.

In terms of the RECIST 1.1, the longest diameter of a target lesion was measured on the transverse (axial) plane in CT by manual method first. Each image analyst examined each target lesion selected by the senior radiologist and determined the transverse (axial) plane with the longest diameter based on visual assessment. Standard window/level (HU) settings were applied in terms of organs or body parts, for instance, lung (1500/−500), liver (310/80), abdomen (400/60), and neck (250/30). The longest diameter of each lesion was measured by placing two endpoints of the diameter on the edge of the lesion without crossing normal tissues.

Following the manual measurement, each image analyst was required to measure the lesion by using the CAC tool, which detected the optimal boundaries in terms of the initial contours given by the image analyst. An image analyst reviewed the CAC-generated contour and corrected it if it was unsatisfactory. The longest diameter (for extranodal disease) or the longest perpendicular diameter (nodal disease) of the lesion was automatically estimated by the CAC tool in either 2D/3D

mode: if only one slice of a lesion was contoured, the longest diameter of the contour was directly calculated; if multiple slices of a lesion were contoured, the contour with the maximum cut-plane area was first selected and the longest diameter of that contour was calculated. This diameter served as the CAC measurement of a lesion.

## Computer-Aided Contouring Method

CAC is a computer-aided contouring toolkit that we developed based on the optimal path search using the dynamic-programming techniques in graph theory (15). A transversal image can be represented by a 2D weighted bi-directed graph, in which one node corresponds to a pixel in the image. Each node (pixel) has 8 connecting links (edges) to its neighboring nodes (pixels), and each node and link has an associated cost. The local boundary of a region-of-interest (ROI) on a 2D transversal image is defined as the optimal path with the minimum cost between two corresponding nodes in the graph, as illustrated in **Figure 2**.

Dynamic-programming theory indicates that the optimal path between node $u$ and $v$ is either a direct link between $u$ and $v$, or going from $u$ to some node $w$ and then directly from node $w$ to $v$, which is also named Dijkstra's observation (16). Based on this theory, Dijkstra's algorithm first computes the minimum cost path from the seed node $u$ to every node in the entire graph. This set of minimum paths can be represented as a tree structure. Once this tree is established, the optimal path from any node to the seed node $u$ can be extracted in the tree in real time by traversing from root $u$ to any node in the tree. For a $512 \times 512$ image, it takes less than 1 s to establish the tree structure from the seed node to the entire image. As long as the tree structure is established, the traversing from the root to a node in the tree structure is very efficient, less than 1 ms. This ensures the real-time interactivity of the CAC algorithm.

Because an optimal path corresponds to a segment of ROI boundaries, pixels or links between neighboring pixels that exhibit strong edge features are made to have low local costs. Edge features such as Laplacian zero-crossing $S_{ZX}(v)$, gradient

magnitude $S_g(v)$, and gradient direction $S_d(u,v)$ (17) are incorporated into the computation of local cost. The cost going from node (pixel) $u$ to node (pixel) $v$ is a weighted sum of each corresponding local cost defined as $c(u, v) = \omega_{ZX} \cdot S_{ZX}(v) + \omega_g \cdot S_g(v) + \omega_d \cdot S_d(u,v)$, where $\omega_{ZX}$, $\omega_g$, and $\omega_d$, are constants to weight features. In **Figure 3**, we demonstrate these cost functions.

In CAC, we assume that the boundary of a ROI is located within the neighborhood of the mouse moving trajectory. This neighborhood is defined as a band on the transversal images centered by the mouse moving trajectory, as shown in **Figure 4A**. The width of the band can be adjusted in terms of the size of the segmented ROIs. In our study, we set the band width to be 20 pixels or 10 pixels on both sides of the initial mouse moving trajectory. The local boundary of a ROI will be searched within this local band. The use of a local band significantly reduces the searching space of optimal path. More importantly, it may improve the stability of CAC detected boundary by converging to the optimal tumor boundary and ignoring nodes or links outside the band, as demonstrated in **Figures 4C1–C4**.

To reduce the influence of inaccurate positions of initial seed points (such as $u$, $v$ given by a user) and ensure that the selected seeds are at or near the boundary of the ROI, we applied a seed selection process *via* a two-pass optimal path searching scheme: after the first optimal path searching using initial given seeds, the points with the lowest cost in every segments of the optimal path between two initial seed points are selected as the new seed points for the second searching of the optimal path. This seed selection process significantly reduces the inaccuracy and the interobserver variability of seed positioning. Because the path is piece-wise optimal (i.e., optimal between seed points), the detected boundary using selected seed points is stable and reproducible, as demonstrated in **Figure 4** wherein the different trajectories in **Figures 4C1–C4** result in the exact same contour shown in **Figure 4B**.

CAC provides the ability to detect (searching and snapping) the ROI boundaries between key slices that have user-defined ROI contours. In general, the user contoured a lesion at the first
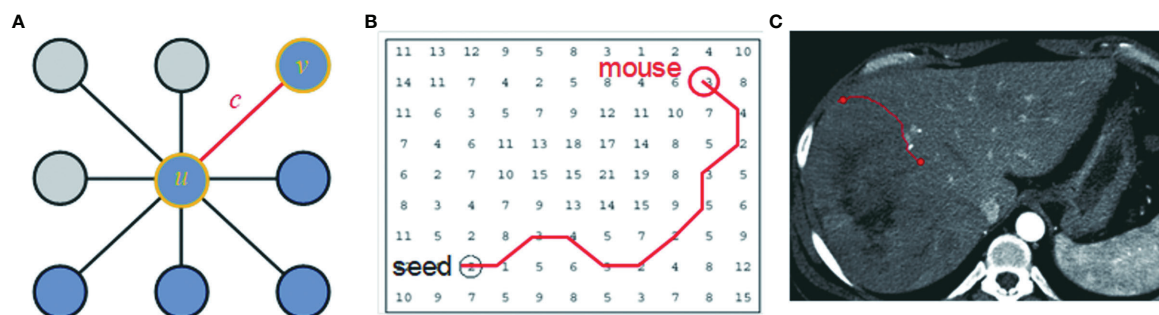


**FIGURE 2** | The local boundary is defined as the optimal path with the minimum cost between two corresponding nodes in the graph. **(A)** An image can be represented as a 2D weighted bi-directed graph, in which one node corresponds to a pixel in the image. Each node (pixel) has 8 links (edge) connecting to its neighborhood pixels. Both node ($u$ or $v$) and link **(C)** have a related cost. **(B)** A contour on a 2D transversal image is defined as the optimal path with the minimum cost between two corresponding nodes in the graph, which can be searched by dynamic-programming methods such as Dijkstra's algorithm. **(C)** A contour in the image corresponds to the optimal path in **(B)**.
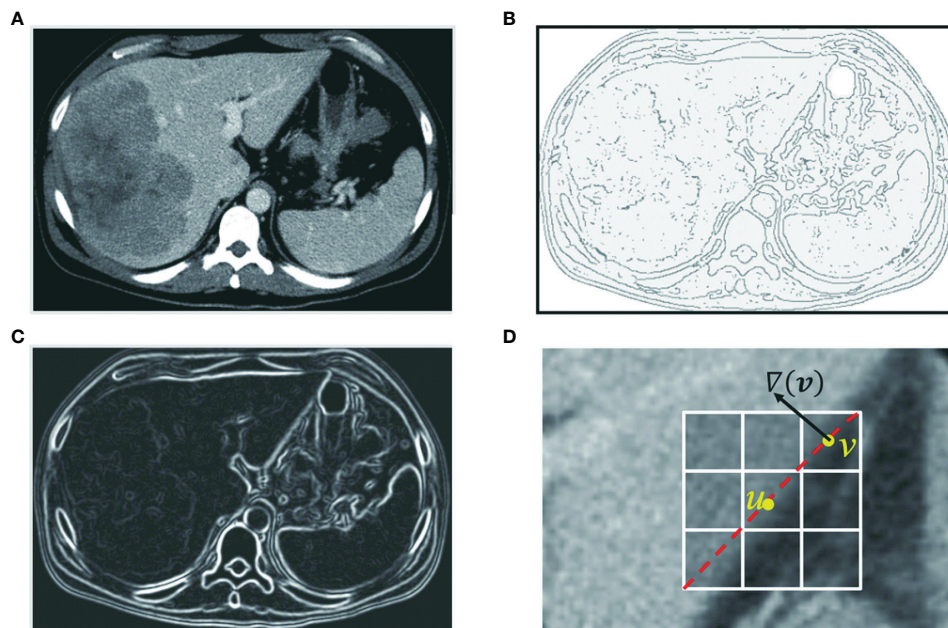
**FIGURE 3** | Edge features such as zero-crossing and gradient are incorporated into the computation of local cost. **(A)** Original axial image. **(B)** Zero-crossing: points with local maximal gradient magnitude corresponds to a local boundary point with lower cost. **(C)** Gradient magnitude: a point with a large gradient magnitude tends to indicate a local boundary and thus a lower cost. **(D)** Gradient direction: a local boundary is perpendicular to the gradient direction. Cost is low if the direction of a link (uv) is perpendicular to the gradient direction [$\nabla(v)$] of the pixel.
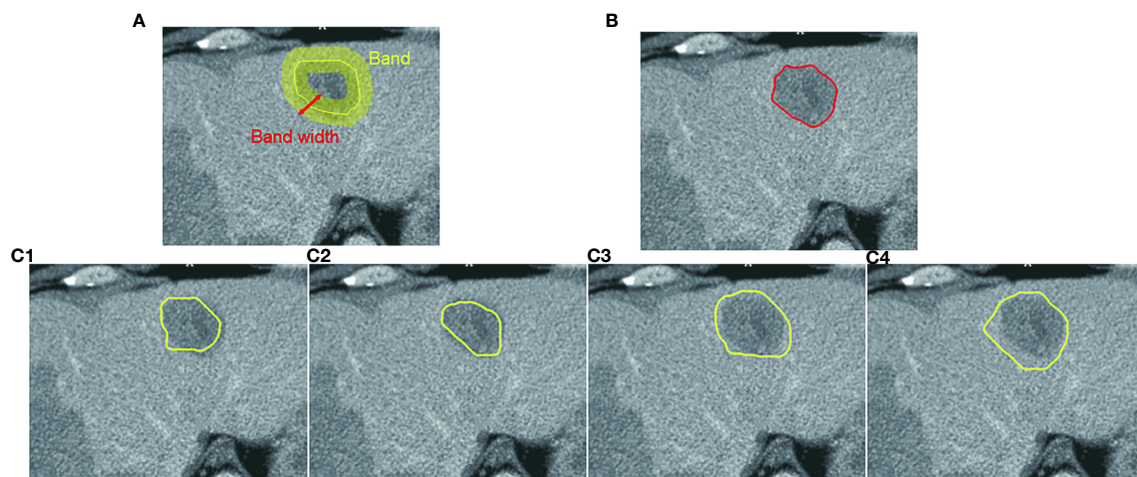


**FIGURE 4** | Computer-aided contouring (CAC) tool. **(A)** Tumor boundary is searched within a band centered by the mouse moving trajectory. **(B)** The resulting boundary is the optimal path searched in terms of the minimum cost of local edges. **(C1–C4)** Different mouse moving trajectories result in the same tumor boundary shown in **(B)** after applying the CAC tool.

and the last slice, and for the middle key slices of a lesion, the user may contour at every 3–5 slice intervals along the scanning direction. The exact number of slices between key slices depended on the complexity (irregularity, infiltration, and inhomogeneity) of the lesion. The contours between these key slices were detected in an automated manner by using the

scheme of between-slice contour detection in CAC, as illustrated in **Figures 5A, B**. This automated between-slice detection scheme consists of three steps below:

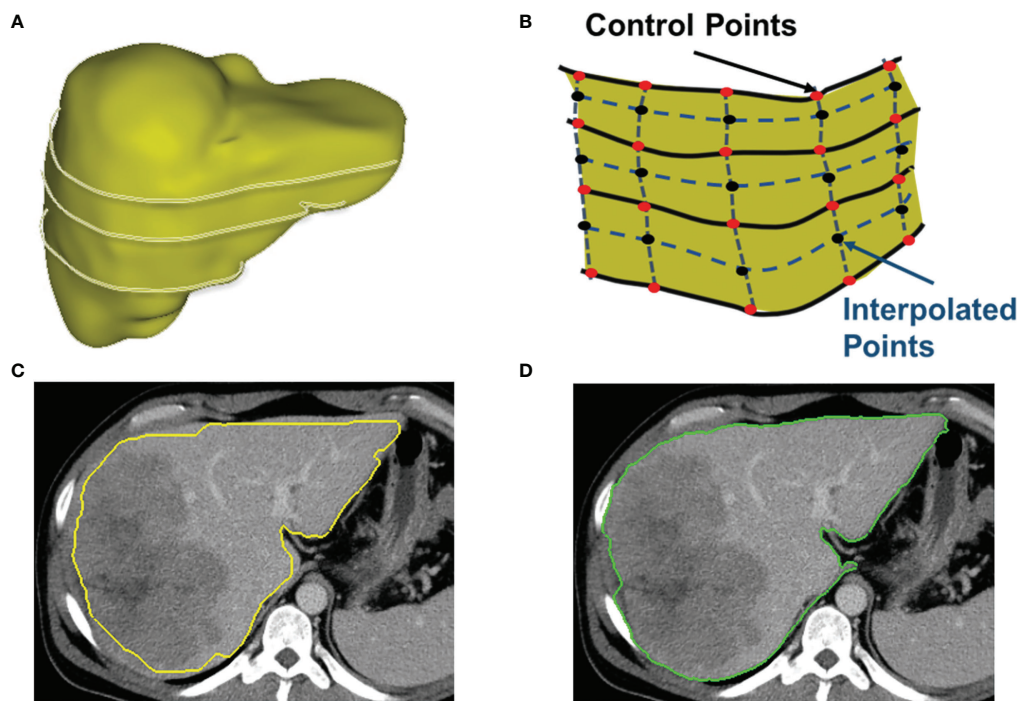(1) Interpolate seed points on each of the between-slices using a pair of neighboring key slices;

**FIGURE 5** | Detection of between-slices contours by interpolation and optimization. **(A)** A 3D shape of a liver with contours on several non-adjacent key slices. CAC toolkit can detect the contours between key slices. **(B)** The interpolation of control points on slices between the key slices. **(C)** The initial contour estimated by using interpolated control points. **(D)** The resulting contour generated by the two-pass optimal path searching scheme using the estimated contour in **(C)**.

(2) Estimate initial contours by connecting these interpolated seed points; and

(3) Search the local boundary using the two-pass optimal path searching scheme (described above) within the band of the estimated contour.

**Figure 5C** shows an initial estimated contour created by interpolation step, and **Figure 5D** is the resulting contour generated by the two-pass optimal path searching scheme in the band of the initial contour in **Figure 5C**. Because our two-pass optimal path searching scheme can find accurate and stable ROI boundaries in an automated manner, the between-slices contour detection can significantly reduce the user time and effort in interactive contouring while still achieving an accurate segmentation. This volumetric contouring ensures that CAC can estimate the longest diameter of the contour with the maximum cut-plane area of a lesion, which is called the 3D mode for longest diameter estimation.

Based on this theory, we developed a CAC toolkit for interactive segmentation of tumors and organs, which includes (1) real-time searching of the local boundary on a 2D transversal image, (2) automated detection of between-slice contours for volumetric contouring, and (3) automated calculation of the longest diameter (RECIST criteria) or the product of perpendicular longest diameters (WHO criteria) of a tumor on either one transversal image (2D mode) or multiple transversal images (3D mode).

## Validation Scheme

These 200 patients were randomly divided into two data sets of 100 patients for experiential learning set (data set 1) and testing (data set 2) of tumor measurement using manual and CAC tools. Five image analysts who had no prior knowledge of tumor measurements and RECIST criteria were first instructed how to measure a lesion diameter in terms of the RECIST 1.1 criteria in the learning set, and then tested in the second data set by using manual and CAC tools on the V3D platform (Quantilogic Healthcare). Tumor responses were evaluated in terms of RECSIT 1.1 criteria by using the resulting measurements of the manual and CAC tools, respectively. The reliability and reproducibility of tumor measurement and response evaluation were compared between manual and CAC tools.

## Data Analysis

To evaluate the agreement of the measurement, we calculated the coefficient of variance (CV) of five measurements of each tumor, the correlation between any two pairs of analysts [Pearson correlation coefficient (PCC)], and the interobserver correlation coefficient [interclass correlation coefficient (ICC)] of five image analysts. To evaluate the agreement of tumor response evaluation, we calculated the Fleiss' kappa coefficients.

Statistical analysis was conducted by R software, version 3.3.2 (https://www.R-project.org/), and SPSS Statistics for Windows, version 17. 0 (SPSS Inc., Chicago, Ill., USA), of which correlations were calculated by SPSS and other statistics were

calculated by R. Quantitative variables were shown as mean ± SD. Statistical group comparisons of data were analyzed by Wilcoxon rank-sum (continuous variables) test and $\chi^2$ (categorical/dichotomous variables) test. A Student's $t$-test was used for the continuous variable. A $p$-value less than 0.05 was considered statistically significant.

## RESULTS

### Clinical Data

Two hundred patients were randomly separated into two sets of 100 patients: data set 1 (71 males; 29 females; mean age 58.7 years; range 34 to 73 years) and data set 2 (67 males; 33 females; mean age 57.1 years; range 21 to 74 years). A total of 249 CT examinations were retrieved from PACS in the first data set, and 573 CT examinations were retrieved in the second data set.

A total of 384 and 583 reported tumors were identified by a senior radiologist (ZN) in data sets 1 and 2, respectively. Because of one or more missing measurements in the resulting measurements by the five image analysts, 76 and 94 tumors were excluded in data sets 1 and 2, respectively. This resulted in 308 lesions in data set 1 and 489 lesions in data set 2 in the measurement study. In the tumor response evaluation study, we excluded patients caused by one time point and missing target lesions in the follow-up examinations. This resulted in 80 evaluations (48 patients) and 231 evaluations (89 patients) in data sets 1 and 2, respectively. **Figure 6** shows the patient selection and distribution in data sets 1 and 2 for the learning and testing studies.

Overall, the lesion diameters ranged from 10.0 to 105.0 mm (mean: 30.7 ± 19.0 mm, median 25.5 mm) in data set 1 and 5.0 to 88.0 mm (mean: 22.5 ± 15.1 mm, median 18.0 mm) in data set 2. In the tumor response evaluation study, the average number of follow-up examinations was 1.6 and 2.6 in the learning and

testing study, and the average number of lesions per patient was 1.94 and 1.92 in two sets, respectively.

The details of the patient characteristics and lesion statistics are summarized in **Table 1**. The example case displayed in **Figure 7** demonstrates a significant reduction of interobserver variability by the CAC tool.

### Agreement of Measurement and Analysis of Coefficient of Variance

**Figures 8A–C** show that the CV in the learning set was reduced by CAC compared to manual measurements, and the percentage of measurements with a CV of <0.2 were 29.9% (manual) vs. 49.0% (CAC), which was statistically significant ($p < 0.001$). **Figures 8D–F** show the CV in the testing set after experiential learning. It indicates that while the mean CV of manual measurement remained constant between the first and second data sets (0.33 vs. 0.32, $p = 0.490$), it decreased for the CAC measurements after learning (0.24 vs. 0.19, $p < 0.001$) as the image analysts became familiar with the CAC tool. In addition, we measured lesions from different body parts in the study to demonstrate the generalizability of the CAC tool. Overall, the CAC tool outperformed manual measurement at all six body parts: CV between manual and CAC tools were 0.27 vs. 0.20 ($p = 0.004$) (lung), 0.23 vs. 0.17 ($p = 0.046$) (liver), 0.30 vs. 0.25 ($p = 0.240$), (abdomen), 0.30 vs. 0.20 ($p < 0.001$) (lymph), and 0.37 vs. 0.27 ($p = 0.003$) (other), respectively. This demonstrated that the CAC tool is a general-purpose imaging tool for tumor measurements, indicating the great potentials for the clinical adoption of the CAC tool.

### Analysis of Interobserver Correlation

**Table 2** presents the PC matrix among five image analysts in data sets 1 and 2, respectively. After experiential learning, we observed that the agreement of each analyst with four other image analysts was improved. However, the CAC tool still outperformed the
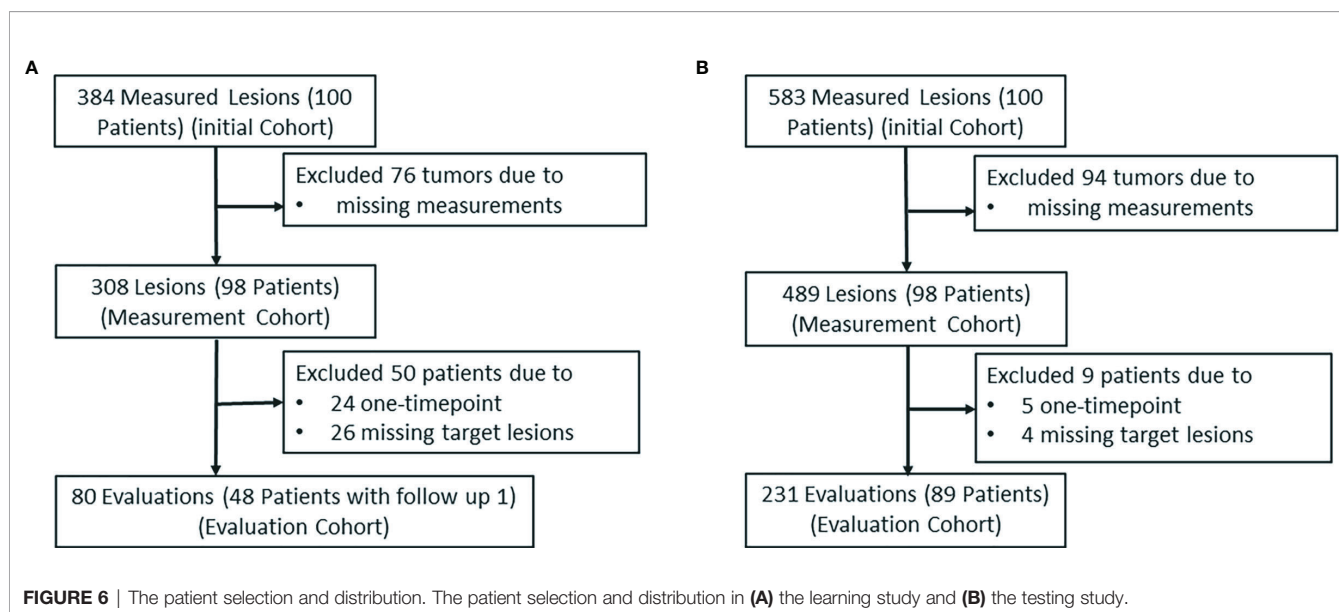


**FIGURE 6** | The patient selection and distribution. The patient selection and distribution in **(A)** the learning study and **(B)** the testing study.

**TABLE 1 |** The demographic and clinical characteristics of the experiential learning data set (data set 1) and testing data set (data set 2).

| Group | Data set 1 (n = 100) | Data set 2 (n = 100) | p-value |
|---|---|---|---|
| Sex | | | |
| Female | 29 | 33 | 0.647 |
| Male | 71 | 67 | |
| Enrolled age | | | |
| Mean | 58.7 | 57.1 | 0.268 |
| Range (Min–Max) | 34–73 | 21–74 | |
| Treatment | | | |
| Chemotherapy only | 8 | 12 | 0.505 |
| TKIs based regimes | 33 | 36 | |
| Antibodies based regimes | 59 | 52 | |
| Target lesions | | | |
| Total | 198 | 208 | |
| Lung | 66 | 78 | 0.052 |
| Lymph nodes | 60 | 58 | |
| Liver | 38 | 20 | |
| Body wall | 10 | 16 | |
| Soft tissue | 7 | 7 | |
| Other | 17 | 29 | |
| Adrenal gland | 1 | 11 | / |
| Peritoneum | 5 | 5 | |
| Brain | 3 | 2 | |
| Breast | 3 | 2 | |
| Kidney | 1 | 2 | |
| Annex | 0 | 2 | |
| Pancreas | 2 | 0 | |
| Spleen | 1 | 1 | |
| Bone | 0 | 1 | |
| Diaphragm | 0 | 1 | |
| Limbs | 0 | 1 | |
| Parotid gland | 0 | 1 | |
| Pharynx | 1 | 0 | |

manual tool. In addition, the two-way mixed interobserver correlation analysis revealed the ICC values of 0.633 (manual) vs. 0.698 (CAC) in data set 1, and 0.716 (manual) vs. 0.824 (CAC) in data set 2. After experiential learning, the CAC tool achieved "excellent" level of agreement (ICC > 0.75).

## Agreement of Tumor Response Evaluation

Tumor response was evaluated in four categories: complete response (CR), partial response (PR), stable disease (SD), and progressive disease (PD). As shown in **Table 3**, the total numbers of RECIST evaluations were 80 and 231 in the two data sets, respectively. The Fleiss' kappa analysis revealed that the percentage overall agreement ($K$) was 58.7% (manual) vs. 58.9% (CAC) in the first learning set, whereas it was 62.9% (manual) vs. 74.5% (CAC) in the second testing set. The agreement increased approximately 4.2% in manual and 15.6% in CAC. The CAC tool approached "excellent" agreement ($K > 0.75$), whereas the manual method remained in "good" agreement.

Among five tumor response evaluations by five image analysts for each patient, we calculated the "excellent" patient-level agreement, which was defined as the same response category as assessed by more than four image analysts assessed the same response categories, i.e., more than 80% evaluation results were the same. The manual measurements achieved a constant level of agreement: 55.0% (learning) and 60.6% (testing) ($p < 0.001$), whereas CAC measurements improved from 66.0% (learning) to 79.7% (testing) ($p < 0.001$). It indicates that when image analysts became familiar with the CAC tool after experiential learning, the difference of agreement between two tools on tumor response evaluation becomes more significant, increasing from 11.0% [learning: 66.0% (CAC) vs. 55.0% (manual), $p < 0.001$] to 19.1% [testing: 79.7% (CAC) vs. 60.6% (manual), $p < 0.001$].

## DISCUSSION

This study shows the improvements in unidimensional tumor measurements that can be gained by utilizing a CAC tool
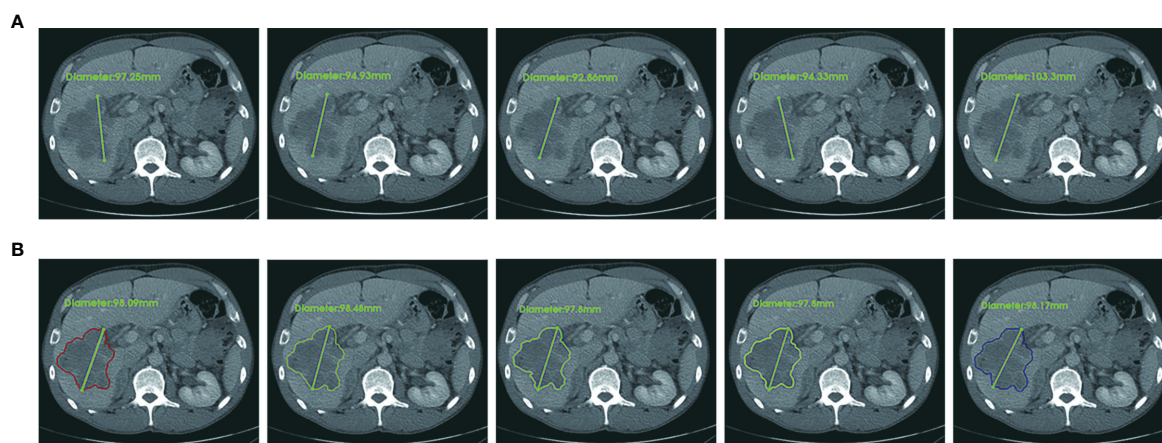


**FIGURE 7 |** An example case of manual and CAC measurements. In terms of RECIST 1.1 criteria, the longest diameter of a target lesion is measured on the image acquisition plane (axial plane in CT). **(A)** Upper row: five manual measurements showed the mean value 96.5 ± 4.1 mm, range 92.7 mm to 103.3 mm, CV (coefficient of variance) 4.3%. **(B)** Lower row: five CAC measurements showed the mean value 98.1 ± 0.3 mm, range 97.8 mm to 98.1 mm, CV 0.3%. The agreement of CAC measurements was substantially better than that of manual measurements.

**A**

|  | Mean STD (MM) | Median STD (MM) | Mean CV (%) | Median CV (%) |
|---|---|---|---|---|
| Manual | 9.96±8.24 | 7.51 | 0.33±0.18 | 0.31 |
| CAC | 8.47±8.82 | 5.06 | 0.24±0.17 | 0.21 |
| Difference | (-1.49) | (-2.45) | (-0.09) | (-0.10) |

**B**

| CV Threshold | Manual | CAC |
|---|---|---|
| < 0.05 | 2.3% | 11.0% |
| < 0.10 | 9.4% | 27.9% |
| < 0.15 | 18.2% | 39.9% |
| < 0.20 | 29.9% | 49.0% |

**C**

**D**

|  | Mean STD (MM) | Median STD (MM) | Mean CV (%) | Median CV (%) |
|---|---|---|---|---|
| Manual | 7.06±6.82 | 4.75 | 0.32±0.21 | 0.28 |
| CAC | 4.62±5.46 | 2.70 | 0.19±0.17 | 0.14 |
| Difference | (-2.44) | (-2.05) | (-0.13) | (-0.14) |

**E**

| CV Threshold | Manual | CAC |
|---|---|---|
| < 0.05 | 1.4% | 14.1% |
| < 0.10 | 9.8% | 36.8% |
| < 0.15 | 19.6% | 52.8% |
| < 0.20 | 30.9% | 64.4% |

**F**

**FIGURE 8** | Analysis of coefficients of variance (CV) in the experiential learning and testing study. Analysis of coefficients of variance (CV) in the experiential learning study (data set 1, **A–C**) and in the testing set (data set 2 **D–F**). The standard deviation (STD) and CV of the manual and CAC measurements (mean and median values given in a and d, respectively) and the CV distribution (**B, E**, respectively) are shown before and after learning. The boxplots in **(C, F)** show the fractions of measurements with good agreement (<0.2). In data set 1, this was 29.9% vs. 49.0% for manual and CAC measurement, respectively, and in data set 2, this was 30.9% vs. 64.4% for manual and CAC measurements, respectively.

compared to manual methods and its impact on response evaluation. In order to prevent bias from prior knowledge or practice, we recruited five image analysts who were unfamiliar with tumor measurement and response assessment, and were instructed to learn and subsequently tested independently the measurement of longest lesion diameter in terms of RECIST criteria using manual and CAC tools. For this purpose, we divided the 200 patients into two sets of 100 patents for learning and testing.

This study demonstrated that the interobserver agreement of the manual measurement had approximately 1/3 of the cases, whereas the CAC tool achieved twice the better performance, which was also indicated by the "excellent" level of ICC of the CAC tool. The novelty of the CAC method is its reproducibility and consistency of segmentation, which can significantly reduce the interobserver variability in tumor measurement, thereby ensuring the quality and repeatability of tumor response evaluation among different radiologists and institutions. Manual

contouring tools are easy to use but very time-consuming, whereas semi-automated interactive tools (such as snake and speedline) are less labor-intensive but highly unstable. We thus developed the CAC tool in the context of tumor response assessment for the purpose of an efficient and consistent tumor measurement. Compared with other open-source segmentation tools in ITK Snap (www.itksnap.org) and Seg3D (www.sci.utah.edu/cibc-software/seg3d.html), the CAC tool has certain advantages on accuracy and reproducibility. A comparison of image contouring tools with the CAC tool is given in **Table S1**.

This study had several limitations. The first limitation was the single CT modality that we used in the validation study. We selected CT examinations for the experiential learning and testing in order to demonstrate the feasibility of the CAC tool. However, the CAC tool is not limited to CT modality. We will investigate the reliability and reproducibility of the CAC tool in MRI and PET/CT examinations. Another limitation is that our study used single-center data from one institute. We plan to collect multi-center

**TABLE 2 |** Pearson correlation matrix among five image analysts in data sets 1 and 2.

| | Data Set 1 (Learning) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | (A) Manual measurement | | | | | (B) CAC measurement | | | | |
| | Analyst 1 | Analyst 2 | Analyst 3 | Analyst 4 | Analyst 5 | Analyst 1 | Analyst 2 | Analyst 3 | Analyst 4 | Analyst 5 |
| Analyst 1 | 1 | 0.385 | 0.418 | 0.367 | 0.401 | 1 | 0.538 | 0.533 | 0.507 | 0.565 |
| Analyst 2 | 0.385 | 1 | 0.418 | 0.367 | 0.401 | 0.538 | 1 | 0.82 | 0.702 | 0.895 |
| Analyst 3 | 0.418 | 0.808 | 1 | 0.736 | 0.865 | 0.533 | 0.82 | 1 | 0.728 | 0.909 |
| Analyst 4 | 0.367 | 0.667 | 0.736 | 1 | 0.673 | 0.506 | 0.702 | 0.728 | 1 | 0.752 |
| Analyst 5 | 0.401 | 0.871 | 0.865 | 0.673 | 1 | 0.565 | 0.895 | 0.909 | 0.752 | 1 |
| Correlation with Others | 0.39 | 0.68 | 0.61 | 0.54 | 0.59 | 0.54 | 0.74 | 0.75 | 0.67 | 0.78 |
| | Data Set 2 (Testing) | | | | | | | | | |
| | (C) Manual measurement | | | | | (D) CAC measurement | | | | |
| | Analyst 1 | Analyst 2 | Analyst 3 | Analyst 4 | Analyst 5 | Analyst 1 | Analyst 2 | Analyst 3 | Analyst 4 | Analyst 5 |
| Analyst 1 | 1 | 0.546 | 0.672 | 0.605 | 0.639 | 1 | 0.760 | 0.815 | 0.815 | 0.845 |
| Analyst 2 | 0.546 | 1 | 0.823 | 0.741 | 0.796 | 0.760 | 1 | 0.850 | 0.795 | 0.828 |
| Analyst 3 | 0.672 | 0.823 | 1 | 0.796 | 0.846 | 0.815 | 0.850 | 1 | 0.830 | 0.883 |
| Analyst 4 | 0.605 | 0.741 | 0.796 | 1 | 0.814 | 0.815 | 0.795 | 0.830 | 1 | 0.931 |
| Analyst 5 | 0.639 | 0.796 | 0.846 | 0.814 | 1 | 0.845 | 0.828 | 0.883 | 0.931 | 1 |
| Correlation with Others | 0.62 | 0.73 | 0.78 | 0.74 | 0.77 | 0.81 | 0.81 | 0.84 | 0.84 | 0.87 |

*(A) The manual measurement: The mean Pearson correlation coefficient (PCC) was 0.56 ± 0.11, range 0.39 to 0.69. (B) The CAC measurement: The mean PCC was 0.69 ± 0.10, range 0.54 to 0.78 in data set 1. The two-way mixed interobserver correlation coefficient (ICC) was 0.633 (manual) vs. 0.698 (CAC). (C) The manual measurement: The mean Pearson correlation coefficient (PCC) was 0.73 ± 0.07, range 0.62 to 0.78. (D) The CAC measurement: The mean PCC was 0.84 ± 0.03, range 0.81 to 0.87 in data set 2. The two-way mixed interobserver correlation coefficient (ICC) was 0.716 (manual) vs. 0.824 (CAC). CAC tool was in "excellent" agreement (>0.75), whereas manual method remains in "good" agreement.*

**TABLE 3 |** Agreement of RECIST 1.1 evaluation.

| | (A) Data set 1 (Learning) | | | | (B) Data set 2 (Testing) | | | |
|---|---|---|---|---|---|---|---|---|
| | Total # | Votes% | Cases # | Cases% | Total # | Votes% | Cases # | Cases% |
| Manual | 80 | 80% | 44 | 55.00% | 231 | 80% | 140 | 60.60% |
| | | 100% | 22 | 27.50% | | 100% | 77 | 33.30% |
| CAC | 80 | 80% | 52 | 66.00% | 231 | 80% | 184 | 79.70% |
| | | 100% | 17 | 21.30% | | 100% | 113 | 48.90% |

*(A) Agreement in data set 1. (B) Agreement in data set 2.*

cases to validate the reproducibility of the CAC tool for tumor measurements and evaluation. The CAC tool needs multi-center and prospective data for further validation and improvements. Although the clinical significance of the CAC tool warrant validation by larger multi-center studies, it may provide a reliable and reproducible solution for radiological tumor measurement, which may further affect the imaging endpoints in tumor response evaluation.

In conclusion, our study demonstrated that the computer-aided contouring method can significantly improve the agreement of radiologic tumor measurements, reduce the interobserver variability of tumor measurement, and thus improve the agreement of tumor response evaluation in oncology clinical trials and clinical care.

# DATA AVAILABILITY STATEMENT

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

# ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics committee of the Sir Run Run Shaw Hospital, Zhejiang University School of Medicine. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements. Written informed consent was not obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

# AUTHOR CONTRIBUTIONS

Conceptualization: HP, WC, YF, and HL. Investigation and data collection: HL, JWS, WH, and LG. Analysis: JYS, YX, and ZN. Image analysist: PC, KW, SZ, CS, and JZ. Wrote original draft: JYS and WC. Review, editing, and final approval: WC, YF, and HP. Supervision: YF.

All authors contributed to the article and approved the submitted version.

## FUNDING

## REFERENCES

1. Hersberger KE, Mendiratta-Lala R, Fischer RK, Kaza IR, Francis MS, Olszewski JF, et al. Quantitative Imaging Assessment for Clinical Trials in Oncology. *J Natl Compr Canc Netw* (2019) 17(12):1505–11. doi: 10.6004/jnccn.2019.7331

2. Cheson BD, Pfistner ME, Juweid RD, Gascoyne L, Specht SJ, Horning B, et al. Revised Response Criteria for Malignant Lymphoma. *J Clin Oncol* (2007) 25 (5):579–86. doi: 10.1200/JCO.2006.09.2403

3. Miller AB, Hoogstraten B, Staquet M, Winkler A. Reporting Results of Cancer Treatment. *Cancer* (1981) 47(1):207–14. doi: 10.1002/1097-0142(19810101) 47:1<207::AID-CNCR2820470134>3.0.CO;2-6

4. Rosenkrantz AB, Mendiratta-Lala M, Bartholmai BJ, Ganeshan D, Abramson RG, Burton KR, et al. Clinical Utility of Quantitative Imaging. *Acad Radiol* (2015) 22(1):33–49. doi: 10.1016/j.acra.2014.08.011

5. Tirkes T, Hollar MA, Tann M, Kohli MD, Akisik F, Sandrasegaran K. Response Criteria in Oncologic Imaging: Review of Traditional and New Criteria. *Radiographics* (2013) 33(5):1323–41. doi: 10.1148/rg.335125214

6. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New Response Evaluation Criteria in Solid Tumours: Revised RECIST Guideline (Version 1.1). *Eur J Cancer* (2009) 45(2):228–47. doi: 10.1016/j.ejca.2008.10.026

7. 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

8. Wolchok JD, Hoos A, O–Day S, Weber JS, Hamid O, Lebbe C, et al. Guidelines for the Evaluation of Immune Therapy Activity in Solid Tumors: Immune-Related Response Criteria. *Clin Cancer Res* (2009) 15 (23):7412–20. doi: 10.1158/1078-0432.CCR-09-1624

9. U.S. Food and Drug Administration. *Guidance Document. Developing Medical Imaging Drug and Biological Products Part 3: Design, Analysis, and Interpretation of Clinical Studies*. Available at: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/developing-medical-imaging-drug-and-biological-products-part-3-design-analysis-and-interpretation (Accessed January 20, 2022).

10. Food and Drug Administration. *Guidance Document. Clinical Trial Imaging Endpoint Process Standards Guidance for Industry*. Available at: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/clinical-trial-imaging-endpoint-process-standards-guidance-industry (Accessed January 20, 2022).

11. Zhao B, Tan Y, Bell DJ, Marley SE, Guo P, Mann H, et al. Exploring 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 Radiol* (2013) 82(6):959–68. doi: 10.1016/j.ejrad.2013.02.018

12. Erasmus JJ, Gladish GW, Broemeling L, Sabloff BS, Truong MT, Herbst RS, et al. Interobserver and Intraobserver Variability in Measurement of non-Small-Cell Carcinoma Lung Lesions: Implications for Assessment of Tumor Response. *J Clin Oncol* (2003) 21(13):2574–82. doi: 10.1200/JCO.2003.01.144

13. Yoon SH, Kim KW, Goo JM, Kim DW, 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

14. Meyer CR, Johnson TD, McLennan G, Aberle DR, Kazerooni EA, Macmahon H, et al. Evaluation of Lung MDCT Nodule Annotation Across Radiologists and Methods. *Acad Radiol* (2006) 13(10):1254–65. doi: 10.1016/j.acra.2006.07.012

15. Cai WL, Hong GB. Quantitative Image Analysis for Evaluation of Tumor Response in Clinical Oncology. *Chronic Dis Transl Med* (2018) 4(1):18–28. doi: 10.1016/j.cdtm.2018.01.002

16. Tacher V, Lin M, Duran R, Yarmohammadi H, Lee H, Chapiro J, et al. Comparison of Existing Response Criteria in Patients With Hepatocellular Carcinoma Treated With Transarterial Chemoembolization Using a 3D Quantitative Approach. *Radiology* (2016) 278(1):275–84. doi: 10.1148/radiol.2015142951

17. Sonka M, Hlavác V, Boyle R. *Image Processing, Analysis and Machine Vision (3. Ed.)*. Toronto:Thompson Learning (2008) DBLP.

## ACKNOWLEDGMENTS

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2021.691638/full#supplementary-material

# Advantages
# of publishing
# in Frontiers

**OPEN ACCESS**
Articles are free to read
for greatest visibility
and readership

**FAST PUBLICATION**
Around 90 days
from submission
to decision

**HIGH QUALITY PEER-REVIEW**
Rigorous, collaborative,
and constructive
peer-review

**TRANSPARENT PEER-REVIEW**
Editors and reviewers
acknowledged by name
on published articles

**Frontiers**
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

**Visit us:** www.frontiersin.org
**Contact us:** frontiersin.org/about/contact

**REPRODUCIBILITY OF
RESEARCH**
Support open data
and methods to enhance
research reproducibility

**DIGITAL PUBLISHING**
Articles designed
for optimal readership
across devices

**FOLLOW US**
@frontiersin

**IMPACT METRICS**
Advanced article metrics
track visibility across
digital media

**EXTENSIVE PROMOTION**
Marketing
and promotion
of impactful research

**LOOP RESEARCH NETWORK**
Our network
increases your
article's readership