MAGNETIC RESONANCE IMAGING FOR RADIATION THERAPY

EDITED BY: Ning Wen, Yue Cao and Jing Cai

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: researchtopics@frontiersin.org

MAGNETIC RESONANCE IMAGING FOR RADIATION THERAPY

Topic Editors:
Ning Wen, Henry Ford Health System, United States
Yue Cao, University of Michigan, United States
Jing Cai, Hong Kong Polytechnic University, Hong Kong

Table of Contents

Editorial: Magnetic Resonance Imaging for Radiation Therapy

Ning Wen[1]*, Yue Cao[2] and Jing Cai[3]

[1] Department of Radiation Oncology, Henry Ford Health System, Detroit, MI, United States, [2] Department of Radiation Oncology, University of Michigan, Ann Arbor, MI, United States, [3] Department of Health Technology and Informatics, Hong Kong Polytechnic University, Kowloon, Hong Kong

Editorial on the Research Topic

Magnetic Resonance Imaging for Radiation Therapy

Since the introduction of magnetic resonance imaging (MRI) to radiation therapy (RT), it has increasingly been adopted in RT treatment planning for target and organ-at-risk (OAR) definition due to superior soft tissue contrasts. Recently, the roles of MRI in RT have been advanced to tumor delineation in a multiparametric format, MRI-based treatment planning and dose calculation, MRI-guided treatment delivery, and outcome assessment using quantitative imaging metrics or radiomic features. These advancements are due to the development of dedicated MRI simulators, integration of MRI scanners with radiation treatment platforms, as well as technical developments and applications of 4D-MRI, tumor tracking, adaptive planning, and treatment response evaluation. The advancement of machine learning and artificial intelligence also brings about tremendous opportunities to transform the application of MRI in RT.

This collection includes 16 articles that cover the following themes:

- Target volume delineation and treatment planning workflow using MRI as the primary imaging modality

Multiparametric MRI (mp-MRI), a combination of morphologic and functional imaging modalities, has shown the potential to increase the accuracy of tumor detection, localization, and characterization of cancer aggression. Integration of mp-MRI techniques into RT offers enormous opportunities to individualize RT adaptation based upon the individual patient's response to treatment.

MR spectroscopy imaging (MRSI) can describe the metabolism of different tissues. However, the spatial resolution is limited due to the very low concentration of the metabolites in tissues. Iqbal et al. developed a densely connected U-Net to create super resolution spectroscopic images by training the T1 weighted images (T1WI) and the low-resolution ^{1}H MR spectroscopic images together. They showed that the ^{1}H spectra were maintained on retrospective *in vivo* data.

Lee et al. performed a volumetric and voxel-wise analysis of the dominant intraprostatic lesions (DIL) defined from T2-weighted imaging (T2WI), diffusion weighted imaging (DWI), and dynamic contrast enhanced (DCE) imaging respectively. The correlation was further classified according to tumor location and Gleason grade group. The data suggested that constructing a Boolean sum volume that incorporated T2WI and apparent diffusion coefficient (ADC) maps were reasonable for delineating the DIL on mp-MRI. The value of adding information provided by K^{trans} maps remains investigational due to the repeatability and consistency of DCE scans. The interobserver variability also indicated the need to develop a consensus guideline on DIL delineation using mp-MRI.

● Techniques to generate synthetic CT from MRI and the clinical implementation of MRI only workflow

Substantial interest has developed around generating synthetic CT (sCT) from MRI in order to use MRI as the only, or primary, imaging modality in the RT workflow. Different methods have been introduced to create sCT images using bulk density, atlas-based, or voxel-based approaches. Deep machine learning algorithms such as a U-Net or Generative Adversarial Network can learn image features among different imaging modalities and have great potentials to generate highly accurate synthetic images.

Choi et al. used a bulk anatomical density approach to develop a method for patient-specific quality assurance for MRI-only prostate RT. The three-class model (bone, muscle, and fat) provided accurate dose calculations for verifying sCT for clinical use in MRI-only workflows. The model has currently been implemented as a quality assurance method in a multi-center trial of prostate stereotactic RT that includes an MRI-only study.

Gupta et al. used a 3-channel U-Net trained on aligned MRI and CT pairs in sagittal planes to generate sCT images. The three channels represented Hounsfield Unit (HU) ranges of voxels containing air, soft tissue, and bone, respectively. The improved soft tissue contrast of sCT was proved with low mean absolute error difference between sCT and actual CT. The improved image quality was also beneficial for the online image registration with cone beam CT.

Wang et al. generated sCT from T2WI of nasopharyngeal carcinoma patients using a 2D U-net algorithm. The deep U-net with 23 convolutional layers was used to generate sCT. The soft tissue, nasal bone, bone marrow, and the interface between bones and soft tissues were carefully evaluated.

Greer et al. described a multi-center study for the implementation of an MRI-only prostate workflow. A sCT was created using an atlas-based method from whole pelvic T2WI scans with an isotropic 1.6 mm voxel. A CT scan was acquired subsequent to MRI only plan approval for patient specific quality assurance. The 3D gamma was calculated to evaluate the dose difference between sCT and actual CT, and gold fiducial marker positions were used to evaluate the image registration accuracy between sCT and actual CT. All 25 patients recruited were treated with MRI only workflow.

Mittauer et al. developed an MRI-guided online adaptive radiotherapy (MRgoART) procedure for palliative care in RT. The electron density information was incorporated with either a bulk density override or deformable image registration of diagnostic CT to the MRI. The plan quality and treatment delivery efficiency were superior than the conventional method. Excellent clinical outcomes were observed and were in line with historical and sampled controls.

● Quantitative analysis of morphological and functional MRI and their applications in treatment response assessment

Quantitative MRI can reflect tissue characteristics. Imaging biomarkers from functional MRI can have prognostic and predictive values for progression free survival, overall survival, and distant metastases etc. Radiomic features, which are defined as the post-processing for extraction of textural information from medical images, can provide tremendous information to analyze and characterize the properties of tumor tissues and their physiological and pathological stages.

In this collection, Cao et al. analyzed MRI-derived gross tumor volume, blood volume, and ADC from pre-treatment and mid-treatment, as well as pre-treatment FDG PET metrics for locally advanced head and neck cancer (HNC) treated with chemoradiation. The mean ADC values from pre-RT and its change rate mid-treatment were significant higher and lower in p16– than p16+ locally advanced HNC tumors, respectively. These biomarkers had predictive values and compared favorably with FDG-PET imaging markers.

van Schie et al. analyzed T2 and ADC changes during treatment and compared patients with and without hormonal therapy, as the hypoFLAME trial patients received ultra-hypofractionated prostate radiotherapy with an integrated boost to the tumor in 5 weekly fractions. Significant ADC changes were observed in the tumor in patients without hormonal therapy. Such early response measured with quantitative MRI holds the potential to predict clinical outcome and guide treatment adaptation.

Bagher-Ebadian et al. extracted discriminant radiomic features in the real radiomics-feature space and the latent-variable space from mp-MRI for prostate cancer. These features were used to construct an artificial neural network to classify the DIL from normal prostatic tissues.

Li Z. et al. analyzed pre-treatment T1WI, T2WI, and DWI for esophageal squamous cell carcinoma patients undergoing concurrent chemoradiotherapy and identified the ADC texture features that can be used to predict the overall survival.

Yu et al. also analyzed pre-treatment T1WI, T2WI to identify tumoral radiomic features that were used to predict patient eligibility for adaptive radiotherapy in advanced nasopharyngeal carcinoma (NPC) patients.

Considering post-treatment changes are often highly heterogeneous, including cellular tumor, fat, necrosis, and cystic tissue compartments, evaluation of the tumors defined using pre-treatment images could be limited to predict treatment response. Blackledge et al. studied 8 commonly used supervised machine-learning algorithms for tissue classification of mp-MRI of soft tissue sarcoma to quantify post-RT changes. Five out of eight algorithms achieved similar performance. Of the five methods, the Naïve-Bayes classifier was chosen for further investigation due to its relatively short training and prediction times.

● Development of MR-Linac

Recent commercial developments of MRI-guided RT platforms provide great opportunities for direct imaging guidance, tumor/OAR tracking during RT, and treatment adaptation. Two systems have received CE and/or FDA 510 k clearance so far: Unity (Elekta, Sweden) and MRIdian system (Viewray, USA). The Australian MRI-linac system is at the research prototype stage and has an inline orientation, with radiation beam parallel to the main magnetic field.

Such inline design can help minimize magnetic field influence on dose deposition. Jelen et al. developed methods to quantify dosimetric characteristics of the Australian MRI-linac system.

● Review papers

There are two review papers in this collection. As MRI-guided RT, including adaptive RT, have advanced in the field, the community needs to develop protocols on how to make clinical decisions with funneling MRIgRT data. Kiser et al. discussed the challenges of interpretability and reproducibility of MRI data, the complexity of a variety of MR sequences, and the corresponding impacts on RT workflow, such as synthetic CT generation, image fusion, dose calculation, and prognostic values using radiomic features.

Li G. et al. reviewed two 4D-MRI techniques—respiratory-correlated (RC) and time-resolved (TR) 4D-MRI. The RC-4DMRI was reconstructed to provide one-breathing-cycle motion, while the TR-4DMRI provided an adequate spatiotemporal resolution to assess tumor motion and motion variation. Both techniques were also discussed in the context of their clinical applications in radiotherapy.

Benefitting from advanced technologies of synthetic CT techniques and MR Linacs, the MR-solely RT workflow has been rapidly evolving and has been clinically implemented widely. It has potential to improve the therapeutic gains for certain disease sites through dose escalation with better tumor delineation and motion management. Randomized clinical trials have been promoted to investigate the effects of dose escalation on normal tissue toxicity, quality of life, as well as overall survival and local control for prostate cancer, locally advanced pancreatic cancer, etc. As MRI is playing an increasingly essential role in RT, opportunities arise to incorporate functional imaging into RT workflow. Considering the response of the ADC maps to radiation dose and relatively robust protocol for DWI acquisition, DWI-derived biomarkers have strong potentials for tumor delineation and response assessment, as evidenced in a series of articles published in this collection.

AUTHOR CONTRIBUTIONS

NW, YC, and JC are the coeditors for this Research Topic.

Volumetric and Voxel-Wise Analysis of Dominant Intraprostatic Lesions on Multiparametric MRI

Joon Lee[1], Eric Carver[1], Aharon Feldman[1], Milan V. Pantelic[2], Mohamed Elshaikh[1] and Ning Wen[1]*

[1] Department of Radiation Oncology, Henry Ford Health System, Detroit, MI, United States, [2] Department of Radiology, Henry Ford Health System, Detroit, MI, United States

Introduction: Multiparametric MR imaging (mpMRI) has shown promising results in the diagnosis and localization of prostate cancer. Furthermore, mpMRI may play an important role in identifying the dominant intraprostatic lesion (DIL) for radiotherapy boost. We sought to investigate the level of correlation between dominant tumor foci contoured on various mpMRI sequences.

Methods: mpMRI data from 90 patients with MR-guided biopsy-proven prostate cancer were obtained from the SPIE-AAPM-NCI Prostate MR Classification Challenge. Each case consisted of T2-weighted (T2W), apparent diffusion coefficient (ADC), and K^{trans} images computed from dynamic contrast-enhanced sequences. All image sets were rigidly co-registered, and the dominant tumor foci were identified and contoured for each MRI sequence. Hausdorff distance (HD), mean distance to agreement (MDA), and Dice and Jaccard coefficients were calculated between the contours for each pair of MRI sequences (i.e., T2 vs. ADC, T2 vs. K^{trans}, and ADC vs. K^{trans}). The voxel wise spearman correlation was also obtained between these image pairs.

Results: The DILs were located in the anterior fibromuscular stroma, central zone, peripheral zone, and transition zone in 35.2, 5.6, 32.4, and 25.4% of patients, respectively. Gleason grade groups 1–5 represented 29.6, 40.8, 15.5, and 14.1% of the study population, respectively (with group grades 4 and 5 analyzed together). The mean contour volumes for the T2W images, and the ADC and K^{trans} maps were 2.14 $\pm$ 2.1, 2.22 $\pm$ 2.2, and 1.84 $\pm$ 1.5 mL, respectively. K^{trans} values were indistinguishable between cancerous regions and the rest of prostatic regions for 19 patients. The Dice coefficient and Jaccard index were 0.74 $\pm$ 0.13, 0.60 $\pm$ 0.15 for T2W-ADC and 0.61 $\pm$ 0.16, 0.46 $\pm$ 0.16 for T2W-K^{trans}. The voxel-based Spearman correlations were 0.20 $\pm$ 0.20 for T2W-ADC and 0.13 $\pm$ 0.25 for T2W-K^{trans}.

Conclusions: The DIL contoured on T2W images had a high level of agreement with those contoured on ADC maps, but there was little to no quantitative correlation of these results with tumor location and Gleason grade group. Technical hurdles are yet to be solved for precision radiotherapy to target the DILs based on physiological imaging. A Boolean sum volume (BSV) incorporating all available MR sequences may be reasonable in delineating the DIL boost volume.

Keywords: prostate cancer, multiparametric MR, dominant intraprostatic lesions, tumor delineation, radiotherapy

INTRODUCTION

Prostate cancer is the most common malignancy of men in the U.S., with an annual incidence of 161,360 cases resulting in 26,730 deaths (1). Most patients are diagnosed with disease localized to the prostate, for which radiation therapy is an important curative treatment modality.

In the modern era of dose-escalated radiation therapy, the entire prostate gland is treated to the same dose of radiation irrespective of the biopsy-proven region of disease. Multiple randomized studies have demonstrated that dose-escalation improves biochemical progression-free survival (2, 3). It has also been reported that local recurrences are dose-dependent and most frequently occur at the site of the dominant intraprostatic lesion (DIL) (4, 5)—defined as the most prominent cancerous lesion within the prostate which also exhibits the most aggressive clinical behavior. Numerous studies have suggested that the addition of a boost to the DIL is safe and efficacious without increased acute or late toxicity (6–14).

Multiparametric magnetic resonance imaging (mpMRI) is rapidly becoming the standard diagnostic imaging modality for prostate cancer. mpMRI can be defined as any functional form of MR imaging which supplements standard anatomical T1- (T1W) and T2-weighted (T2W) MR sequences. Namely, this includes diffusion-weighted imaging (DWI), which measures the Brownian motion of water molecules in tissue; dynamic contrast-enhanced (DCE) sequences, which assess tumor angiogenesis and detect microvascular vessel wall permeability; and MR spectroscopy (MRS), which analyzes the chemical composition of prostatic tissue, and compares it to that of cancerous tissue.

mpMRI has shown potential to increase the accuracy of tumor detection, localization, and characterization of prostate cancer (15–18). It has been demonstrated to have a negative predictive value of up to 95% for clinically significant prostate cancer (defined as the presence of Gleason pattern 4 or greater) (19, 20). Whole amount histopathology has been used a gold standard reference to evaluate DIL detection and localization accuracy using mpMRI (21).

The correlation of tumor volume defined by pathology and mpMRI was also investigated and it showed strong dependence on both imaging techniques and specimen processing workflow (22). There are still lack of studies to investigate whether a specific MR sequence is optimal or if a combination of MR sequences is mandatory in accurately delineating the DIL for radiotherapy planning. In this study, we performed volumetric and voxel-wise analyses of tumor foci delineated in three MR sequences and report the level of concordance between them. Furthermore, we quantitatively correlated these results with tumor location and Gleason grade group.

MATERIALS AND METHODS

Robust mpMRI data from 90 patients with MRI-guided biopsy-proven prostate cancer were obtained from the SPIE-AAPM-NCI Prostate MR Classification Challenge (23, 24). All images were acquired using two different types of Siemens 3-Tesla MR scanners (the Magnetom Trio and Skyra) without an endorectal coil. Each dataset consisted of T2W, ADC, and volume transfer coefficient (K^{trans}) images computed from DCE sequences.

The T2W images were acquired using a turbo spin echo sequence (TE/TR: 5,660/104 ms, Flip Angle: 90° with image resolution of $0.5 \times 0.5 \times 3.0$ mm^3). The DWI was acquired with a single-shot echo planar imaging sequence with diffusion-encoding gradients in three directions (TR/TE: 2,700/63 ms, with image resolution of $2.0 \times 2.0 \times 3.0$ mm^3). The ADC map was calculated from three b-values of 50, 400, and 800 s/mm^2. The DCE series were acquired using a 3-D turbo flash gradient echo sequence (TR/TE: 3.4/1.5 ms, with image resolution of $1.5 \times 1.5 \times 3.0$ mm^3 and a temporal resolution of 3.5 s). The standard Tofts model was used for pharmacokinetic modeling of the contrast concentration curves. An automated reference tissue method was used to estimate the arterial input function (25). The transfer constant (K^{trans}) parametric maps were calculated from the contrast concentration curves.

An experienced radiologist annotated suspicious lesions on each MR modality, and MRI-guided biopsies were performed to confirm the aggressiveness of the disease (i.e., Gleason grade grouping). The tissue specimens were examined by expert pathologists and the results were defined as the ground truth in this study. Both the ADC and K^{trans} image sets were rigidly co-registered and resampled using linear interpolation to match those of the T2W images. For example, resampling transformed the resolution from $2.0 \times 2.0 \times 3.0$ mm (ADC) and $1.5 \times 1.5 \times 3.0$ mm (K^{trans}) to $0.5 \times 0.5 \times 3.0$ mm (T2W). The intraprostatic lesions were then identified and contoured on each MR sequence separately for every patient by a radiation oncologist based on the radiologist's annotation following criteria of hypointense values on the T2W images (window 718, level 360) and ADC maps (window 3,000, level 1,500) and high values on the K^{trans} maps (window 39, level 21). The DIL was separately contoured by a second radiation oncologist for a subset of MR images (19 patients) to assess for interobserver variability. Representative images of an intraprostatic lesion contoured on an ADC map, K^{trans} map and T2W image are shown in **Figure 1**.

The anatomic location of the intraprostatic lesions as well as their corresponding Gleason grade group (1–5) were available for each patient. Due to the small number of data points available, Gleason grade groups 4 and 5 were analyzed together. To evaluate the quantitative correlation between contours on each imaging modality and its statistical dependence on tumor location and Gleason grade group, the 95 percentiles of Hausdorff distance (HD), mean distance to agreement (MDA), Dice coefficient, and Jaccard index were calculated between the contours for each pair of MR sequences (i.e., T2W vs. ADC, T2W vs. K^{trans}, and ADC vs. K^{trans}). These variables are defined in **Table 1**.

For the voxel-wise analysis, a Boolean sum volume (BSV) was defined as a combination of the contours from all three image modalities for each patient. This additional step was performed to ensure that an equal number of representative voxels from each MR sequence were included in the analysis. Fractional ranks were then obtained for each voxel of the BSV and the Spearman correlation was calculated. It is worth noting that the Spearman correlation was selected because a monotonic relationship was assumed between each pair of contours, as opposed to a linear
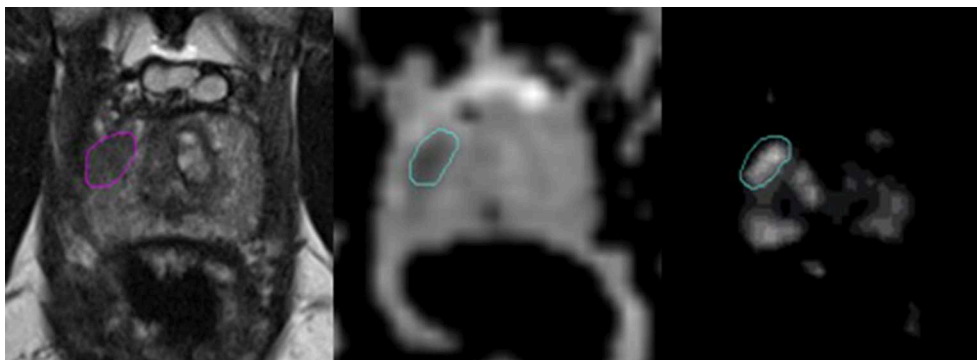
FIGURE 1 | Representative images of a dominant intraprostatic lesion contoured on an T2W image, ADC map, and K^{trans} map (from left to right). The axial T2W images were carefully scrutinized along with those in the coronal, and sagittal planes (not shown) to confirm the presence of a hypodense lesion corresponding to the location of the DIL annotated by an experienced radiologist. In the axial T2W image above, a hypointense lesion is demonstrated in the right peripheral zone. This same area was then assessed for values that were lower and higher than the surrounding normal prostate tissue in the rigidly-registered axial ADC and K^{trans} maps, respectively. Of note, areas of hypervascularity outside of the contoured region in the K^{trans} map shown above were assumed to represent normal vasculature within the central zone of the prostate gland, as a corresponding area suspicious for cancer was not visualized in either the T2W image or the ADC map.

TABLE 1 | Definitions of contour evaluation metrics.

Definitions
Hausdorff distance (HD): The distance from one point in a subset to the closest point in another subset.
$$d_H = \max[\sup \inf d(x, y), \sup \inf d(x, y)]$$
Mean distance to agreement (MDA): The average of the Hausdorff distances within a defined metric space.
Dice coefficient: Measure of the degree of overlap between sample sets, with a value of 1.0 representing complete overlap (range: 0–1.0).
$$DICE = \frac{2 * A \bigcap B}{A + B}$$
Jaccard index: Comparison of the similarity and diversity of sample sets, with a value of 1.0 representing unity (range 0–1.0).
$$Jaccard = \frac{A \bigcap B}{A + B - A \bigcap B}$$

relationship in which case a Pearson correlation may have been more appropriate.

RESULTS

The DILs were located in the anterior fibromuscular stroma, central zone, peripheral zone, and transition zone in 35.2, 5.6, 32.4, and 25.4% of patients, respectively. Gleason grade groups 1–5 represented 30.3, 39.4, 17.2, and 13.1% of the study population, respectively (with group grades 4 and 5 analyzed together). The mean contour volumes for the T2W images, and the ADC and K^{trans} maps were 2.14 ± 2.1, 2.22 ± 2.2, and 1.84 ± 1.5 mL, respectively. K^{trans} values were indistinguishable between cancerous regions and normal prostatic tissue for 19 patients.

The Dice coefficient and Jaccard index were 0.74 ± 0.13, 0.60 ± 0.15 for T2W-ADC, and 0.61 ± 0.16, 0.46 ± 0.16 for T2W-K^{trans}. For the voxel-based portion of the study, the Spearman correlations were 0.20 ± 0.20 for T2W-ADC and 0.13 ± 0.25 for T2W-K^{trans}.

Tables 2, **3** summarize the Spearman correlation, Dice coefficient, Jaccard index, as well as the HD and MDA by DIL location and Gleason grade groups, respectively.

The DIL was separately contoured for 19 patients by a second radiation oncologist to assess for interobserver variability (these results were not analyzed with respect to Gleason Grade and

TABLE 2 | Hausdorff distance (95%), Mean distance to agreement, Dice coefficient, Jaccard index, and Spearman-rank order by tumor location.

Location		AFS	Peripheral	Cent/Tran	Total
T2-ADC	HD	4.46 ± 1.69	4.11 ± 1.39	4.36 ± 2.43	4.32 ± 1.89
	MDA	0.96 ± 0.53	0.94 ± 0.45	1.00 ± 0.77	0.97 ± 0.60
	Dice	0.73 ± 0.12	0.69 ± 0.16	0.74 ± 0.16	0.74 ± 0.13
	Jaccard	0.59 ± 0.14	0.55 ± 0.17	0.61 ± 0.18	0.60 ± 0.15
	Spearman	0.15 ± 0.16	0.26 ± 0.26	0.18 ± 0.13	0.20 ± 0.20
T2-K^{trans}	HD	6.46 ± 2.42	5.77 ± 2.82	6.57 ± 4.18	6.21 ± 3.25
	MDA	1.65 ± 0.65	1.53 ± 0.97	1.63 ± 1.13	1.59 ± 0.95
	Dice	0.61 ± 0.15	0.59 ± 0.17	0.61 ± 0.19	0.61 ± 0.16
	Jaccard	0.45 ± 0.16	0.44 ± 0.17	0.46 ± 0.18	0.46 ± 0.16
	Spearman	0.18 ± 0.22	0.06 ± 0.26	0.19 ± 0.27	0.13 ± 0.25

Cent/Tran, central/transition; AFS, anterior fibromuscular stroma.

location). For this second set of contours, the Dice coefficient was 0.51 ± 0.19 for T2W-ADC and 0.42 ± 0.13 for T2W-K^{trans}. A comparison between the Dice and Jaccard coefficients, MDA, and HD for the 19 patients contoured by the two different physicians is shown in **Figure 2**.

Table 4 shows the mean, minimum, and maximum pixel values within the ADC and K^{trans} contours of the 90 patients split according to the Gleason grade group. The results of grade groups 4 and 5 were combined together due to the smaller sample sizes.

The mean ADC contoured pixel values ranged from 964.14 to 1007.74 mm^2/s among different groups while the K^{trans} values ranged from 3.27 to 6.21 min^{-1}.

DISCUSSION

Current national guidelines recommend dose-escalated radiation therapy to the entire intact prostate gland for men receiving definitive radiation therapy for prostate cancer. This approach has demonstrated clear benefits in biochemical progression-free survival across multiple randomized controlled studies. Unfortunately, biochemical recurrence rates can exceed 25% at 10 years necessitating further salvage therapy, which can adversely affect quality of life. Up to 90% of local recurrences after conventional radiation therapy have been shown to occur at the site of the DIL (4, 26). This coupled with the tremendous technological advancements in diagnostic imaging (27) and modern radiation therapy techniques such as treatment under image-guidance (28–30), has led to emerging interest in more accurately targeting the intraprostatic lesion and delivering a further boost to the dominant site of disease.

While the majority of these efforts have been realized using intensity-modulated radiation therapy (7, 9, 11, 31–33), studies have also included dose-escalation using brachytherapy with biologically equivalent doses of around 200 Gy to the DIL (6, 34–36), and more recently with stereotactic body radiotherapy using a simultaneous integrated boost technique (14, 32, 37, 38).

Clinical Outcomes of Radiation Therapy Boost to the DIL

Early results have demonstrated efficacy with low acute and late toxicities with either treatment approach. A recent systematic review of dose-escalated radiation therapy to the DIL

TABLE 3 | Hausdorff distance (95%), Mean distance to agreement, Dice coefficient, Jaccard index, and Spearman-rank order by Gleason grade group.

Gleason grade group		1	2	3	4, 5
T2-ADC	HD	4.33 ± 2.24	4.21 ± 1.96	4.20 ± 1.26	4.61 ± 1.28
	MDA	0.94 ± 0.68	0.96 ± 0.61	0.94 ± 0.37	1.03 ± 0.57
	Dice	0.73 ± 0.15	0.73 ± 0.13	0.71 ± 0.14	0.70 ± 0.21
	Jaccard	0.59 ± 0.15	0.59 ± 0.15	0.56 ± 0.15	0.57 ± 0.22
	Spearman	0.21 ± 0.16	0.25 ± 0.20	0.20 ± 0.19	0.04 ± 0.19
T2-K^{trans}	HD	6.07 ± 3.53	5.89 ± 2.87	6.07 ± 2.44	7.51 ± 4.62
	MDA	1.65 ± 1.29	1.50 ± 0.68	1.40 ± 0.66	2.02 ± 1.33
	Dice	0.60 ± 0.22	0.60 ± 0.14	0.65 ± 0.12	0.57 ± 0.22
	Jaccard	0.46 ± 0.21	0.44 ± 0.15	0.50 ± 0.12	0.43 ± 0.22
	Spearman	0.15 ± 0.24	0.10 ± 0.23	0.12 ± 0.32	0.15 ± 0.27

FIGURE 2 | Dice, Jaccard, Mean Distance to Agreement (MDA), and 95% Hausdoff Distance (HD) for 19 patients contoured by the two different physicians. One physician's contour comparisons (i.e., T2-ADC vs. T2-K^{trans}) are denoted by blue and orange, respectively, while the other physician's contour comparisons (i.e., T2-ADC vs. T2-K^{trans}) are denoted by gray and yellow, respectively. The data suggests strong dependence on physician performance and relatively high interobserver variability.

TABLE 4 | Mean, minimum, and maximum contoured pixel values for ADC and K^{trans}.

Gleason grade group	Number of patients (%)	ADC Mean	ADC min	ADC max	ADC SD	K^{trans} mean	K^{trans} min	K^{trans} max	K^{trans} SD
		mm/s	mm/s	mm/s		Min^{-1}	Min^{-1}	Min^{-1}	
1	30.3	964.14	661.55	1247.63	136.70	3.27	0	13.54	3.52
2	39.4	988.43	822.13	1380.78	130.65	3.81	0	10.54	3.19
3	17.2	1007.74	656.63	1334.02	163.89	6.21	0	17.29	4.92
4+5	13.1	999.62	770.11	1806.99	251.47	3.56	0	16.28	5.37

SD, standard deviation.

demonstrated that the average grade 3+ gastrointestinal and genitourinary late toxicity was ∼2–3% for intensity-modulated radiation therapy, 6–10% for stereotactic body radiotherapy, and 2–6% for brachytherapy (39). The median 5-years biochemical progression-free survival was reported to be 85%. However, the study population included patients of all risk groups with heterogenous use of androgen deprivation therapy. These factors need to be taken into consideration when interpreting the results of these studies.

mpMRI as a Tool for Target Volume Delineation

mpMRI demonstrates high sensitivity and specificity in the diagnosis and staging of prostate cancer, and its utility in this context has been extensively investigated (40, 41). However, its use as a tool in target volume delineation for the purposes of radiation treatments has not been adequately elucidated. Several barriers exist to incorporating mpMRI to define adequate radiation treatment volumes, one of the most significant being a lack of sufficient data to determine which mpMRI sequence is most accurate in defining the DIL. Groenendaal et al. developed a logistic regression model on DCE and DWI images to predict tumor presence and validated on whole-mount section histopathological images for 12 patients. The model achieved a receiver operating characteristic curve of 0.70 (41). However, the study was limited to peripheral zone and low Gleason score lesions (6 and 7). In addition, each image modality reflects different biological characteristics and may be individually inconsistent in tumor delineation particularly at the voxel level. It is unclear whether a combination of MR sequences would confer any advantage compared to a single mpMRI sequence when contouring the DIL, especially with respect to clinical outcomes such as biochemical control.

Considering the technical variability and lack of consensus on ADC and K^{trans} values for intraprostatic lesions, we did not use automatic threshold values for segmentation. ADC and K^{trans} values can have significant variations across different scanning protocols and MR scanners which makes quantitative analysis difficult. The monoexponential ADC model was used to describe the water diffusion behavior in this study. The b value selection as well as the duration and strength of diffusion sensitizing gradients could have impact on the ADC value. And the ADC values depended on many factors including cell density, size, shape, permeability, and perfusion effects. The complex diffusion dynamics of biological tissue required more advanced

compartment models such as intravoxel incoherent motion and vascular, extracellular, and restricted diffusion for cytometry in tumors. It was difficult to achieve an optimal balance of spatial and temporal resolution of the DCE scans in the pelvic region. In the past decade, several models have been recruited in pharmacokinetic analysis of clinical trial data and animal studies to calculate the plasma volume fraction, extravascular and extracellular volume fraction, and K^{trans} (42–45). However, few have examined whether the models are appropriate to the data (46, 47) and the variances and co-variances of parametric estimates, as well as the biases introduced by systematic errors, is generally lacking. Model selection, which is a potential solution since it defines the region of leaky microvasculature, a tumor signature, allows delineating different tumor regions and the temporal evolution of the local model and producing approximately unbiased estimate of vascular parameters that are relatively independent of variation in the details of image acquisition and equipment (48).

Tumor Size and Location Are Important Considerations

Smaller lesions pose a challenge to using mpMRI to accurately and reproducibly target the DIL as imaging precision is known to become less accurate as volume decreases. This was previously reported by Groenendaal et al. (49), citing the impact of noise and geometrical distortions induced by MRI machines in complicating the validity of functional MRI techniques for smaller volumes. In this study, the mean contour volume for the T2W images, and the ADC and K^{trans} maps were 2.14 ± 2.1, 2.22 ± 2.2, and 1.84 ± 1.5 mL, respectively. It is reasonable to assume that with such small volumes, even slight differences in contours can result in significantly altered results.

The location of the DIL also plays a role in precise target delineation with mpMRI as lesions involving different zones of the prostate gland can present unique challenges. Prostate cancer most commonly involves the peripheral zone of the gland and appears as a region of homogenous low-signal intensity on T2W. Tumor involving of the central gland can be more difficult to discern (e.g., due to benign prostatic hyperplasia), but cross-observer consensus can be reached in up to 80% of cases (50). Similarly, K^{trans} does not reliably differentiate prostate cancer from benign prostatic hyperplasia within the central zone of the prostate gland due to similarities in microvascular density exhibited by both conditions. In fact, K^{trans} values were

indistinguishable between tumor foci and the normal prostate gland for 19 patients in this study suggesting that the value of this mpMRI sequence may be limited to more peripheral lesions.

Quantitative Correlation Between the DIL and Tumor Location

To evaluate the quantitative correlation between contours on each imaging modality and its statistical dependence on tumor location and Gleason grade group, the Hausdorff distance (HD), mean distance to agreement (MDA), Dice coefficient, and Jaccard index were calculated between the contours for each pair of MR sequences (i.e., T2W vs. ADC, T2W vs. K^{trans}, and ADC vs. K^{trans}). **Table 2** summarizes the results of the statistical analysis based on tumor location. Between T2W-ADC and T2W-K^{trans}, the Dice coefficient was 0.74 ± 0.13 and 0.61 ± 0.16, respectively, and the Jaccard index was 0.60 ± 0.15 and 0.46 ± 0.16, respectively. This suggests that there was a relatively high level of overlap between the contours regardless of tumor location, and that the contours were slightly more similar than they were divergent. Furthermore, the results were consistently better between T2W-ADC vs. T2W-K^{trans} which may reflect the fact that T2W images provide anatomical information whereas K^{trans} maps reflect the permeability of regional vasculature; consequently, although we expect to appreciate a certain level of correlation between the two MR sequences, it is understandable that a more substantial overlap between the contours was not observed. Conversely, the voxel-based Spearman correlation was 0.21 ± 0.18 and 0.13 ± 0.25, respectively, suggesting that the strength of the association between the contours was not very robust.

Quantitative Correlation Between the DIL and Gleason Grade Group

Table 3 summarizes the results of the statistical analysis based on Gleason grade group. Between T2W-ADC and T2W-K^{trans}, the overall Dice coefficient and Jaccard index were identical to the results based on tumor location. Furthermore, the voxel-based Spearman correlation between T2W-ADC was similarly low, especially for Gleason grade groups 4 and 5 (0.04 ± 0.19) suggesting a very poor correlation between anatomical imaging and diffusion-weighted and perfusion-based imaging in poorly-differentiated prostate cancer. Again, the results were consistently better between T2W-ADC vs. T2W-K^{trans}.

Incorporating a Boolean Sum Volume (BSV) to Better Delineate the DIL

As previously mentioned, the Spearman correlation between tumor location and Gleason grade group for the MR sequences was rather weak. This was particularly so between the T2W images and ADC maps for lesions with Gleason grade groups 4 and 5 (although lower Gleason grade group did not necessarily predict for a higher correlation). This data would suggest that constructing a BSV that incorporates T2W images and ADC maps may be reasonable for delineating the DIL on mpMRI, as the BSV would adequately represent radiographic disease that is both anatomically- and functionally-defined. This is supported by the fact that the level of correlation between T2W images and ADC maps was relatively high but far from reaching unity.

This would, in theory, allow the entire DIL to be included in the radiation boost volume reducing the probability of a marginal miss especially with an adequately designed margin. The value of adding information provided by K^{trans} maps to the BSV remains investigational at this time as this mpMRI sequence was not reliably and consistently detectable as elaborated on above. A larger study population and a community consensus on quantitative analysis of K^{trans} may be warranted prior to its systematic incorporation into tumor delineation.

Interobserver Variability

Nineteen cases were contoured by two radiation oncologists in an effort to assess for interobserver variability. There was a large difference in the Dice coefficient between the contoured DILs (23 and 19% for T2W-ADC and T2W-K^{trans}, respectively). This is not surprising as significant interobserver variability is a known limitation in the interpretation of mpMRI images. As previously mentioned, the small volumes of the contours in this study (mean volumes ranging from 1.84 to 2.14 mL) may have amplified even the smallest of differences in tumor delineation, and whether these marginal statistical discrepancies would translate into meaningful differences in clinical outcome is debatable. Furthermore, it would be impractical for more than one radiation oncologist to delineate the DIL in clinical practice. A more pragmatic approach would be to develop an expert consensus guideline on DIL delineation coupled with suggestions for optimal clinical target volume margins to ensure adequate coverage.

Contoured Pixel Values

The mean, minimum, and maximum contoured pixel values for ADC and K^{trans} are tabulated in **Table 4**. This information is intended as a baseline threshold recommendation for automatic segmentation of ADC and K^{trans} maps based on Gleason grade group. The contours used to obtain this data were delineated by the original physician on 90 patients. Of note, the K^{trans} mean pixel value is relatively high compared to reported tumor regions in previous studies (40, 51–53). Ktrans images used in this study were procured by a method explained in Huisman et al. (54), which results with differing pixel values than other commonly used methods. Since this study has shown that there is a large variation of ADC and K^{trans} values in each Gleason grade group, future work is needed to recommend specific thresholds for automatic delineation with the verification of whole-mount histopathologic section findings.

Study Limitations

The limitations of this study include its retrospective design, inherent inconsistencies between functional MR images (e.g., different institutional imaging protocols such as contrast inject rate, variations in patient body mass index, and differences in spatial and temporal resolution), lack of histopathological validation, maximum b-value of 800 in calculating the ADC map, and tumor delineation by only two radiation oncologists. A prospectively designed study using standardized imaging with up-to-date protocols and contouring by a team of experienced radiation oncologists allowing for interobserver variability would strengthen the validity of these results.

CONCLUSIONS

Using mpMRI to delineate a target volume for a radiation boost is an emerging area of interest and one that may improve clinical outcomes without increasing the toxicity associated with external beam radiation therapy. The intraprostatic lesions contoured on T2W images had a high level of agreement with those contoured on ADC maps, but there was little to no quantitative correlation of these results with tumor location and Gleason grade group. As shown in the study, there have been many technical hurdles to be solved for precision radiotherapy to target the tumor based on physiological imaging and understand its corresponding treatment outcome. A BSV incorporating all available MR sequences may be reasonable at the current stage in delineating the DIL boost volume for clinical practice. A larger study population and a community consensus on quantitative analysis of K^{trans} is warranted prior to its systematic incorporation into tumor delineation.

DATA AVAILABILITY

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

This study uses a public cohort. There is no need to request the approval from the IRB Committee.

AUTHOR CONTRIBUTIONS

JL is responsible for study design, implementation, manuscript writing. EC is responsible for statistical analysis. AF is responsible for the implementation. MP and ME is responsible for the study conception, implementation. NW is responsible for study conception, study design, and manuscript editing.

ACKNOWLEDGMENTS

REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer statistics. *CA Cancer J Clin.* (2017) 67:7–30. doi: 10.3322/caac.21387

2. Dearnaley DP, Sydes MR, Graham JD, Aird EG, Bottomley D, Cowan RA, et al. Escalated-dose versus standard-dose conformal radiotherapy in prostate cancer: first results from the MRC RT01 randomised controlled trial. *Lancet Oncol.* (2007) 8:475–87. doi: 10.1016/S1470-2045(07)70143-2

3. Zietman AL, DeSilvio ML, Slater JD, Rossi CJ Jr, Miller DW, Adams JA, et al. Comparison of conventional-dose vs high-dose conformal radiation therapy in clinically localized adenocarcinoma of the prostate: a randomized controlled trial. *JAMA.* (2005) 294:1233–9. doi: 10.1001/jama.294.10.1233

4. Pucar D, Hricak H, Shukla-Dave A, Kuroiwa K, Drobnjak M, Eastham J, et al. Clinically significant prostate cancer local recurrence after radiation therapy occurs at the site of primary tumor: magnetic resonance imaging and step-section pathology evidence. *Int J Rad Oncol Biol Phys.* (2007) 69:62–9. doi: 10.1016/j.ijrobp.2007.03.065

5. Chopra S, Toi A, Taback N, Evans A, Haider MA, Milosevic M, et al. Pathological predictors for site of local recurrence after radiotherapy for prostate cancer. *Int J Radiat Oncol Biol Phys.* (2012) 82:e441–8. doi: 10.1016/j.ijrobp.2011.05.035

6. Vigneault E, Mbodji K, Racine LG, Chevrette E, Lavallee MC, Martin AG, et al. Image-guided high-dose-rate brachytherapy boost to the dominant intraprostatic lesion using multiparametric magnetic resonance imaging including spectroscopy: results of a prospective study. *Brachytherapy.* (2016) 15:746–51. doi: 10.1016/j.brachy.2016.09.004

7. Sundahl N, De Meerleer G, Villeirs G, Ost P, De Neve W, Lumen N, et al. Combining high dose external beam radiotherapy with a simultaneous integrated boost to the dominant intraprostatic lesion: analysis of genito-urinary and rectal toxicity. *Radiother Oncol.* (2016) 119:398–404. doi: 10.1016/j.radonc.2016.04.031

8. von Eyben FE, Kiljunen T, Kangasmaki A, Kairemo K, von Eyben R, Joensuu T. Radiotherapy boost for the dominant intraprostatic cancer lesion-A systematic review and meta-analysis. *Clin Genitourin Cancer.* (2016) 14:189–97. doi: 10.1016/j.clgc.2015.12.005

9. Schild MH, Schild SE, Wong WW, Vora SA, Silva AC, Silva AM, et al. Early outcome of prostate intensity modulated radiation therapy (IMRT) incorporating a simultaneous intra-prostatic MRI directed boost *OMICS. J Radiol.* (2014) 3:170. doi: 10.4172/2167-7964.1000170

10. Aluwini S, van Rooij P, Hoogeman M, Kirkels W, Kolkman-Deurloo IK, Bangma C. Stereotactic body radiotherapy with a focal boost to the MRI-visible tumor as monotherapy for low- and intermediate-risk prostate cancer: early results. *Radiat Oncol.* (2013) 8:84. doi: 10.1186/1748-717X-8-84

11. Fonteyne V, Villeirs G, Speleers B, De Neve W, De Wagter C, Lumen N, et al. Intensity-modulated radiotherapy as primary therapy for prostate cancer: report on acute toxicity after dose escalation with simultaneous integrated boost to intraprostatic lesion. *Int J Radiat Oncol Biol Phys.* (2008) 72:799–807. doi: 10.1016/j.ijrobp.2008.01.040

12. Schick U, Popowski Y, Nouet P, Bieri S, Rouzaud M, Khan H, et al. High-dose-rate brachytherapy boost to the dominant intra-prostatic tumor region: hemi-irradiation of prostate cancer. *Prostate.* (2011) 71:1309–16. doi: 10.1002/pros.21347

13. Singh AK, Guion P, Sears-Crouse N, Ullman K, Smith S, Albert PS, et al. Simultaneous integrated boost of biopsy proven, MRI defined dominant intra-prostatic lesions to 95 Gray with IMRT: early results of a phase I NCI study. *Radiat Oncol.* (2007) 2:36. doi: 10.1186/1748-717X-2-36

14. Herrera FG, Valerio M, Berthold D, Tawadros T, Meuwly JY, Vallet V, et al. 50 Gy stereotactic body radiation therapy to the dominant intra-prostatic nodule: results from a phase Ia/b trial. *Int J Radiat Oncol Biol Phys.* (2018) 103:320–34. doi: 10.1016/j.ijrobp.2018.09.023

15. Hoeks CM, Hambrock T, Yakar D, Hulsbergen-van de Kaa CA, Feuth T, Witjes JA, et al. Transition zone prostate cancer: detection and localization with 3-T multiparametric MR imaging. *Radiology.* (2013) 266:207–17. doi: 10.1148/radiol.12120281

16. Selnaes KM, Heerschap A, Jensen LR, Tessem MB, Schweder GJ, Goa PE, et al. Peripheral zone prostate cancer localization by multiparametric magnetic resonance at 3 T: unbiased cancer identification by matching to histopathology. *Invest Radiol.* (2012) 47:624–33. doi: 10.1097/RLI.0b013e318263f0fd

17. Delongchamps NB, Rouanne M, Flam T, Beuvon F, Liberatore M, Zerbib M, et al. Multiparametric magnetic resonance imaging for the detection and localization of prostate cancer: combination of T2-weighted,

dynamic contrast-enhanced, and diffusion-weighted imaging. *BJU Int.* (2011) 107:1411–8. doi: 10.1111/j.1464-410X.2010.09808.x

18. Dickinson L, Ahmed HU, Allen C, Barentsz JO, Carey B, Futterer JJ, et al. Magnetic resonance imaging for the detection, localisation, and characterisation of prostate cancer: recommendations from a European consensus meeting. *Eur Urol.* (2011) 59:477–94. doi: 10.1016/j.eururo.2010.12.009

19. Puech P, Potiron E, Lemaitre L, Leroy X, Haber GP, Crouzet S, et al. Dynamic contrast-enhanced-magnetic resonance imaging evaluation of intraprostatic prostate cancer: correlation with radical prostatectomy specimens. *Urology.* (2009) 74:1094–9. doi: 10.1016/j.urology.2009.04.102

20. Villers A, Puech P, Mouton D, Leroy X, Ballereau C, Lemaitre L. Dynamic contrast enhanced, pelvic phased array magnetic resonance imaging of localized prostate cancer for predicting tumor volume: correlation with radical prostatectomy findings. *J Urol.* (2006) 176:2432–7. doi: 10.1016/j.juro.2006.08.007

21. Isebaert S, Van den Bergh L, Haustermans K, Joniau S, Lerut E, De Wever L, et al. Multiparametric MRI for prostate cancer localization in correlation to whole-mount histopathology. *J Magn Reson Imaging.* (2013) 37:1392–401. doi: 10.1002/jmri.23938

22. Fedorov A, Penzkofer T, Hirsch MS, Flood TA, Vangel MG, Masry P, et al. The role of pathology correlation approach in prostate cancer index lesion detection and quantitative analysis with multiparametric MRI. *Acad Radiol.* (2015) 22:548–55. doi: 10.1016/j.acra.2014.12.022

23. Litjens G, Debats O, Barentsz J, Karssemeijer N, Huisman H. Computer-aided detection of prostate cancer in MRI. *IEEE Trans Med Imaging.* (2014) 33:1083–92. doi: 10.1109/TMI.2014.2303821

24. Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P, et al. The cancer imaging archive (TCIA): maintaining and operating a public information repository. *J Digit Imaging.* (2013) 26:1045–57. doi: 10.1007/s10278-013-9622-7

25. Vos PC, Hambrock T, Barentsz JO, Huisman HJ. Computer-assisted analysis of peripheral zone prostate lesions using T2-weighted and dynamic contrast enhanced T1-weighted MRI. *Phys Med Biol.* (2010) 55:1719–34. doi: 10.1088/0031-9155/55/6/012

26. Cellini N, Morganti AG, Mattiucci GC, Valentini V, Leone M, Luzi S, et al. Analysis of intraprostatic failures in patients treated with hormonal therapy and radiotherapy: implications for conformal therapy planning. *Int J Radiat Oncol Biol Phys.* (2002) 53:595–9. doi: 10.1016/S0360-3016(02)02795-5

27. Sciarra A, Barentsz J, Bjartell A, Eastham J, Hricak H, Panebianco V, et al. Advances in magnetic resonance imaging: how they are changing the management of prostate cancer. *Eur Urol.* (2011) 59:962–77. doi: 10.1016/j.eururo.2011.02.034

28. Zelefsky MJ, Kollmeier M, Cox B, Fidaleo A, Sperling D, Pei X, et al. Improved clinical outcomes with high-dose image guided radiotherapy compared with non-IGRT for the treatment of clinically localized prostate cancer. *Int J Radiat Oncol Biol Phys.* (2012) 84:125–9. doi: 10.1016/j.ijrobp.2011.11.047

29. Xie Y, Djajaputra D, King CR, Hossain S, Ma L, Xing L. Intrafractional motion of the prostate during hypofractionated radiotherapy. *Int J Radiat Oncol Biol Phys.* (2008) 72:236–46. doi: 10.1016/j.ijrobp.2008.04.051

30. Klayton T, Price R, Buyyounouski MK, Sobczak M, Greenberg R, Li J, et al. Prostate bed motion during intensity-modulated radiotherapy treatment. *Int J Radiat Oncol Biol Phys.* (2012) 84:130–6. doi: 10.1016/j.ijrobp.2011.11.041

31. De Meerleer G, Villeirs G, Bral S, Paelinck L, De Gersem W, Dekuyper P, et al. The magnetic resonance detected intraprostatic lesion in prostate cancer: planning and delivery of intensity-modulated radiotherapy. *Radiother Oncol.* (2005) 75:325–33. doi: 10.1016/j.radonc.2005.04.014

32. Miralbell R, Molla M, Rouzaud M, Hidalgo A, Toscas JI, Lozano J, et al. Hypofractionated boost to the dominant tumor region with intensity modulated stereotactic radiotherapy for prostate cancer: a sequential dose escalation pilot study. *Int J Radiat Oncol Biol Phys.* (2010) 78:50–7. doi: 10.1016/j.ijrobp.2009.07.1689

33. Ippolito E, Mantini G, Morganti AG, Mazzeo E, Padula GD, Digesu C, et al. Intensity-modulated radiotherapy with simultaneous integrated boost to dominant intraprostatic lesion: preliminary report on toxicity. *Am J Clin Oncol.* (2012) 35:158–62. doi: 10.1097/COC.0b013e318209cd8f

34. Quivrin M, Loffroy R, Cormier L, Mazoyer F, Bertaut A, Chambade D, et al. Multiparametric MRI and post implant CT-based dosimetry after prostate brachytherapy with iodine seeds: the higher the dose to the dominant index lesion, the lower the PSA bounce. *Radiother Oncol.* (2015) 117:258–61. doi: 10.1016/j.radonc.2015.08.020

35. Gomez-Iturriaga A, Casquero F, Urresola A, Ezquerro A, Lopez JI, Espinosa JM, et al. Dose escalation to dominant intraprostatic lesions with MRI-transrectal ultrasound fusion High-Dose-Rate prostate brachytherapy. *Prospect Phase II Trial Radiother Oncol.* (2016) 119:91–6. doi: 10.1016/j.radonc.2016.02.004

36. Crook J, Ots A, Gaztanaga M, Schmid M, Araujo C, Hilts M, et al. Ultrasound-planned high-dose-rate prostate brachytherapy: dose painting to the dominant intraprostatic lesion. *Brachytherapy.* (2014) 13:433–41. doi: 10.1016/j.brachy.2014.05.006

37. Feng Y, Welsh D, McDonald K, Carruthers L, Cheng K, Montgomery D, et al. Identifying the dominant prostate cancer focal lesion using image analysis and planning of a simultaneous integrated stereotactic boost. *Acta Oncol.* (2015) 54:1543–50. doi: 10.3109/0284186X.2015.1063782

38. Murray LJ, Lilley J, Thompson CM, Cosgrove V, Mason J, Sykes J, et al. Prostate stereotactic ablative radiation therapy using volumetric modulated arc therapy to dominant intraprostatic lesions. *Int J Radiat Oncol Biol Phys.* (2014) 89:406–15. doi: 10.1016/j.ijrobp.2014.01.042

39. Bauman G, Haider M, Van der Heide UA, Menard C. Boosting imaging defined dominant prostatic tumors: a systematic review. *Radiother Oncol.* (2013) 107:274–81. doi: 10.1016/j.radonc.2013.04.027

40. Kozlowski P, Chang SD, Jones EC, Berean KW, Chen H, Goldenberg SL. Combined diffusion-weighted and dynamic contrast-enhanced MRI for prostate cancer diagnosis–correlation with biopsy and histopathology. *J Magn Reson Imaging.* (2006) 24:108–13. doi: 10.1002/jmri.20626

41. Groenendaal G, Borren A, Moman MR, Monninkhof E, van Diest PJ, Philippens ME, et al. Pathologic validation of a model based on diffusion-weighted and dynamic contrast-enhanced magnetic resonance imaging for tumor delineation in the prostate peripheral zone. *Int J Radiat Oncol Biol Phys.* (2012) 82:e537–44. doi: 10.1016/j.ijrobp.2011.07.021

42. Tofts PS, Berkowitz B, Schnall MD. Quantitative analysis of dynamic Gd-DTPA enhancement in breast tumors using a permeability model. *Magnet Reson Med.* (1995) 33:564–8. doi: 10.1002/mrm.1910330416

43. Johnson TN, Rostami-Hodjegan A. Resurgence in the use of physiologically based pharmacokinetic models in pediatric clinical pharmacology: parallel shift in incorporating the knowledge of biological elements and increased applicability to drug development and clinical practice. *Paediatr Anaesth.* (2011) 21:291–301. doi: 10.1111/j.1460-9592.2010.03323.x

44. Tofts PS. Modeling tracer kinetics in dynamic Gd-DTPA MR imaging. *J Magn Reson Imaging.* (1997) 7:91–101. doi: 10.1002/jmri.1880070113

45. Brix G, Kiessling F, Lucht R, Darai S, Wasser K, Delorme S, et al. Microcirculation and microvasculature in breast tumors: pharmacokinetic analysis of dynamic MR image series. *Magnet Reson Med.* (2004) 52:420–9. doi: 10.1002/mrm.20161

46. Bagher-Ebadian H, Jain R, Nejad-Davarani SP, Mikkelsen T, Lu M, Jiang Q, et al. Model selection for DCE-T1 studies in glioblastoma. *Magnet Reson Med.* (2012) 68:241–51. doi: 10.1002/mrm.23211

47. Ewing JR, Bagher-Ebadian H. Model selection in measures of vascular parameters using dynamic contrast-enhanced MRI: experimental and clinical applications. *NMR Biomed.* (2013) 26:1028–41. doi: 10.1002/nbm.2996

48. Jain RK. Normalizing tumor microenvironment to treat cancer: bench to bedside to biomarkers. *J Clin Oncol.* (2013) 31:2205–18. doi: 10.1200/JCO.2012.46.3653

49. Groenendaal G, Moman MR, Korporaal JG, van Diest PJ, van Vulpen M, Philippens ME, et al. Validation of functional imaging with pathology for tumor delineation in the prostate. *Radiother Oncol.* (2010) 94:145–50. doi: 10.1016/j.radonc.2009.12.034

50. Vargas HA, Akin O, Franiel T, Goldman DA, Udo K, Touijer KA, et al. Normal central zone of the prostate and central zone involvement by prostate

cancer: clinical and MR imaging implications. *Radiology*. (2012) 262:894–902. doi: 10.1148/radiol.11110663

51. Ocak I, Bernardo M, Metzger G, Barrett T, Pinto P, Albert PS, et al. Dynamic contrast-enhanced MRI of prostate cancer at 3 T: a study of pharmacokinetic parameters. *AJR Am J Roentgenol*. (2007) 189:849. doi: 10.2214/AJR. 06.1329

52. van Dorsten FA, van der Graaf M, Engelbrecht MR, van Leenders GJ, Verhofstad A, Rijpkema M, et al. Combined quantitative dynamic contrast-enhanced MR imaging and (1)H MR spectroscopic imaging of human prostate cancer. *J Magn Reson Imaging*. (2004) 20:279–87. doi: 10.1002/jmri.20113

53. Groenendaal G, van den Berg CA, Korporaal JG, Philippens ME, Luijten PR, van Vulpen M, et al. Simultaneous MRI diffusion and perfusion imaging for tumor delineation in prostate cancer patients. *Radiother Oncol*. (2010) 95:185–90. doi: 10.1016/j.radonc.2010.02.014

54. Huisman HJ, Engelbrecht MR, Barentsz JO. Accurate estimation of pharmacokinetic contrast-enhanced dynamic MRI parameters of the prostate. *J Magnet Reson Imaging*. (2001) 13:607–14. doi: 10.1002/jmri.1085

A Multi-center Prospective Study for Implementation of an MRI-Only Prostate Treatment Planning Workflow

Peter Greer[1,2]*, Jarad Martin[1,2], Mark Sidhom[3,4], Perry Hunter[1], Peter Pichler[1], Jae Hyuk Choi[2], Leah Best[5], Joanne Smart[1], Tony Young[3,6], Michael Jameson[3,4,7,8], Tess Afinidad[3], Chris Wratten[1,2], James Denham[1,2], Lois Holloway[3,4,7], Swetha Sridharan[1], Robba Rai[3,4,7], Gary Liney[3,4,7], Parnesh Raniga[9] and Jason Dowling[2,4,9]

[1] Department of Radiation Oncology, Calvary Mater Newcastle, Newcastle, NSW, Australia, [2] School of Mathematical and Physical Sciences, University of Newcastle, Newcastle, NSW, Australia, [3] Liverpool Hospital Cancer Therapy Centre, South West Sydney Local Health District, Sydney, NSW, Australia, [4] South Western Sydney Clinical School, University of New South Wales, Sydney, NSW, Australia, [5] Hunter New England Imaging, HNE Health Service, Newcastle, NSW, Australia, [6] School of Physics, University of Sydney, Sydney, NSW, Australia, [7] Ingham Institute for Applied Medical Research, Sydney, NSW, Australia, [8] Centre for Medical Radiation Physics, University of Wollongong, Wollongong, NSW, Australia, [9] The Australian E-Health Research Centre, Commonwealth Scientific and Industrial Research Organisation, Brisbane, QLD, Australia

Purpose: This project investigates the feasibility of implementation of MRI-only prostate planning in a prospective multi-center study.

Method and Materials: A two-phase implementation model was utilized where centers performed retrospective analysis of MRI-only plans for five patients followed by prospective MRI-only planning for subsequent patients. Feasibility was assessed if at least 23/25 patients recruited to phase 2 received MRI-only treatment workflow. Whole-pelvic MRI scans (T2 weighted, isotropic 1.6 mm voxel 3D sequence) were converted to pseudo-CT using an established atlas-based method. Dose plans were generated using MRI contoured anatomy with pseudo-CT for dose calculation. A conventional CT scan was acquired subsequent to MRI-only plan approval for quality assurance purposes (QA-CT). 3D Gamma evaluation was performed between pseudo-CT calculated plan dose and recalculation on QA-CT. Criteria was 2%, 2 mm criteria with 20% low dose threshold. Gold fiducial marker positions for image guidance were compared between pseudo-CT and QA-CT scan prior to treatment.

Results: All 25 patients recruited to phase 2 were treated using the MRI-only workflow. Isocenter dose differences between pseudo-CT and QA-CT were $-0.04 \pm 0.93\%$ (mean $\pm$ SD). 3D Gamma dose comparison pass-rates were $99.7\% \pm 0.5\%$ with mean gamma 0.22 ± 0.07. Results were similar for the two centers using two different scanners. All gamma comparisons exceeded the 90% pass-rate tolerance with a minimum gamma pass-rate of 98.0%. In all cases the gold fiducial markers were correctly identified on MRI and the distances of all seeds to centroid were within the tolerance of 1.0 mm of the distances on QA-CT (0.07 ± 0.41 mm), with a root-mean-square difference of 0.42 mm.

Conclusion: The results support the hypothesis that an MRI-only prostate workflow can be implemented safely and accurately with appropriate quality assurance methods.

Keywords: MRI-only, synthetic CT, pseudo-CT, MRI-alone, prostate

INTRODUCTION

The benefit of MRI scanning for prostate radiation therapy planning is-well established with studies demonstrating lower inter-observer variation in contours, and smaller contours than CT with subsequent lower doses to normal tissues such as the penile bulb (1–3). The use of MRI for prostate delineation therefore potentially allows for more accurate and more consistent treatment. Typically for prostate planning as well as other treatment sites the MRI scans are registered to CT scans to allow for dose computation using the electron density or physical density map that is generated from simple calibration of Hounsfield Units (HU). The registration can be performed using MRI sequences that visualize implanted gold fiducial markers or less accurately using prostate soft-tissue. The MRI scans are often not acquired in the treatment position and do not encompass the patient external contour. The major limitation of this approach is that systematic registration uncertainties can result in the prostate contour from MRI (the target which is used to generate the high dose region) being misaligned to the gold fiducial positions on CT which are used for image guidance. These uncertainties have been estimated to be up to 2 mm in standard deviation which are significant given the small margins used for modern high dose treatments. These are systematic targeting uncertainties present for every treatment fraction.

Recently a new paradigm for treatment planning has emerged that of MRI-alone or MRI-only planning (4–13). In this approach the HU map for dose calculation is generated from one or more MRI sequences that encompass the typical field of view of a planning CT scan (pseudo/synthetic CT). A variety of methods have been developed to convert MRI data to HU including calibration and classification methods using the MRI voxel values, atlas-based methods that use deformable image registration, hybrid voxel, and atlas methods and deep-learning algorithms (convolutional neural networks and generative adversarial networks) (14, 15). The performance of these algorithms are similar and meet the requirements for dose calculation accuracy. Clinical acceptance is assessed by comparison of dose calculation on CT and pseudo-CT for individual patients. The increase in dose calculation uncertainty is regarded as a worthwhile trade-off to eliminate the systematic registration uncertainty (4).

While there have been many investigations performed retrospectively comparing new pseudo-CT methods to CT dose calculations, there has been less attention to clinical implementation of MRI-only workflows and in particular how these can be performed and assessed to ensure safe clinical use. Tyagi et al. presented a clinical workflow for MRI-only simulation (16). Their workflow included an initial CT simulation appointment where orthogonal x-ray scout images were used to determine patient dimensions and acceptance for use by the commercial MRCAT synthetic CT software. If the patient had prior brachytherapy a small field-of-view CT scan was acquired to distinguish brachytherapy seeds from fiducial markers. Forty-two patients from an initial cohort of 48 received this workflow. Tenhunen et al. presented their experience with

TABLE 1 | Patient details.

Patient detail	Mean [range]
Age (years)	73.4 [58–83]
Gleason score	3 + 3 = 6 (n = 3), 3 + 4 = 7 (n = 13), 4 + 3 = 7 (n = 10), 4 + 4 = 8 (n = 1), 4 + 5 = 9 (n = 3)
Pre-treatment PSA	9.0 [0.88–33.8]
Weight (kg)	84.4 [62–122]
Body mass index (BMI)	28.5 [19–39]

MRI-only prostate planning for a large cohort at Helsinki hospital (17). They found that 92% of patients were suitable for MRI-only workflow. To date these reports are for single institution studies.

MRI-only treatment planning is an entirely new approach for treatment centers and does entail potential risks. Recently a failure modes and effects analysis (FMEA) of MRI-only treatment planning was reported which demonstrated multiple failure modes that need to be considered (18). To gain benefit from these techniques it is important that MRI-only workflows be implemented in a rigorous and safe manner with appropriate quality assurance methods. In this work a multi-center study was initiated for the implementation of an MRI-only prostate workflow. Two different treatment centers participated and 30 patients in total were recruited, 15 at each center. The study was designed to enable and assess safe implementation of this new technique for radiation therapy departments.

METHODS AND MATERIALS

Patients

Thirty patients receiving radical radiation oncology treatment for prostate cancer were recruited across two treatment centers. The study title was High precision Prostate Substitute CT based External beam Radiotherapy (HIPSTER). The study was ethically approved by the Hunter New England Human Research Ethics Committee (HREC Registration No: 16/07/20/3.01, NSW HREC Reference No: HREC/16/HNE/298, Australian New Zealand Clinical Trials Registry ACTRN12616001653459) and informed consent was obtained from all patients. The study opened for recruitment 6 April 2017 and closed to recruitment 16 April 2019 with 15 patients recruited at each center. Eligibility criteria were men >18 years, low, intermediate or high risk prostate cancer, fiducial gold markers inserted and prostate or prostate and seminal vesicle irradiation. The exclusion criteria were inability to undergo MRI scanning, prior pelvic radiation therapy, unsafe for or refusal to undergo fiducial marker insertion, presence of hip prostheses, men highly dependent on medical care or men with mental or intellectual impairment that would have difficulty giving informed consent to the study. Patient details are listed in **Table 1**. Three fiducial markers were implanted at least 1 week prior to MRI scanning. Treatment details are listed in **Table 2**. Patients were scanned and treated according to local guidelines except for the MRI-only planning requirements outlined below.

TABLE 2 | Details of the centers equipment and techniques.

	Center 1	Center 2
CT scanner	Toshiba Acquilion or GE LightSpeed-RT or Siemens Confidence	Philips Brilliance Big Bore
MRI scanner	Siemens Magnetom Skyra 3T	Siemens Magnetom Skyra 3T
Treatment planning system	Varian Eclipse	Philips Pinnacle
Record and verify system	Varian Aria	Elekta Mosaiq
Linear accelerators	Varian Clinac or Truebeam	Elekta Synergy (Agility MLC) or Versa
Fiducial markers	1.0×3.0 mm gold	1.2×3.0 mm gold
Treatment technique	7-field IMRT ($n = 11$) 2-arc VMAT ($n = 4$)	1-arc VMAT ($n = 12$) 2-arc VMAT ($n = 3$)
Prescribed dose	60 Gy in 20 fractions ($n = 14$) 78 Gy in 39 fractions ($n = 1$)	60 Gy in 20 fractions ($n = 9$) 78 Gy in 39 fractions ($n = 6$)
Beam energy	6 MV ($n = 13$) 10 MV ($n = 2$)	6 MV

FIGURE 1 | Phase 1 design for retrospective analysis.
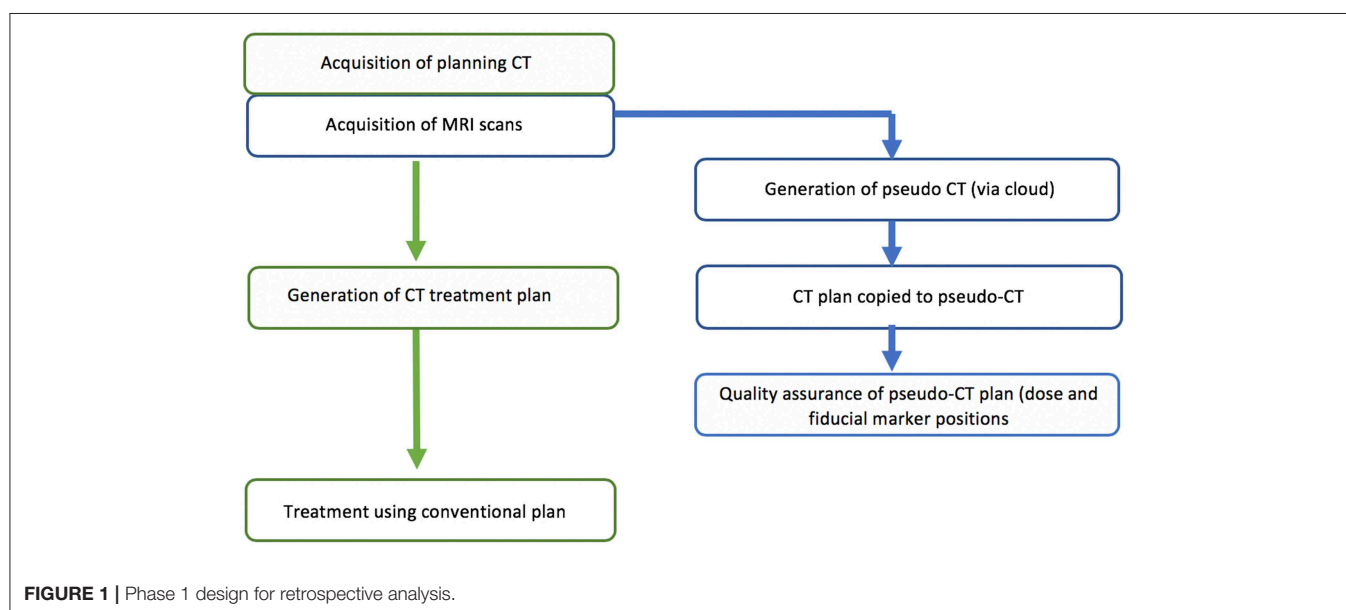
Centers and Equipment

While the centers had the same make and model of 3T MRI scanner (MAGNETOM Skyra, Siemens Healthineers, Erlangen, Germany) they differed in all other radiation therapy equipment. Both MRI scanners were fully equipped as MRI simulators with radiation therapy flat couch tops (CIVCO Medical Solutions, Coralville, USA), laser bridges (LAP Laser, Luneburg, Germany) and pelvic coil bridges (CIVCO). Both scanners had regular quality assurance procedures for image quality and distortion.

Study Design

The study was designed as a two-phase implementation model where centers performed retrospective analysis of MRI-only plans for five patients followed by prospective MRI-only planning for subsequent patients. The first phase is commensurate with literature studies to determine the accuracy of the pseudo-CT generation (**Figure 1**). The second phase is designed as a transition to MRI-only planning without CT where the MRI-only workflow is implemented but with final quality assurance

to ensure accuracy and safety provided by comparison to CT scanning. In this phase the CT scan (QA-CT) is only imported into the TPS following preliminary radiation oncologist approval of the MRI-only plan (**Figure 2**). The study aimed to recruit and treat 25 patients with phase 2 MRI-only prospective planning. As center 1 had previously performed retrospective analysis for 39 patients (19) they began at phase 2. Center 1 recruited 15 patients to phase 2 while center 2 recruited five patients to phase 1 and a further 10 patients to phase 2.

The major endpoint of the study was feasibility of MRI-only implementation with the aim achieved if >90% of patients received MRI-only treatment. This allowed for 2/25 patients to have their MRI-only plans deemed unacceptable. From previous experience 39/39 patients would have achieved the dose calculation criteria therefore a 25 patient sample size was regarded as reasonable to recruit and to demonstrate feasibility. Secondary endpoints were the assessment of the dose and image-guidance quality assurance metrics.
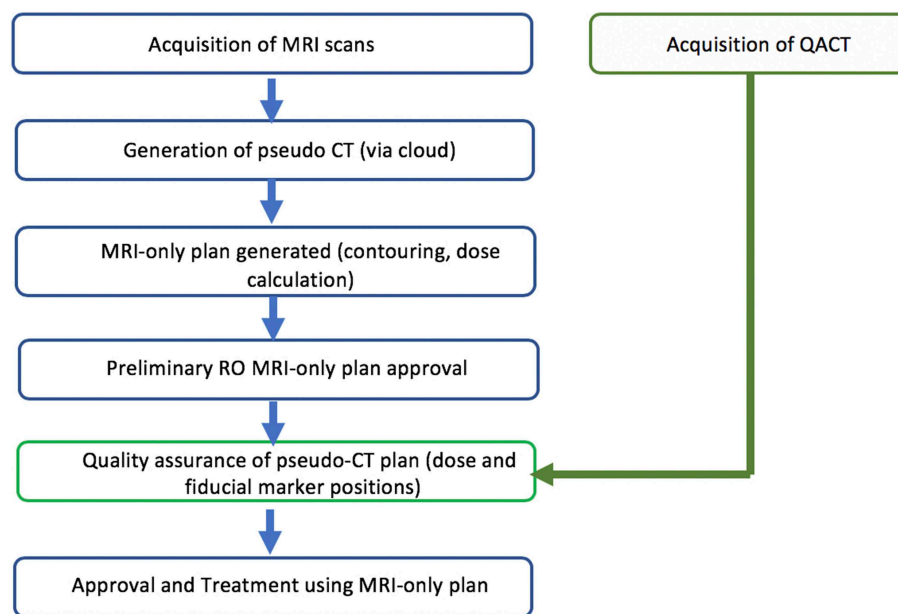
FIGURE 2 | Phase 2 study design for prospective MRI-only planning.

FIGURE 3 | Example patient setup for MRI simulation.

MRI Simulation

The patients were setup in exactly the same position as for treatment with a radiation therapist in attendance for patient setup. Patients were aligned using the lasers and MRI visible skin markers (Liquimark, Suremark) were placed on the patient's skin along with temporary tattoo marks. The coil mount was placed over the patient's pelvis, without compressing their contour. All patients were positioned head-first supine and had full bladder and empty rectum. An example of patient positioning for MRI simulation is shown in **Figure 3**.

TABLE 3 | Details of the MRI scans acquired and their function for MRI-only planning.

Scan type	Function
Small field-of view T2 TSE	Prostate delineation (CTV) and urethral delineation
Small field-of-view T1 GRE Flip 80	Fiducial marker delineation
Large field-of-view T2 SPACE	Organ delineation Generation of pseudo- CT
Pseudo-CT	Dose calculation Image guidance using fiducial marker contours transferred from T1 flip 80

A large field of view 3D sequence was utilized for pseudo-CT generation. Both centers used the same T2-weighted SPACE isotropic 3D sequence with 1.6 mm voxel side dimensions and scanning parameters as previously reported (19). The manufacturers 3D distortion correction was used for all scans.

Routine sequences used at each center were also acquired for prostate contouring and fiducial marker visualization. These were not altered for this study as the aim was to follow the conventional workflow as closely as possible but with MRI replacing the functionality of CT for treatment dose calculation, contouring and image-guidance. Details of these sequences have been reported earlier (19). The functionality of the three main sequences acquired along with the pseudo-CT are shown in **Table 3**. A checklist was designed to ensure adequate MRI scanning for treatment planning shown in **Table 4**.

Pseudo-CT Generation

Details of the pseudo-CT method have been reported in detail previously (19). The method is a hybrid atlas-voxel method using

an atlas of 39 previously acquired patients. The LFOV SPACE sequence was de-identified and the patient details replaced with a study ID before cloud upload to a secure site. The pseudo-CT was generated and downloaded to the treatment center where the patient ID and details were entered into the DICOM header of the scan, replacing the study ID. For the first eight patients the pseudo-CT generation was identical to the method described in Dowling et al. (19) including the addition of an extra 1.0 mm "skin" expansion due to the lack of visibility of this layer on MRI. However, this was discontinued due to erroneous generation of this layer for patient number 9 of the study and it was decided that it was clinically more robust to subsequently exclude this additional layer calculation from the algorithm.

CT Scanning

All patients received CT scans (QA-CT) for quality assurance and analysis of the MRI-only workflow performed as close as possible in time as the MRI scan and preferably after the MRI scan.

Slice thickness was 2.0 mm or 2.5 mm at Center 1 and 2.0 mm at Center 2.

MRI-Only Treatment Planning

The MRI sequences along with the pseudo-CT were imported into the treatment planning system (TPS). Alignment of all scans was visually checked by a radiation therapist. Following prostate, organ and fiducial marker delineation these contours were transferred to the pseudo-CT for treatment plan generation following incorporation of a couch-model. Imbedding of fiducials into the pseudo-CT scan pixel values was not used. Treatment plans were then defined according to routine department protocols. The pseudo-CT with attached fiducial marker contours was then transferred to the linear accelerator for image guidance with either cone-beam CT based image registration to pseudo-CT based on the markers or orthogonal kilovoltage x-ray image based image registration to digitally reconstructed radiographs generated from pseudo-CT.

Quality Assurance

A quality assurance procedure was designed for assessment of MRI-only treatment plans prior to acceptance of the plan for treatment. This included verification that the scans were consistent, pseudo-CT appearance and field-of-view, seeds were correctly identified, and dose and image-guidance metrics as described below (**Table 5**). This procedure is designed for an implementation phase for MRI-only planning where a MRI-only workflow is used but a gold-standard CT scan is still acquired for final verification before MRI-only plan is used for treatment.

Following full preparation of the MRI-only treatment plan and preliminary radiation oncologist approval, the QA-CT scan was imported into the TPS. This scan was registered to the pseudo-CT using automatic registration and the MRI plan transferred to the QA-CT. Dose was recalculated on the QA-CT using the same fluences and monitor units as the MRI

TABLE 4 | MRI simulation checklist.

MRI skin markers placed on tattoos, patient level with lasers (scanned HFS)	☐
LFOV MRI acquired first	☐
LFOV MRI—skin markers visible on LFOV MRI and patient is leveled (within 0.5 cm)	☐
LFOV MRI—covers external body contour in all directions and inferior superior extent according to CT scanning guidelines	☐
LFOV MRI—T2 SPACE, 1.6 mm isotropic voxels	☐
LFOV MRI—3D distortion correction is active	☐
SFOV T2 TSE—3D distortion correction is active	☐
SFOV T1 GRE flip80—three gold markers are visible on the scan, 2D distortion correction is active	☐

TABLE 5 | Quality assurance checklist for MRI-only plan.

Distortion correction	Confirm that 3D distortion correction was activated for the whole-pelvic scan. Check distortion corrections for other scans.	☐
Image transfer	Confirm that pseudo-CT corresponds to the MRI scan and conventional CT scan to verify that correct pseudo-CT has been assigned to the patient.	☐
Image orientation and appearance	Confirm that pseudo-CT is correctly oriented by comparison to conventional CT scan. Visually inspect the entire pseudo-CT volume and compare to conventional CT for any missing tissue or major differences.	☐
Field of view	Ensure that the pseudo-CT has sufficient field-of-view to cover all external contours and sufficient extension superiorly and inferiorly for dose calculations.	☐
Fiducial marker visibility	Verify that the fiducial marker structures generated on the pseudo-CT correspond to the fiducial markers determined from the conventional CT (i.e., all fiducial marker locations have been correctly identified).	☐
Femoral heads	Confirm visually that MRI generated bone contours visually correspond to CT bone contours.	☐
Dose at isocenter	Verify that isocenter dose on pseudo-CT is within 2% of conventional CT	☐
Dose distribution	Verify that 3D Gamma comparison at 2%, 2 mm criteria > 90% pass-rate for the entire body volume (−1.5 cm to avoid skin region where dose is uncertain).	☐
Fiducial marker positions	Verify that fiducial marker contours on pseudo-CT are within 1 mm from centroid of the locations on conventional CT from centroid (accounting for prostate rotation).	☐
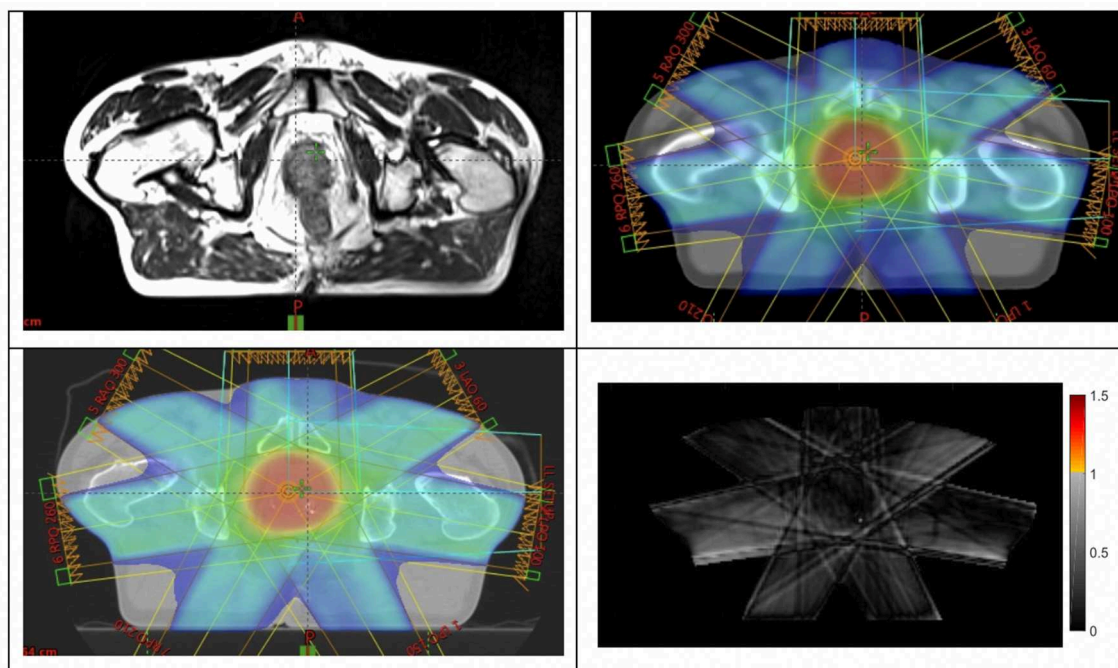
FIGURE 4 | Example of a patient (top-left) large-field-of-view MRI scan; (top-right) dose plan developed on pseudo-CT; (bottom-left) dose recalculated on QA-CT scan; (bottom-right) gamma analysis result at 2%, 2 mm criteria.

plan. Following alignment using the isocenters the doses were interpolated onto a 1.5 mm voxel size and compared with a three-dimensional gamma calculation. A 20 mm region close to the skin was excluded from the comparison using a two-dimensional erosion operation on each axial plane to avoid the large dose discrepancies due to differences in the external body contour at CT and MRI. A dose threshold of 20% of the maximum dose was used and gamma criteria of 3%, 3 mm, 2%, 2 mm, and 2%, 1 mm with the QA-CT as the reference dose for the comparison. Doses at the isocenter were also compared. Acceptance criteria for the dose calculation on pseudo-CT were isocenter dose within 2% and gamma pass-rate > 90% at 2%, 2 mm criteria.

Locations of fiducial markers as identified on MRI were also compared to locations on the QA-CT scan. The x, y, and z locations of the markers were carefully measured on the scans and entered into an Excel spreadsheet which calculated the centroid of the markers for each scan. The distances of each marker to the centroid were calculated and compared for the scans. If all distances were within 1.0 mm then the MRI locations were accepted.

RESULTS

The primary outcome of the study was achieved with all 25 patients in phase 2 having their MRI-only plans accepted by the radiation oncologist and passing all quality assurance criteria. These patients were all treated using the MRI-only workflow.

Figure 4 shows an example patient MRI scan, pseudo-CT dose calculation, and QA-CT dose recalculated for comparison.

For the secondary endpoints all 30 patient results were assessed including the five patients for phase 1 at center 2 as the assessment methodology is the same as phase 2. The results for the ratio of isocenter dose on pseudo-CT and QA-CT are shown in **Figure 5** along with the Bland-Altman levels. The mean difference in isocenter doses was −0.04% with a standard deviation of 0.93%. The effect of the first eight patients calculated with the 1.0 mm skin expansion can be seen with lower pseudo-CT doses. The mean difference for the first eight patients was −0.64% (0.90%) while for the subsequent patients it was 0.17% (0.85%). All isocenter dose differences were within 2.0% and only 3 (10%) had more than 1.5% difference.

The results for the gamma evaluations of the dose on pseudo-CT and QA-CT are shown in **Table 6** for the three gamma criteria. The average gamma pass-rates and the average of the mean gamma values for all patients are shown.

The results for the comparison of fiducial marker distances to centroid on MRI and QA-CT are shown in **Figure 6**. The average difference between MRI and QA-CT was 0.07 mm (1 SD = 0.41 mm) and the root-mean-square difference was 0.42 mm. The maximum difference was 1.00 mm.

DISCUSSION

This study demonstrates that MRI-only workflows can be implemented in a multi-center setting with appropriate quality assurance measures to ensure accurate and safe treatment. The
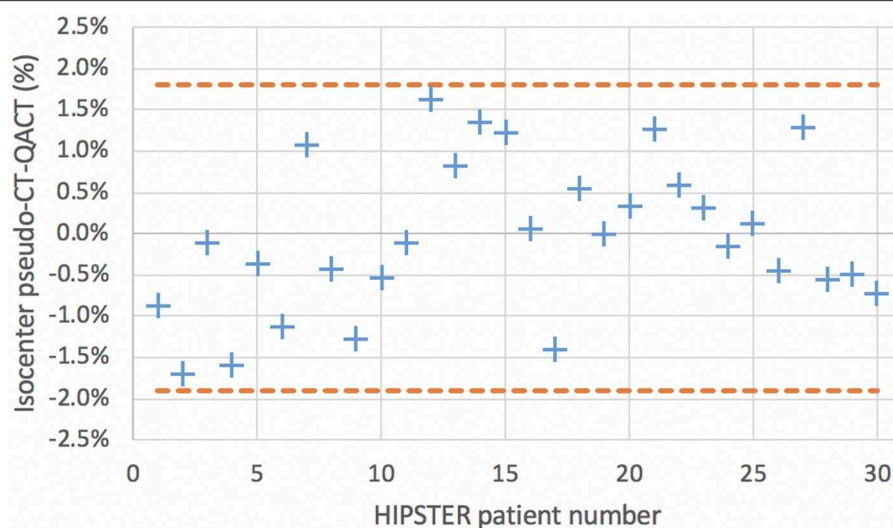
FIGURE 5 | Isocenter dose comparison on MRI only pseudo-CT scan and QA-CT scan.

TABLE 6 | Results of gamma analysis for comparison of dose calculation using pseudo-CT and CT.

	3%, 3 mm	2%, 2 mm	2%, 1 mm
Gamma pass-rate (%)	100.0	99.7	99.2
Standard deviation (%)	0.1	0.5	1.0
Mean gamma	0.145	0.218	0.221
Standard deviation	0.05	0.07	0.07

study is distinct from most other reported studies in that it is prospective and the 25 patients received an MRI-only workflow and treatment.

The QA-CT scan was only used for quality assurance purposes and this was imported following full generation and preliminary approval of the MRI only plan. This ensures that an MRI-only planning workflow is fully implemented but also allows for verification against the gold standard for dose calculation and image-guidance. This prepares the center for MRI-only workflow and ensures safe practice. Subsequent to this implementation phase the center could then use the MRI-only workflow without CT acquisition. There are two major potential approaches to this; consider that there is now adequate confidence in the process that no specific quality assurance techniques are required; or to utilize separate quality assurance techniques, and these decisions will be center-dependent. To provide a method for the latter approach, in parallel with this study a simple bulk-density calculation method was developed to compare to the pseudo-CT dose calculation. This was based on MRI bone and body contour anatomy and the results will be reported separately. This method can provide confidence in the integrity of the pseudo-CT and is robust and easy to perform.

Quality assurance methods to validate fiducial marker positions as identified on MRI scans would also be beneficial.
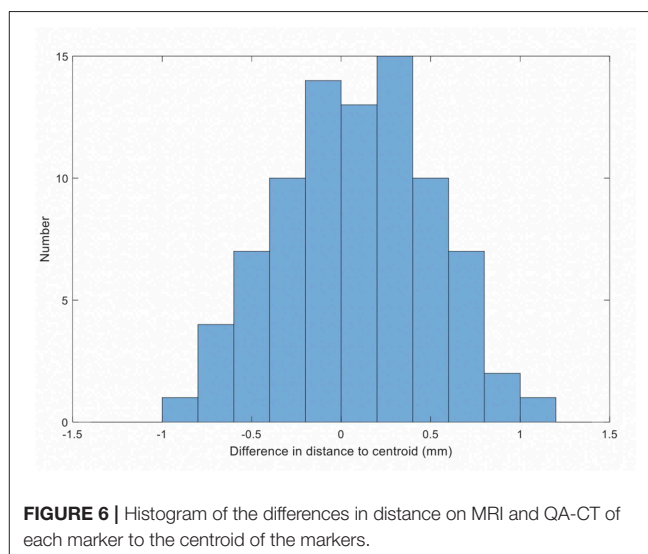
FIGURE 6 | Histogram of the differences in distance on MRI and QA-CT of each marker to the centroid of the markers.

Prostate calcifications can in some cases be difficult to differentiate from fiducial markers in the sequences used here. Although potentially problematic this misidentification is unlikely to lead to image guidance errors as the seed positions are clearly identified with cone-beam CT scans or x-ray images prior to treatment and misidentified seed positions are obvious and can be corrected. However, this is not an ideal scenario as it could delay treatment. Several methods have been proposed to ensure fiducial marker identification using MRI techniques or planar x-ray imaging (16, 20–25).

Patient movement between and during MRI scans is also a potential source of error in MRI-only planning as the scans can take several minutes to acquire. It is critical to ensure that movement has not occurred between the small field-of-view acquisitions used for CTV definition and fiducial marker

delineation. If there is a shift of position between them this will introduce a systematic error in dose delivery to the CTV/PTV. Visual inspection of alignment of the prostate contour on the two sequences should be performed. Note that this is not a problem specific to MRI-only planning. This problem also exists for MRI-CT registration based treatment planning as is currently performed. Movement of the patient for the large-field-of-view MRI scan is not as critical for dose calculation but it will result in systematic errors of normal tissues that are delineated on this scan and hence potential mismatch of planned and delivered doses to these organs.

For patient 9 an error in the pseudo-CT scan was detected visually during plan generation. This was due to the algorithm component that introduces an additional "skin" expansion to compensate for the lack of visibility of the skin on MRI. Although this correction was introduced in earlier method development to improve dose calculation accuracy it was felt that it would be clinically safer to exclude this additional layer for this and subsequent patients. This patients pseudo-CT was recalculated with the modified algorithm which generated a new pseudo-CT that was used for the treatment plan. This has a small effect on the dose calculation when compared to QA-CT. The patients prior to patient 9 that included this layer had on average slightly lower dose calculation on pseudo-CT compared to CT whereas the patients subsequent to the change had on average slightly higher dose on pseudo-CT when compared to CT. The patients prior to the change could be recalculated with the modified algorithm however the patient plans were developed using the prior algorithm so this would not reflect the reality for this prospective study.

CONCLUSION

An MRI-only workflow was introduced in a prospective multi-center trial setting and all recruited (25 patients) received the MRI-only workflow. MRI-only planning workflow can be implemented in a safe manner with appropriate testing and quality assurance.

DATA AVAILABILITY

The datasets generated for this study are available on request to the corresponding author.

AUTHOR CONTRIBUTIONS

PG: principal investigator, project manager, and wrote manuscript. JM and MS: patient recruitment, treatment planning, and clinical guidance. PH and PP: patient data acquisition, treatment planning, and protocol development. JC: data analysis and reporting of results. LB and GL: patient data acquisition and imaging development. JS: patient recruitment, data management, and ethics management. TY: patient data acquisition, data analysis, and quality assurance. MJ and LH: patient data acquisition, data analysis, and protocol development. TA: patient recruitment, data management, and quality assurance. CW and JDe: patient recruitment and clinical guidance. SS: patient recruitment, clinical guidance, and procedures development. RR: patient data acquisition, imaging development, and protocol development. PR and JDo: synthetic CT generation and technical development.

FUNDING

REFERENCES

1. Roach III M, Faillace-Akazawa P, Malfatti C, Holland J, Hricak H. Prostate volumes defined by magnetic resonance imaging and computerized tomographic scans for three-dimensional conformal radiotherapy. *Int J Radiat Oncol Biol Phys.* (1996) 35:1011–8. doi: 10.1016/0360-3016(96)00232-5

2. Debois M, Oyen R, Maes F, Verswijvel G, Gatti G, Bosmans H, et al. The contribution of magnetic resonance imaging to the three-dimensional treatment planning of localized prostate cancer. International *J Radiat Oncol Biol Phys.* (1999) 45:857–65. doi: 10.1016/S0360-3016(99)00288-6

3. Rasch C, Barillot I, Remeijer P, Touw A, Van Herk M, Lebesque JV. Definition of the prostate in CT and MRI: a multi-observer study. *Int J Radiat Oncol Biol Phys.* (1999) 43:57–66. doi: 10.1016/S0360-3016(98)00351-4

4. Owrangi AM, Greer PB, Glide-Hurst CK. MRI-only treatment planning: benefits and challenges. *Phys Med Biol.* (2018) 63:05TR1. doi: 10.1088/1361-6560/aaaca4

5. Lee YK, Bollet M, Charles-Edwards G, Flower MA, Leach MO, McNair H, et al. Radiotherapy treatment planning of prostate cancer using magnetic resonance imaging alone. *Radiother Oncol.* (2003) 66:203–16. doi: 10.1016/S0167-8140(02)00440-1

6. Chen L, Price RA Jr, Nguyen TB, Wang L, Li JS, Qin L, et al. Dosimetric evaluation of MRI-based treatment planning for prostate cancer. *Phys Med Biol.* (2004) 49:5157–70. doi: 10.1088/0031-9155/49/22/010

7. Pasquier D, Betrouni N, Vermandel M, Lacornerie T, Lartigau E, Rousseau J. MRI alone simulation for conformal radiation therapy of prostate cancer: technical aspects. *Ann Int Confer IEEE Eng Med Biol Proc.* (2006) 1:160–3. doi: 10.1109/IEMBS.2006.260341

8. Wang C, Chao M, Lee L, Xing L. MRI-based treatment planning with electron density information mapped from CT images: a preliminary study. *Technol Cancer Res Treat.* (2008) 7:341–7. doi: 10.1177/153303460800700501

9. Nyholm T, Karlsson MG, Karlsson M. MR based treatment workflow for external radiotherapy of prostate cancer. In: Dössel O, Schlegel WC, editors. *World Congress on Medical Physics and Biomedical Engineering, September 7 – 12, 2009, Munich, Germany. IFMBE Proceedings*, Vol 25/1. Berlin, Heidelberg: Springer. (2009). doi: 10.1007/978-3-642-03474-9_16

10. Dowling J, Lambert J, Parker J, Greer PB, Fripp J, Denham J, et al. Automatic MRI atlas-based external beam radiation therapy treatment planning for prostate cancer. In: *International Workshop on Prostate Cancer Imaging: Computer-Aided Diagnosis, Prognosis, and Intervention Held in Conjunction With MICCAI.* Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Beijing (2010). p. 25–33.

11. Jonsson JH, Karlsson MG, Karlsson M, Nyholm T. Treatment planning using MRI data: an analysis of the dose calculation accuracy for different treatment regions. *Radiat Oncol.* (2010) 5:62. doi: 10.1186/1748-717 X-5-62

12. Greer PB, Dowling JA, Lambert JA, Fripp J, Parker J, Denham JW, et al. A magnetic resonance imaging-based workflow for planning radiation therapy for prostate cancer. *Med J Aust.* (2011) 194:S24–7.

13. Kerkmeijer LGW, Maspero M, Meijer GJ, van der Voort van Zyp JRN, de Boer HCJ, van den Berg CAT. Magnetic resonance imaging only workflow for radiotherapy simulation and planning in prostate cancer. *Clin Oncol.* (2018) 30:692–701. doi: 10.1016/j.clon.2018.08.009

14. Edmund JM, Nyholm T. A review of substitute CT generation for MRI-only radiation therapy. *Radiat Oncol.* (2017) 12. doi: 10.1186/s13014-016-0747-y

15. Johnstone E, Wyatt JJ, Henry AM, Short SC, Sebag-Montefiore D, Murray L, et al. Systematic review of synthetic computed tomography generation methodologies for use in magnetic resonance imaging–only radiation therapy. *Int J Radiat Oncol Biol Phys.* (2018) 100:199–217. doi: 10.1016/j.ijrobp.2017.08.043

16. Tyagi N, Fontenla S, Zhang J, Cloutier M, Kadbi M, Mechalakos J, et al. Dosimetric and workflow evaluation of first commercial synthetic CT software for clinical use in pelvis. *Phys Med Biol.* (2017) 62:2961. doi: 10.1088/1361-6560/aa5452

17. Tenhunen M, Korhonen J, Kapanen M, Seppälä T, Koivula L, Collan J, et al. MRI-only based radiation therapy of prostate cancer: workflow and early clinical experience. *Acta Oncol.* (2018) 57:902–7. doi: 10.1080/0284186X.2018. 1445284

18. Kim J, Miller B, Siddiqui MS, Movsas B, Glide-Hurst C. FMEA of MR-only treatment planning in the pelvis. *Adv Radiat Oncol.* (2019) 4:168–76. doi: 10.1016/j.adro.2018. 08.024

19. Dowling JA, Sun J, Pichler P, Rivest-Hénault D, Ghose S, Richardson H, et al. Automatic substitute computed tomography generation and contouring for magnetic resonance imaging (MRI)-alone external beam radiation therapy from standard MRI sequences. *Int J Radiat Oncol Biol Phys.* (2015) 93:1144–53. doi: 10.1016/j.ijrobp.2015.08.045

20. Ghose S, Mitra J, Rivest-Hénault D, Fazlollahi A, Stanwell P, Pichler P, et al. MRI-alone radiation therapy planning for prostate cancer: automatic fiducial marker detection. *Med Phys.* (2016) 43:2218–28. doi: 10.1118/1.4944871

21. Maspero M, Seevinck PR, Willems NJW, Sikkes GG, de Kogel GJ, de Boer HCJ, et al. Evaluation of gold fiducial marker manual localisation for magnetic resonance-only prostate radiotherapy. *Radiat Oncol.* (2018) 13:105. doi: 10.1186/s13014-018-1029-7

22. Maspero M, Van Den Berg CAT, Zijlstra F, Sikkes GG, De Boer HCJ, Meijer GJ, et al. Evaluation of an automatic MR-based gold fiducial marker localisation method for MR-only prostate radiotherapy. *Phys Med Biol.* (2017) 62:7981–8002. doi: 10.1088/1361-6560/aa875f

23. Dinis Fernandes C, Dinh CV, Steggerda MJ, Ter Beek LC, Smolic M, Van Buuren LD, et al. Prostate fiducial marker detection with the use of multi-parametric magnetic resonance imaging. *Phys Imag Radiat Oncol.* (2017) 1:14–20. doi: 10.1016/j.phro.2017.02.001

24. Gustafsson C, Korhonen J, Persson E, Gunnlaugsson A, Nyholm T, Olsson LE. Registration free automatic identification of gold fiducial markers in MRI target delineation images for prostate radiotherapy. *Med Phys.* (2017) 44:5563–74. doi: 10.1002/mp.12516

25. Jonsson JH, Garpebring A, Karlsson MG, Nyholm T. Internal fiducial markers and susceptibility effects in MRI - simulation and measurement of spatial accuracy. *Int J Radiat Oncol Biol Phys.* (2012) 82:1612–8. doi: 10.1016/j.ijrobp.2011.01.046

Conflict of Interest Statement: 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.

Generation of Synthetic CT Images From MRI for Treatment Planning and Patient Positioning Using a 3-Channel U-Net Trained on Sagittal Images

Dinank Gupta*, Michelle Kim, Karen A. Vineberg and James M. Balter

Department of Radiation Oncology, University of Michigan, Ann Arbor, MI, United States

A novel deep learning architecture was explored to create synthetic CT (MRCT) images that preserve soft tissue contrast necessary for support of patient positioning in Radiation therapy. A U-Net architecture was applied to learn the correspondence between input T1-weighted MRI and spatially aligned corresponding CT images. The network was trained on sagittal images, taking advantage of the left-right symmetry of the brain to increase the amount of training data for similar anatomic positions. The output CT images were divided into three channels, representing Hounsfield Unit (HU) ranges of voxels containing air, soft tissue, and bone, respectively, and simultaneously trained using a combined Mean Absolute Error (MAE) and Mean Squared Error (MSE) loss function equally weighted for each channel. Training on 9192 image pairs yielded resulting synthetic CT images on 13 test patients with MAE of 17.6+/−3.4 HU (range 14–26.5 HU) in soft tissue. Varying the amount of training data demonstrated a general decrease in MAE values with more data, with the lack of a plateau indicating that additional training data could further improve correspondence between MRCT and CT tissue intensities. Treatment plans optimized on MRCT-derived density grids using this network for 7 radiosurgical targets had doses recalculated using the corresponding CT-derived density grids, yielding a systematic mean target dose difference of 2.3% due to the lack of the immobilization mask on the MRCT images, and a standard deviation of 0.1%, indicating the consistency of this correctable difference. Alignment of MRCT and cone beam CT (CBCT) images used for patient positioning demonstrated excellent preservation of dominant soft tissue features, and alignment comparisons of treatment planning CT scans to CBCT images vs. MRCT to CBCT alignment demonstrated differences of −0.1 (σ 0.2) mm, −0.1 (σ 0.3) mm, and −0.2 (σ 0.3) mm about the left-right, anterior-posterior and cranial-caudal axes, respectively.

Keywords: synthetic CT, MRCT, deep learning, MRI, radiation oncology

INTRODUCTION

While MRI has shown significant value for Radiation Oncology treatment of intracranial tumors due to its superior soft tissue contrast and ability to map quantitative biological features such as diffusion and perfusion, it has inherent limitations in providing electron density maps necessary to support calculation of radiation dose distributions, as well as in supporting most existing clinical workflows for patient positioning that rely on alignment of treatment planning CT images with Cone Beam CT (CBCT) scans acquired at the time of patient positioning. While the former issue has been reasonably resolved by a variety of synthetic CT approaches (1–6), the latter has received little attention.

Many CBCT-CT alignment mechanisms rely on reasonably similar intensity distributions, especially those that align soft tissue features. Recent reports have demonstrated the potential of "machine learning" approaches to generate synthetic CT ("MRCT") scans, but have shown rather large errors in intensity differences of the soft tissues of the brain. While not specifically analyzed in most of these investigations, the structural details of soft tissue features are often misrepresented, thus potentially confounding alignment with similar features displayed on CBCT image volumes. This may present challenges for precise local alignment of tissues, as the potential for local changes between simulation and treatment is enhanced due to the temporal periods associated with frameless radiosurgery techniques (7).

The objective of this investigation was to investigate whether a Neural Network could be optimized to preserve the soft tissue contrast features necessary for precision alignment of intracranial tumors. Attempts to maximize local contrast include use of a U-Net architecture trained on aligned MR and CT pairs, training on sagittal planes to increase data diversity for the same number of input patients, and separation of the CT images into three intensity regions, preserving the narrow intensity range wherein most of the soft tissue contrast falls on CT. The impact of numbers of training images is briefly explored.

MATERIALS AND METHODS

Training Data

Under an Institutional Review Board approved protocol, 60 patients who underwent CT-based simulation for treatment of intracranial tumors further underwent an MR simulation scan while immobilized with their fixation devices. CT image volumes were all acquired using the same in-house CT simulator (Brilliance big bore, Philips Medical Systems, Andover MA) and had initial voxel sizes ranging from 0.6 by 0.6 by 1 mm to 1.17 by 1.17 by 3 mm. MR images were acquired on an in-house 3 Tesla MR Simulator (Skyra, Siemens Healthineers, Erlangen, Germany), and included a T1-weighted acquisition as the in phase images of a Dixon scan series. These images, with sampled voxel sizes of ~1 by 1.25 by 1.25 mm, were used for training.

CT image volumes were rigidly aligned to corresponding MR images using an open source package *dipy* (8) and the resulting transforms applied to the CT and resampled to match the native MR image resolution.

The range of Hounsfield Units of typical human tissues is roughly −1,000 to 2,000. The majority of this intensity range is occupied by air and/or skeletal tissues. Most soft tissue falls within a narrow subset of HU values (~−100 to 100 HU). As a result of this very limited region of the intensity range wherein soft tissue contrast lies, training loss functions will have HU differences an order of magnitude higher in air or bone regions than in locations consisting of primarily soft tissue contrast. As a result, the training might prioritize errors in bone or air over those of soft tissue. This leads to a potential challenge to preserving local soft tissue structures, especially with limited amounts of training data. To attempt to capture soft tissue contrast, we split training CT images into 3 separate output "channels" that can facilitate easier learning from limited data sets (**Figure 1**). These channels were defined by using intensity thresholds of < −100 HU to define voxels containing air, −100 HU to 100 HU to primarily identify soft tissue and >100 HU for voxels containing bone. The regions outside of the threshold masks were set to 0 HU for each channel. This 3-channel approach forces the network to learn the 3 regions of interests separately, thus capturing the tissue intensity contrasts independently for air, tissue and bone. As the tissue intensities were consistent across the MR scans due to a standard image acquisition methodology and coil configuration, and the HU value ranges of tissues were similarly consistent, a fixed normalization was applied to the input and separately each of the output channels.
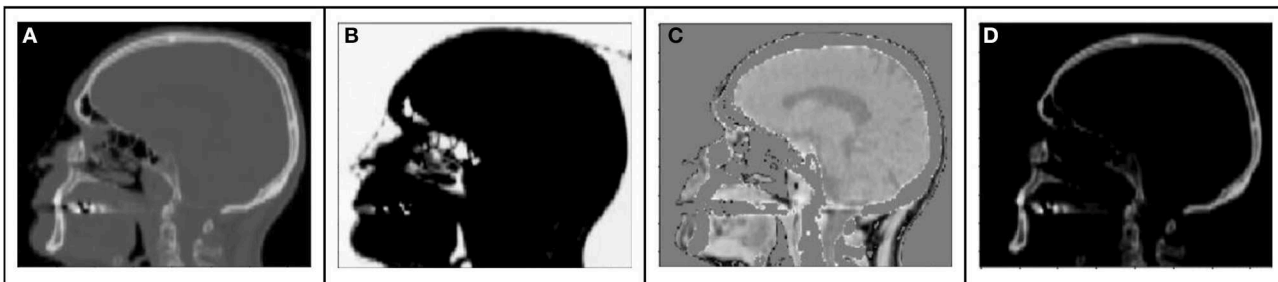
FIGURE 1 | Example sagittal CT image with tissue windows: **(A)** original CT image **(B)** air window **(C)** tissue window, **(D)** bone window.

FIGURE 2 | U-net architecture. Each block represents a convolution operation, followed by batch normalization and Leaky ReLU. Last convolution operation converts 64 dimensional channel to 3 channel synthetic CT.

FIGURE 3 | Normalized training and validation loss curves.
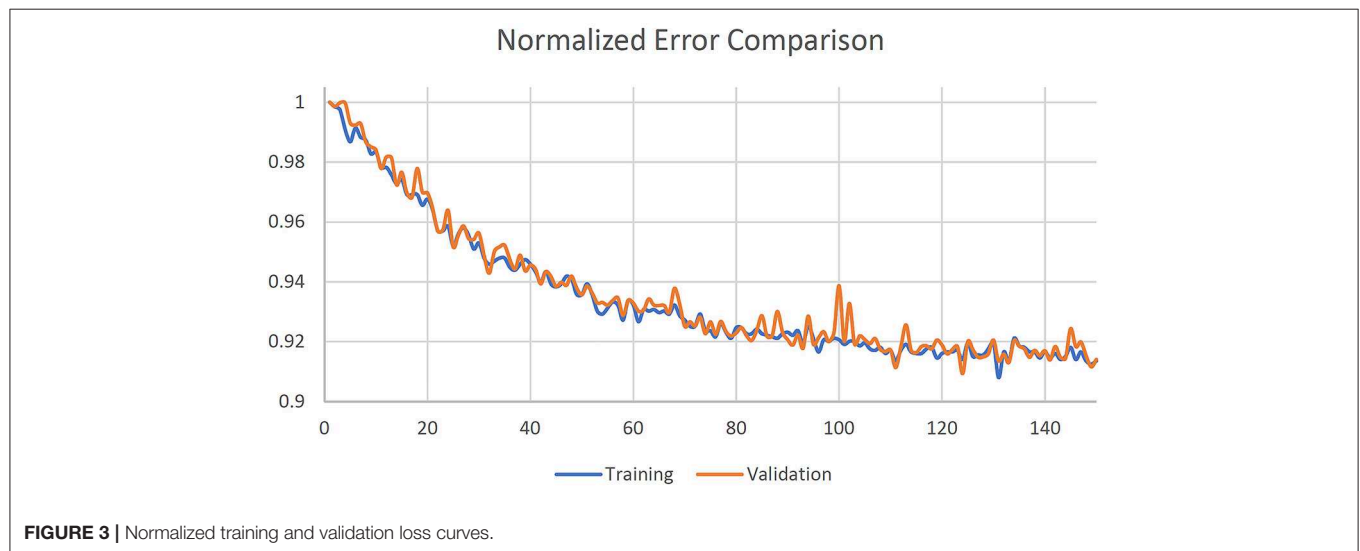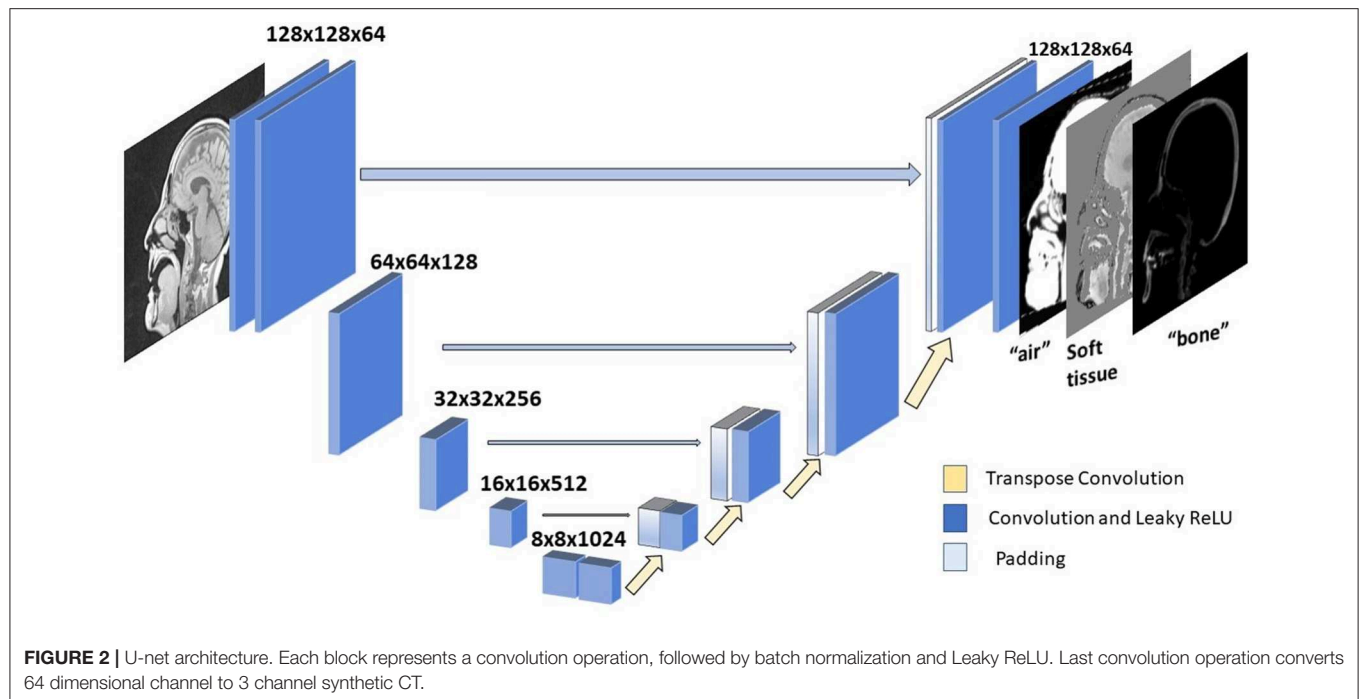
Network Architecture

A U-Net neural network architecture (1, 9) (**Figure 2**) was implemented for translating T1-weighted MRI images into corresponding MRCT images. This network involves a series of downsampling operations that squeezes the input image by factors of two while increasing the number of filters by factors of two. Once this downsampling shrinks the input image 5 times, the same number of upsampling operations successively increase the image dimension by factors of 2, while reducing the number of channels by factors of two. This upsampling is also supported by padding of weights from the corresponding dimension image in the downsampled layer. This allows for

easy flow of gradient information and avoids the "vanishing gradient" problem (10). Each convolution layer is followed by a Batch Normalization (11) and Leaky ReLU (12) activation. We perform downsampling in our convolution operation and upsampling with a transpose convolution operation. The very last convolution layer converts a 64 channel input to a 3-channel output image. We employ Adaptive stochastic gradient descent (Adam) (13) as our optimizer. The U-Net architecture was chosen due to its lower complexity and data requirements than recently used adversarial networks that might overfit the training dataset, since the data sufficiency problem has not been addressed in deep learning based synthetic CT literature.
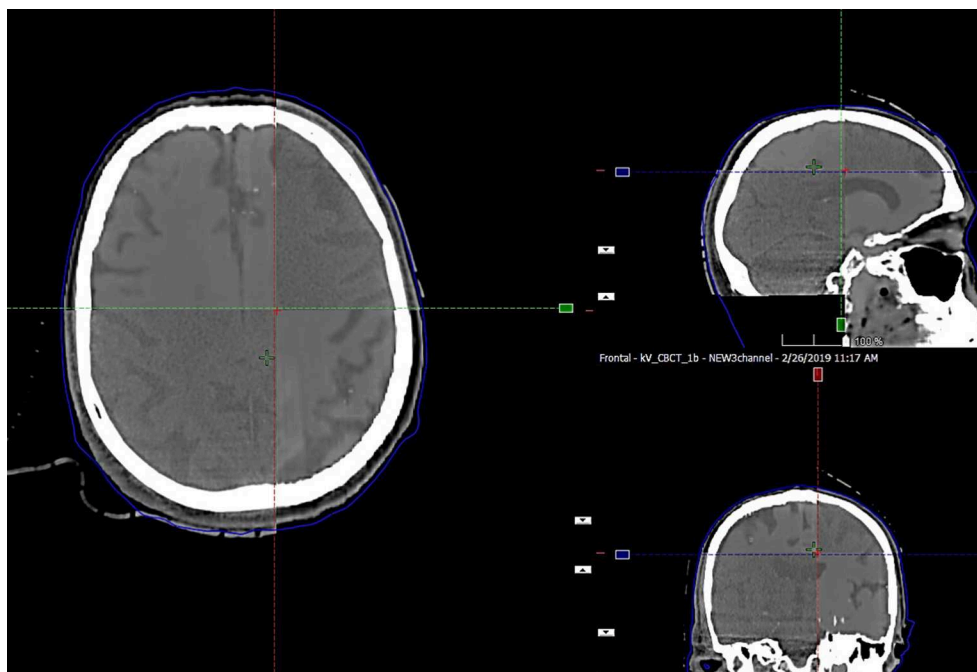
FIGURE 4 | Split window display of MRCT aligned with CBCT for an example patient, demonstrating the preservation of dominant soft tissue interfaces such as major sulci and ventricles as seen in axial, coronal, and sagittal cross sections through the image volumes.

Of the 60 patients, 47 were used for training. To increase the diversity of imaging features on similar anatomic cross sections, sagittal planes were used for training. A total of 9,192 images were used for training. We also implemented data augmentation by random rotation of image by 90 degrees and also by randomly cropping a section of each image for training. To explore the impact of magnitude (and by implication, diversity) of training data, subsets of 10, 20, and 50% of total images were also tested, and the resulting MRCT images from test subjects qualitatively reviewed.

Loss Function

The choice of loss function (L) for our task was a combination of mean absolute error (MAE) and mean squared error (MSE) losses between the CT and MRCT images.

$$MAE\,(X, Y) = \frac{\sum_{i=1}^{n} |X_i - Y_i|}{n} \qquad (1)$$

$$MSE\,(X, Y) = \frac{\sum_{i=1}^{n} (X_i^2 - Y_i^2)}{n} \qquad (2)$$

$$L\,(X, Y) = MAE\,(X, Y) + MSE\,(X, Y) \qquad (3)$$

where X, Y are the images being compared, n is the total number of pixels in the image and X_i represents the ith pixel for image X.

We compare the loss L for each region separately which is backprojected for training the network:

$$L_{tot}(X, Y) = L(X_{air}, Y_{air}) + L(X_{tissue}, Y_{tissue}) + L(X_{bone}, Y_{bone})$$

$$\qquad (4)$$

TABLE 1 | MAE values between MRCT and CT image volumes from the fully sampled network.

Patient	MAE (HU) all voxels	MAE (HU) air	MAE (HU) tissue	MAE (HU) bone
Patient 1	73.80	213.5	13.97	170.02
Patient 2	76.88	227.54	14.39	182.83
Patient 3	88.31	234.27	17.11	217.24
Patient 4	77.53	229.34	17.87	174.08
Patient 5	99.28	271.21	17.8	227.25
Patient 6	69.17	220.9	15.11	169.53
Patient 7	118.14	293.7	26.5	302.54
Patient 8	83.20	218.69	23.51	188.98
Patient 9	58.13	197.96	15.56	154.99
Patient 10	71.45	216.57	17.15	181.51
Patient 11	75.66	200.88	16.7	158.48
Patient 12	89.04	274.47	16.97	214.17
Patient 13	72.66	239.77	16.29	169.13
Mean	81.02	233.75	17.61	193.13
Standard deviation	14.60	28.02	3.41	38.33
Minimum	58.13	197.96	13.97	154.99
Maximum	118.14	293.7	26.5	302.54

Network Training

The U-net was initialized with a normal distribution with mean 0 and standard deviation 0.01. Training was done in mini-batches of 32 random slices. Five-fold cross validation was used, and
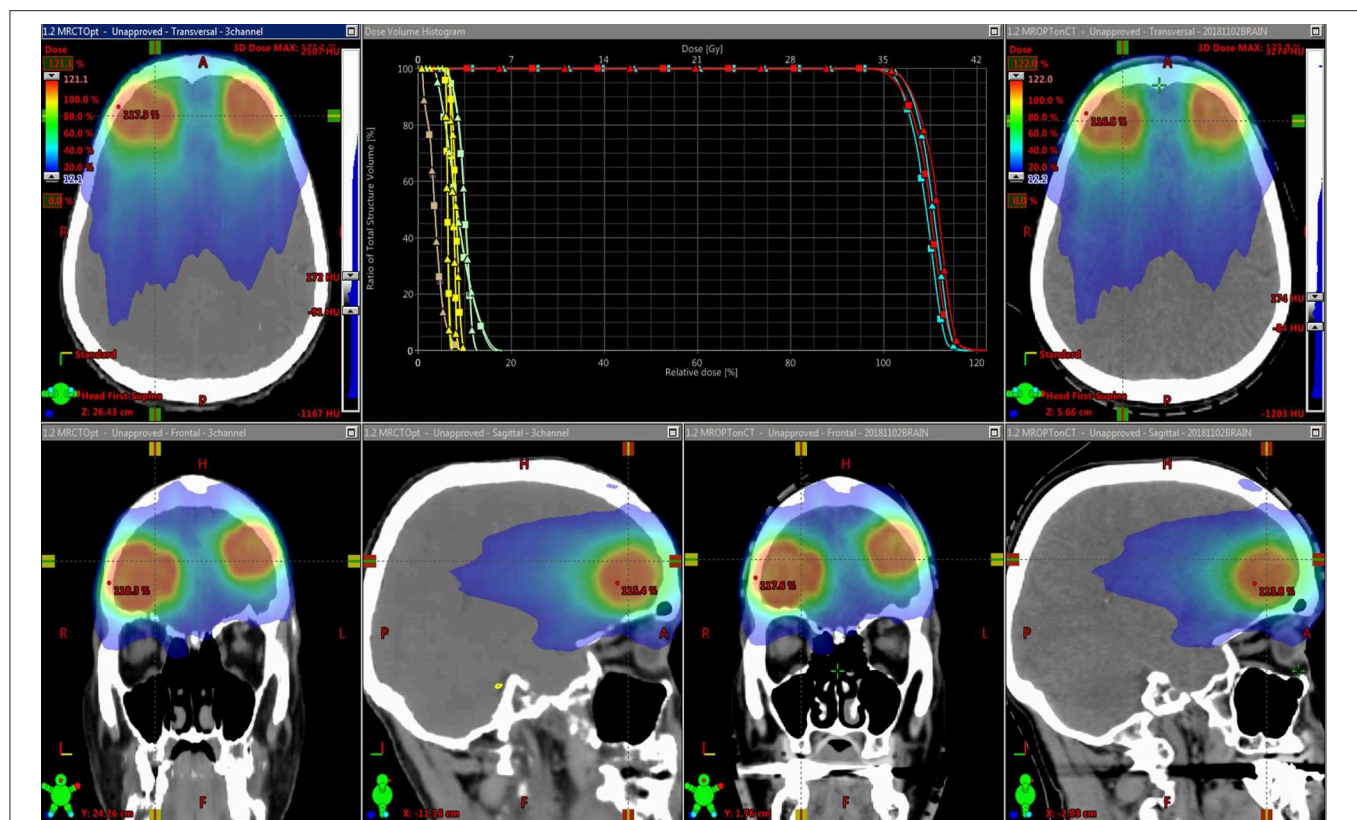
FIGURE 5 | Dose distributions for intensity modulated treatment plans for two targets. The original plan was optimized using the MRCT-derived density grid (left), and the resulting beam fluences were used to recalculate doses on the CT-derived density grid (right). Dose volume histograms (DVHs) for the Brainstem (yellow), Optic chiasm (brown), eyes (green), and two targets (light and dark blue) are shown. Squares represent MRCT plan DVH curves, and triangles come from recalculated plans using CT.

training was stopped after 150 epochs where loss function was observed to reach a plateau as shown in **Figure 3**. The 3 channel images were summed along the channels dimension to generate corresponding MRCT slices.

MRCT Evaluation

MRCT volumes were compared with corresponding CT volumes by various methods. MAE comparisons were done on voxel wise basis, as well as for voxels primarily containing air, soft tissue and bone. These regions were defined within an automatically generated mask that encompassed the head to the inferior border of the skull by using *dipy* (8).

Dosimetric comparisons were made on 11 targets from 7 of the test patients. Using a commercial treatment planning system (Eclipse, Varian Medical Systems, Palo Alto, CA), treatment plans for these radiosurgical targets were generated using the clinical treatment planning directives and with electron density maps derive using MRCT images. The beam fluences generated from these plans were used to recalculate doses by applying the aligned treatment planning CT image volumes as attenuation maps.

For these patients, the MRCT and treatment planning CT image volumes were individually aligned to the Cone Beam CT (CBCT) images acquired for treatment positioning. The

alignment transformations were subsequently applied to the center of the planned treatment targets, and the differences in transformed coordinates compared.

RESULTS

The network training times were 928, 634, 302, and 161 min using 9192, 4096 (50%), 1838 (20%), and 919 (10%) image pairs, respectively, on 2 NVIDIA K40 GPUs. Generation of 3-channel MRCT images took ~1 s.

The preservation of major soft tissue interfaces is demonstrated in example images in **Figure 4**, which further shows support for soft tissue-based alignment between MRCT and CBCT. The MAE for the 13 test patients is reported in **Table 1**. The MAE for all voxels ranged from 58.1–118.1 HU with mean 81.0 HU and standard deviation 14.6 HU. Error values for each of the 3 channels are reported in **Table 1**. Mean MAE values for air, tissue and bone were 234, 22, and 193 HU, respectively.

Figure 5 shows an example treatment plan comparison. The PTV mean dose values had a systematic difference of 2.3% (σ 0.1%) between the plans generated using the MRCT-defined density grids and recalculated using the CT-defined grids. As can
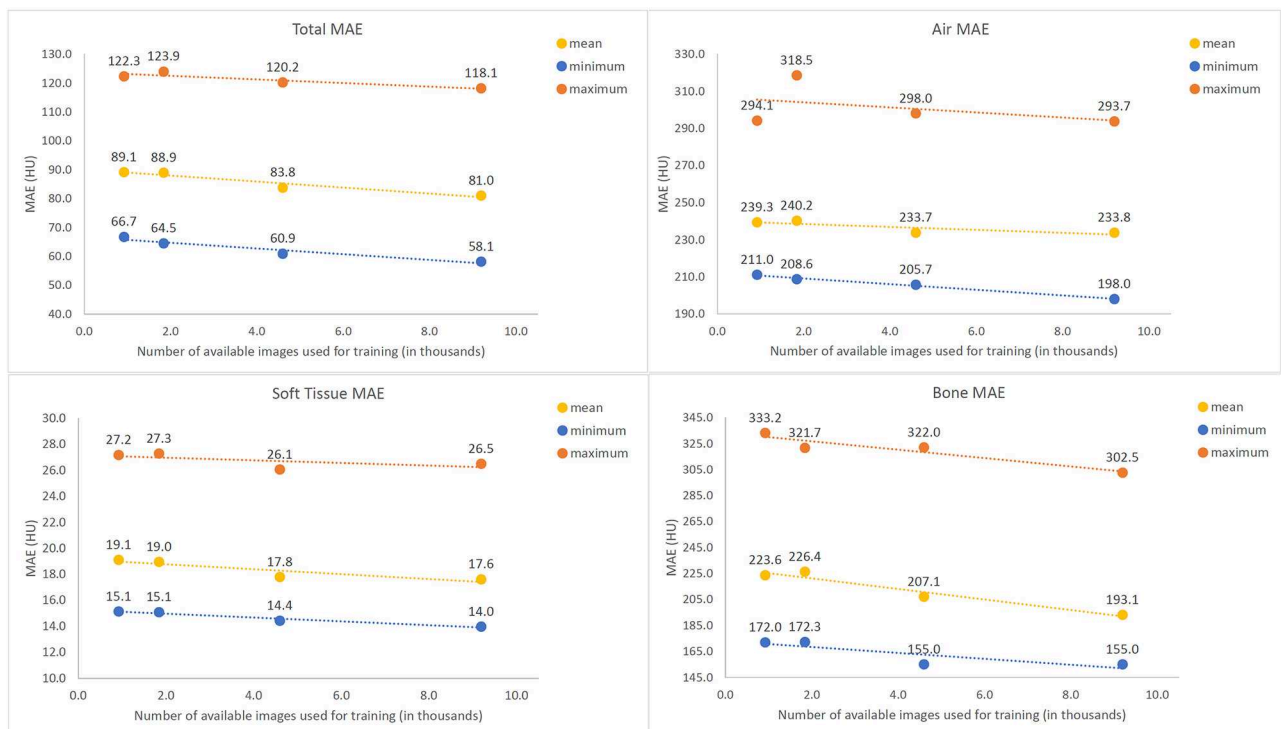
FIGURE 6 | Mean Absolute Error (MAE) values between Synthetic and actual CT image volumes as a function of amount of training data used, tested across 13 tested patients. The maximum (orange), mean (purple), and minimum (blue) values all demonstrate a gradual decreasing trend with increased numbers of training images.

be seen on the images, the MRCT was not trained to reproduce the immobilization device present on the CT, and thus these differences are expected due to the added attenuation of the mask.

Alignment results from CT to CBCT as well as corresponding MRCT-CBCT alignment showed a mean difference of −0.1 (σ 0.2) mm, −0.1 (σ 0.3) mm, and −0.2 (σ 0.3) mm about the left-right, anterior-posterior and cranial-caudal axes, respectively. The range of differences was (−0.3, 0.4), (−0.4, 0.3), and (−0.7, 0.2) mm about the same axes.

Figure 6 shows error for training the U-net with subsampled data for air, tissue and bone, respectively. While the difference between training on 10% (912) and 20% (1824) of available images is not clearly discernable, increasing the number of training images beyond 20% of the total 9192 available samples for training yielded a gradual decrease in average MAE for all three classes intensities, with the most significant trend observed in bony tissues. No plateau was observed, indicating that potential further improvements might be possible with a larger base set of training images.

DISCUSSION

In this report, we suggest an update to the design of neural networks used for generating synthetic CT from MRI. The goal of a 3-channel network is to allow learning of subtle contrast changes in HU values that might not be accurately learned due to the vast range of intensities in CT images. We implemented

the 3-channel structure in a U-Net architecture and saw that soft-tissue contrast can be learned with good precision.

Two previous investigations reported MAE differences between synthetic and actual CT images within soft tissue regions. Emami (4) reported a MAE of 41.85 +/− 8.58 HU in soft tissue using a GAN trained on 15 patients, and Dinkla (6) reported a MAE of 22 +/− 3 HU using a dilated convolutional neural network trained on 52 patients. While we observed error values that are comparable or better (at least in soft tissue) than those reported in these and other investigations (1, 14), we would nonetheless argue that low MAE values are not enough for clinical implementation of MRI-only radiotherapy. Alignment of CT and CBCT is a crucial step that requires correct soft-tissue contrast, and a 3-channel network optimizes for it. We show that the 3-channel output network potentially reduces the problem of faithfully preserving soft tissue features by separately training on CT images within an appropriate intensity range. This process also allows us to scale the loss function to incur heavier penalties separately for errors for each of the different intensity regions.

We observed a gradual trend toward decreasing MAE with increasing amounts of training data. Many prior investigations used far fewer patient images for training than the 47 we had available, and it may be possible that their results are potentially limited by the amount of data available. It is likely that our results are limited by the amount of available data as well, and future investigations will focus on increasing the training data set to incorporate ideally hundreds of patients. A critical question for future investigations will be the elucidation of

the necessary complexity of training information and robust estimation of resulting uncertainty from trained networks. While we chose to focus on a U-Net for training our data in part to limit the potential overfitting due to degeneracy associated with optimizing a network with a larger number of degrees of freedom from limited data, it is also possible that use of a generative adversarial network (GAN) may better reveal the relationship between volume and by inference complexity of training data and accuracy of final results. We will explore the use of GANs as we increase our training data in the future. Of note, a recently published study using a GAN trained on 77 patients with mutual information as a loss function reported an average MAE of 47.2, compared to a MAE of 60.2 when MAE was used as the loss function using the same network (14). While we combined L1 (MAE) and L2 (MSE) in our loss function, we clearly see the value in evaluating loss functions that are better designed to preserve local features, and will consider optimizing such functions in future investigations.

While we chose to train on 2-dimensional images in this investigation, other investigators have shown interesting results using "2.5 dimensional" groupings of multiple images in the same or orthogonal orientations, as well as through training on several 3-dimensional patches. These techniques, as well as nominally fully three-dimensional training, will be part of our future focus.

CONCLUSION

A deep learning approach, consisting of simultaneous training of conversion of T1-weighted MR images to 3 separate intensity regions of corresponding spatially aligned CT images representing HU values typically found in voxels containing mostly air, soft tissue and bone, respectively, was investigated.

Results indicate potential promise in preserving local soft tissue features. Furthermore, the potential advantage of increasing the volume of training data indicated potential further improvements with additional number of patients.

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

AUTHOR CONTRIBUTIONS

DG designed the study, conceived of the network design, collated the training data, and performed all training and related MAE analyses. MK accrued patients to the protocol for analyses, and provided the clinical treatment planning directives used for dose comparisons. KV performed the treatment planning, dose calculations and comparisons, and alignments with cone beam CT images. JB conceived the investigation, oversaw all aspects of the study design and execution, and collated analyses of data. DG and JB wrote the manuscript.

FUNDING

REFERENCES

1. Han X. MR-based synthetic CT generation using a deep convolutional neural network method. *Med Phys*. (2017) 44:1408–19. doi: 10.1002/mp. 12155

2. Johnstone E, Wyatt JJ, Henry AM, Short SC, Sebag-Montefiore D, Murray L, et al. Systematic review of synthetic computed tomography generation methodologies for use in magnetic resonance imaging-only radiation therapy. *Int J Radiat Oncol Biol Phys*. (2018) 100:199–217. doi: 10.1016/j.ijrobp.2017. 08.043

3. Chen S, Qin A, Zhou D, Yan D. Technical Note: U-net-generated synthetic CT images for magnetic resonance imaging-only prostate intensity-modulated radiation therapy treatment planning. *Med Phys*. (2018) 45:5659–65. doi: 10.1002/mp. 13247

4. Emami H, Dong M, Nejad-Davarani SP, Glide-Hurst CK. Generating synthetic CTs from magnetic resonance images using generative adversarial networks. *Med Phys*. (2018) 45:3627–36. doi: 10.1002/mp. 13047

5. Maspero M, Savenije MHFF, Dinkla AM, Seevinck PR, Intven MPW, Jurgenliemk-Schulz IM, et al. Dose evaluation of fast synthetic-CT generation using a generative adversarial network for general pelvis MR-only radiotherapy. *Phys Med Biol*. (2018) 63:185001. doi: 10.1088/1361-6560/ aada6d

6. Dinkla AM, Wolterink JM, Maspero M, Savenije MHF, Verhoeff JJC, Seravalli E, et al. MR-only brain radiation therapy: dosimetric evaluation of synthetic cts generated by a dilated convolutional neural network. *Int J Radiat Oncol*. (2018) 102:801–12. doi: 10.1016/j.ijrobp.2018.05.058

7. Salkeld AL, Hau EKC, Nahar N, Sykes JR, Wang W, Thwaites DI. Changes in brain metastasis during radiosurgical planning. *Int J Radiat Oncol Biol Phys*. (2018) 102:727–33. doi: 10.1016/j.ijrobp.2018.06.021

8. Garyfallidis E, Brett M, Amirbekian B, Rokem A, van der Walt S, Descoteaux M, et al. Dipy, a library for the analysis of diffusion MRI data. *Front Neuroinform*. (2014) 8:8. doi: 10.3389/fninf.2014.00008

9. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells W, Frangi A, editors. *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015. Lecture Notes in Computer Science*. Vol. 9351. Cham: Springer (2015). p. 234–41.

10. Bengio Y, Simard P, Frasconi P. Learning long-term dependencies with gradient descent is difficult. *IEEE Trans Neural Networks*. (1994) 5:157–66. doi: 10.1109/72. 279181

11. Ioffe S, Szegedy C. *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*. (2015). Available online at: http://arxiv.org/abs/1502.03167 (accessed May 9, 2018).

12. Maas AL, Hannun AY, Ng AY. Rectifier nonlinearities improve neural network acoustic models. In: *ICML'13*. Vol 28.

(2013). Available online at: https://www.semanticscholar.org/paper/Rectifier-Nonlinearities-Improve-Neural-Network-Maas/367f2c63a6f6a10b3b64b8729d601e69337ee3cc (accessed April 18, 2019).

13. Kingma DP, Ba J. *Adam: A Method for Stochastic Optimization*. (2014). Available online at: http://arxiv.org/abs/1412.6980 (accessed April 18, 2019).

14. Kazemifar S, McGuire S, Timmerman R, Wardak Z, Nguyen D, Park Y, et al. MRI-only brain radiotherapy: Assessing the dosimetric accuracy of synthetic CT images generated using a deep learning approach. *Radiother Oncol*. (2019) 136:56–63. doi: 10.1016/j.radonc.2019.03.026

"Après Mois, Le Déluge": Preparing for the Coming Data Flood in the MRI-Guided Radiotherapy Era

Kendall J. Kiser [1,2,3], Benjamin D. Smith [3], Jihong Wang [4] and Clifton D. Fuller [3*]

[1] John P. and Kathrine G. McGovern Medical School, University of Texas Health Science Center, Houston, TX, United States, [2] School of Biomedical Informatics, University of Texas Health Science Center, Houston, TX, United States, [3] Department of Radiation Oncology, University of Texas MD Anderson Cancer Center, Houston, TX, United States, [4] Department of Radiation Physics, University of Texas MD Anderson Cancer Center, Houston, TX, United States

Magnetic resonance imaging provides a sea of quantitative and semi-quantitative data. While radiation oncologists already navigate a pool of clinical (semantic) and imaging data, the tide will swell with the advent of hybrid MRI/linear accelerator devices and increasing interest in MRI-guided radiotherapy (MRIgRT), including adaptive MRIgRT. The variety of MR sequences (of greater complexity than the single parameter Hounsfield unit of CT scanning routinely used in radiotherapy), the workflow of adaptive fractionation, and the sheer quantity of daily images acquired are challenges for scaling this technology. Biomedical informatics, which is the science of information in biomedicine, can provide helpful insights for this looming transition. Funneling MRIgRT data into clinically meaningful information streams requires committing to the flow of inter-institutional data accessibility and interoperability initiatives, standardizing MRIgRT dosimetry methods, streamlining MR linear accelerator workflow, and standardizing MRI acquisition and post-processing. This review will attempt to conceptually ford these topics using clinical informatics approaches as a theoretical bridge.

Keywords: MRI, MRI-guided radiotherapy, MR LINAC, informatics, biomedical informatics, clinical informatics, imaging informatics, radiomics

INTRODUCTION

Use of magnetic resonance imaging (MRI) rather than computed tomography (CT) for radiotherapy (RT) planning can be highly desirable because MRI visualizes soft tissues with superior contrast and resolution (1), introduces unique sequences and contrast agents for delineating specific tumors and anatomic subsites (1, 2), and permits daily adaptive radiotherapy (ART) without added CT radiation dose (3–5). MRI-guided ART (MRIgART) machines have advanced from low-field (0.35 Tesla) magnets with Cobalt-60 radiation sources (6) to diagnostic-strength magnetic fields (1.5 Tesla) fully integrated with linear accelerators (7, 8) in <5 years. Over the coming decade, MRI-guided RT (MRIgRT) may change clinical practice paradigms (9). The earliest adopter of MRIgART, Washington University in St. Louis (WUSTL), has already altered its management of breast and abdominal malignancies (10). However, to scale MRIgRT, workflow and standardization challenges that do not exist in CT-guided planning need be resolved.

First, MR scan reproducibility is more complicated than for CT. Consider a T1-weighted scan: pixel intensities are predominately derived from longitudinal relaxation time (T1), an intrinsic tissue property. Nevertheless, proton density (H) and transverse relaxation time (T2) (which are

also intrinsic tissue properties) may greatly influence overall signal intensity (11) depending on repetition time (TR) and echo time (TE) parameters (see Equation 1).

$$S = K \cdot [H] \cdot \left(1 - e^{-\frac{TR}{T1}}\right) \cdot e^{-\frac{TE}{T2}} \qquad (1)$$

These parameters are not standardized across institutions or vendors, so a T1-weighted scan acquired by a given vendor's machine is not necessarily equivalent in terms of observed intensity as one acquired by another manufacturer. Similarly, MRI acquisition suffers from geometric distortions that are model-, vendor-, software-, shim-, and coil-dependent. Proper correction also depends on variable user-driven acquisition parameters (12, 13).

Second, there are more steps in MRIgRT planning than CT-guided planning. MRI does not convey electron density information necessary for standard photon dosimetry, so either (1) MRI data must be registered to CT Hounsfield unit values (14–17), or (2) a synthetic CT (sCT) must be algorithmically generated from MRI (18, 19), or (3) tissue types must be assigned a single, indiscriminate density (18) (**Figure 1**). Additionally, MRIgART fractionation requires far more time of the patient, radiation oncologist, and staff than traditional RT treatment courses.

Third, RT generates seas of imaging data (20, 21) and structured and unstructured clinical data (22–24) that will deepen with multiparametric MRI sequences, unique contrast agents, and radiomics features and MRIgART daily images, contours, and plans (**Figure 2**). At our institution, MRIgRT generates roughly four times as many bytes of data as CT-guided RT (1 Gb per patient per day vs. 250 Mb). *Not all data are fit for clinical decision-making or scientific inquiry.* For example, MRIgART could quantitatively track soft tissue tumor shrinkage, but the results would only be clinically actionable if the segmentation method were systematic and reproduceable. Interpretability and reproducibility of MRI data across institutions and vendors is not a given.

Effective use of "biomedical data, information, and knowledge for scientific inquiry, problem-solving and decision making" formally defines the field of biomedical informatics (BMI) (25). The raison d'être of BMI is to reduce data (which are meaningless symbols) into information (which is data plus meaning), and further into knowledge (which is information that is justifiably believed to be true) (26). This paper considers BMI concepts in the context of scaling MRIgRT (see **Table 1**) and critiques existing literature from the perspective of how it increases information and knowledge to streamline MRIgRT workflow and ensure the consistency and usability of MRIgRT data.

HOW IS BIOMEDICAL INFORMATICS RELEVANT TO MRIGRT CURRENTLY?

MRI is already an established modality for image-guided RT of nasopharynx, brain, spine, liver, pancreas, prostate, and female genital tract cancers (1, 2, 32). In each case, standardization preserves the integrity of critical decision-making information. Consider MRIgRT for prostate cancer

(33, 34). Radiology standards exist for MRI acquisition, interpretation, and reporting (35). These improve reporting among radiologists of varying experience levels (36), lest anatomic delineation suffer poor consistency and patient outcomes comparison data be meaningless. At the MR-CT co-registration step, co-registration between limited field-of-view images is the recommended standard because error is increased when the field-of-view includes the anatomically variable bladder and rectum (37). At the RT planning step, guidelines from the European Society for Radiotherapy and Oncology (ESTRO) (38) and Radiation Therapy Oncology Group (RTOG) (39) standardize MRI-based clinical target and organ-at-risk (OAR) contour volumes. Ostensibly, these steps culminate in more conformal prostate RT, but MRIgRT has proved only modest decreases in OAR toxicity compared to CT-guided RT (40, 41), especially with the development of rectal spacer hydrogel (42). Evaluating data quality and the assumptions used to establish the clinical value of MRIgRT will be a critical BMI task in the coming decade, one that should exploit emerging consumer health informatics approaches.

BMI Considerations for MRIgRT Dosimetry

As already noted, MRIgRT requires either MRI-CT co-registration, sCT generation, or bulk density assignment to calculate tissue radiation dose. MR-only workflows employ either the second or third approach (with the caveat that atlas-based sCT generation techniques employ MRI-CT co-registration to generate an MRI atlas). Improving MR-only RT is strongly motivated by the desire to simplify adaptive workflow for integrated magnetic resonance linear accelerators (MRLs). **Figure 3** exemplifies an imputed electron density map in a patient treated on an MRL. We refer the reader to Table 1 in (34) for an overview of current MRL platforms.

Bulk Density Override and Synthetic CT

Homogenous bulk density override is crude but achieves reasonable dosimetric accuracy if specific structures (e.g., cortical bone) are contoured by a radiation oncologist and separately assigned a unique density (43–45). In contrast, sCT generation by voxel-based or atlas-based methods obviates the need for time-intensive contouring and therefore may be preferred. Johnstone et al. extensively discussed sCT generation methods in a systematic review (18). Many sCT results appear clinically comparable to CT. In the brain, sCT-derived digitally reconstructed radiographs were as geometrically robust as those derived from CT (46). In the prostate, sCT gamma passing rates have been comparable with CT gamma passing rates (median 1%/1 mm pass rate of 100% for almost all regions of interest across 29 patient scans) (47). Nevertheless, MR-only workflow introduces unique BMI considerations. For example, prostate RT plans are more precise with setup to intraprostatic gold or titanium fiducial implants (48), but these are visualized as signal void on conventional MRI sequences and poorly differentiated from calcifications. Maspero et al. (49) reported that 3/48 fiducial implants were imprecisely and inaccurately identified by five radiation technologists when visualized only on MRI. On the other hand, new setup techniques based on MR daily imaging

FIGURE 1 | RT planning with MRI requires assignment of electron density to pixels or tissues to calculate dose, a step that is not part of CT-guided RT planning.

FIGURE 2 | MRI-guided radiotherapy may introduce a deluge of new image sequences, optimization needs, image post-processing needs, contrast agents, prognostic and predictive radiomics features, and adaptive imaging and clinical data.

might obviate the need for fiducials. Thoughtful consideration of parameters like these are needed to ensure not only the safety of the method but the quality and reproducibility of the data. Consensus is also needed to establish the standard metrics by which sCT quality should be gauged (18).

MRI-CT Co-registration

MRIs can be registered to CT rigidly (without warping the MRI) or by a deformation vector field. Deformable registration confers a more concordant result than rigid registration between diagnostic CT and simulation CT (50–53), but recent work from

TABLE 1 | Biomedical informatics concepts.

Concept	Explanation	Example
Data, Information, Knowledge, Wisdom (DIKW) pyramid	Data, information, and knowledge are not synonymous terms. Information is data plus meaning. Knowledge is information plus a justifiable belief in its veracity. Some models also include wisdom as a tier above knowledge (27), but Bernstam et al. (26) omit this.	Radiation oncologist expert interpretation refines dose-volume histogram (DVH) data into information. Statistical analyses then extract and justify knowledge.
The cycle of clinical information flow	Clinical care generates data, which are used in biomedical research, the results of which develop prevention and treatment standards, which are built into clinical protocols, which are built into clinical decision support and order-entry systems, which directly influence clinical care, etc. (28).	MRIgRT outcomes data are recorded in electronic health records, then leveraged in clinical research, which begins to establish the role of MRIgRT, which dictates clinical protocols, which are built into clinical decision support and other health information technology systems, which collaborate with physicians during MRIgRT treatment evaluation and planning, etc.
Data standards	Data standards define and describe "common and repeated use, rules, guidelines or characteristics for activities or their results, aimed at the optimum degree of order" (29).	Digital Imaging and Communications (DICOM), DICOM-RT, Fast Healthcare Interoperable Resources (FHIR), AAPM Task Group 263 consensus nomenclature for dosimetric structures (24).
Interoperability	Interoperability is "the ability of a system or product to work with other systems or products without special effort on the part of the customer" (30). Data standards precede interoperability.	FHIR-conforming electronic health record applications are interoperable between different vendors that also conform with FHIR.
Consumer health informatics	Consumer health informatics is a subfield of biomedical informatics focused on the interactions of patients and consumers with health information systems, catalyzed by mobile technologies and the Internet (31).	Patients log acute and late toxicities during and after MRIgRT in applications built for their phones.

our group did not demonstrate the same advantage between simulation MRI and simulation CT, at least in the head and neck (17). This should not imply that rigid registration is adequately accurate, since we also found that the registration error (whether by deformable or rigid means) may not be within the target tolerance recommended by the American Association of Physicists in Medicine (AAPM) Task Group 132 (Dice similarity coefficient > 0.8) (54). Perhaps registration was poor because of MR geometric distortion, or perhaps because

not all OARs are clearly delineated on both CT and MRI. Regardless, this informs our view that sCT may be preferred to CT-MRI co-registration for RT dose deposition calculation, pending needed standardizations as discussed above.

WORKFLOW CONSIDERATIONS FOR INTEGRATED MR LINEAR ACCELERATORS

The "holy grail" of MRL RT is to see the target at setup, adapt the plan as needed, and gate by watching anatomic movement while the beam is on. The experience of the Department of Radiation Oncology at WUSTL, which introduced the first 0.35 T tri-cobalt-60 MRIgRT system (ViewRay, Oakwood Village, OH, USA) in the USA (6), provides great insight into adaptive MRL clinical informatics challenges. In a Phase I trial intended to demonstrate the temporal feasibility of MRI-guided stereotactic body radiation (SBRT), median on-table time per fraction was 79 min and consisted of MR set-up, physician arrival, patient localization, re-segmentation, re-planning, quality assurance (QA), and beam on-time (3). Almost all fractions (81/97) were adapted based on the patient's anatomy-of-day to avoid irradiating OARs. Despite fear that patients would not tolerate fractions longer than 80 min, all 20 patients completed their treatments as prescribed.

MRL RT has evolved into a dominant indication for abdominal and breast cancers at WUSTL, primarily because motion gating and daily adaptation prevent OAR dose constraint violations (10). The MRL has also prevented violations in hypofractionated lung tumor stereotactic radiotherapy, and enabled adaptive GTV reductions by as much as 65% (55). However, adaptation remains time-intensive. Current systems require physician attendance during every fraction (56), which would not be sustainable at sites that lack sufficient physician and support staff. Three studies from the University of Alberta examined whether automated ROI segmentation can decrease the burden on physician time. In the first, a pulse-coupled neural network (PCNN) was developed to segment lung tumors in the context of adaptive MRL RT (57). The PCNN achieved a strong Dice Similarity Index (DSI) of 0.87–0.92, but it required training on a unique dataset of manually-generated contours per patient. A follow-up study improved DSI (58) with a pre-segmentation deformable registration methodology, but still required a physician to segment lung tumor across multiple image frames. In the third study, DSI and other conformality metrics improved using a fully convolutional neural network (FCNN), but the FCNN still needed to be trained on 30 manual contours per patient. While these studies demonstrate that automated segmentations of lung tumors for MRL RT can achieve high fidelity, they may not hasten adaptive, online MRL workflow. In contrast, a WUSTL novel tri-convolutional neural network architecture capable of segmenting liver, kidneys, stomach, bowel, and duodenum did reduce manual segmentation time by 75% at WUSTL (59).

Intra and inter-observer variation in segmentation quality has been documented using many imaging modalities in pelvic
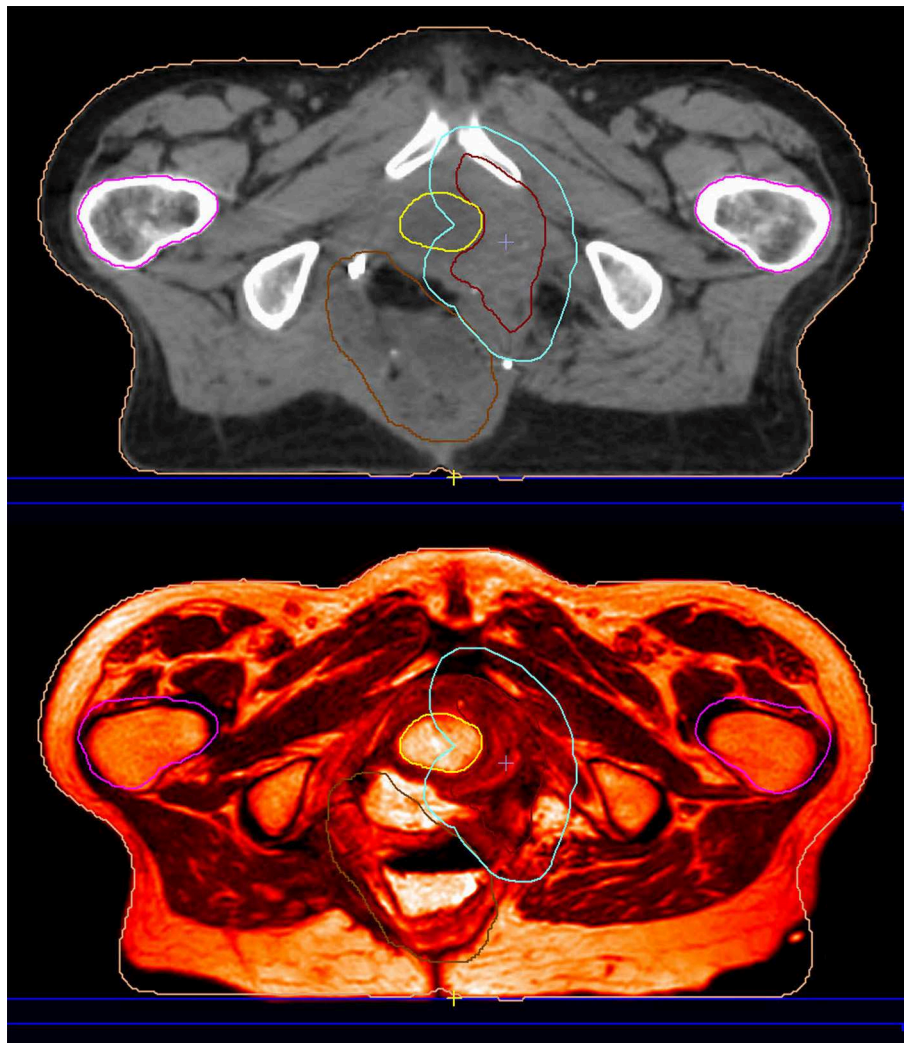
FIGURE 3 | This patient was reirradiated for a rectal adenocarcinoma recurrence on a hybrid magnetic resonance-guided linear accelerator. The planning image **(top)** is an electron density imputation based on the simulation MRI **(bottom)**.

(60, 61), lung (62), breast (63), head-and-neck (60, 64), and brain (65) RT planning. In the specific context of MRL for lung stereotactic body RT, Wee et al. (66) found no significant intra or inter-observer variation in manual segmentations of images acquired on a 0.35 T MRL. However, only two radiation oncologist observers were compared, for only one ROI (gross tumor volume) (66), limiting the generalizability of the study conclusion.

To hasten MRL re-planning, WUSTL simplified the number of planning objectives by grouping OAR structures into a single structure (67). This both increased PTV coverage and simplified re-planning by reducing the computational burden of satisfying a greater number of competing objectives. This work was specific to pancreatic cancer planning objectives, but the approach may be amenable to re-planning for other sites.

Intrafraction motion management/gating is a hotly anticipated MRL advantage. Han et al. (68) applied 3D-Rotating Cartesian K-space MRI (4D-ROCK-MRI) in an MRL RT workflow to improve lung tumor motion tracking. 4D-ROCK-MRI improved image quality and motion tracking and decreased lung cancer GTV variability compared with 4D-CT, which suffers from 2D-slice "stitching" artifact. The authors reason that it might capture motion better than 2D-CINE MRI because it acquires data over a 7 min interval, while the latter screens less than a minute of data. Cusumano et al. (69) compared 4D-CT and 2D-cine MR motion data acquired at the time of simulation with complete 2D-cine MR datasets acquired over entire MRIgRT treatment courses. Simulation 2D-cine MR appeared better than simulation 4D-CT, though not significantly. Patients with large motion amplitudes at the time of simulation tended to have more variable amplitudes throughout their treatment course, but even targets with steady amplitudes frequently drifted from the motion trajectory calculated at simulation. Drift was as severe as 1.6 cm craniocaudally and 1.2 cm anteroposteriorly,

which highlights the importance of continual IGRT monitoring throughout treatment. Palacios et al. (70) tracked adrenal metastases and discovered that one-third of the time anatomy positioning violated OAR or target dose constraints. van Sornsen de Koste et al. (71) followed lung, adrenal, and pancreatic tumor GTVs with 2D-cine MRI. In 90% of cases these tumors oscillated no more than 6 mm anteroposteriorly and 9 mm craniocaudally. Mean coverage was better than 94% of the GTV volume for all three tumor types (coverage was defined as a 3 mm isotropic GTV expansion).

ART discussions encompass many other considerations beyond the scope of this paper, but we highlight one more: both US commercial hybrid MRL systems use what Heukelom et al. (5) define as "serial ART" (i.e., daily images are registered to a planning scan serially without interval dose accumulation) but can conceivably be utilized for "triggered ART" (when fixed interval re-planning offline occurs) or "cascade ART" (when serial deformed dose is integrated from prior treatments). Consequently, a need for new ways of visualizing and reporting dose and morphometric alterations will soon arise. Centers that lack MRL machines but are interested in MRIgRT for abdominal cancers may find the workflow outlined in Heerkens et al. (72) informative. This phase I trial demonstrated a favorable toxicity profile (no treatment-attributable grade 3 acute or late toxicities) in 20 patients with unresectable pancreatic cancer who received 24 Gy/3 fx SBRT planned with multiparametric MRI sequences and sagittal cine MRI.

RADIOMICS STANDARDIZATION: A PRESSING INFORMATICS CHALLENGE

The use of imaging biomarkers for diagnosis and prognosis is the field known as radiomics (73), or radiogenomics if the biomarkers are both radiomic and genomic (74, 75). MRI radiomics features have predicted tumor histopathology (76, 77), improved region-of-interest (ROI) auto-segmentation (78), automated radiotherapy planning (79) and predicted outcomes (e.g., survival, toxicity) (80–82). However, standardization of radiomics feature parameters is needed across radiation oncology, radiology, and nuclear medicine disciplines (83). In a systematic review of MRI radiomics applications, Jethanandani et al. concluded that MRI radiomics studies suffer from lack of standardization at multiple stages of image acquisition and processing, including MRI scanner sequence, scanner vendor, and scan acquisition parameters. There is currently no way to reliably compare between MRI radiomics studies (84). MRI has not nearly enjoyed the attention given to CT and PET radiomics standardization. Traverso et al. (85) systematically reviewed studies that assessed the repeatability and reproducibility radiomics features, finding only 1/41 papers (86) that investigated MRI.

Radiomics models should be commissioned from their ideation with a clinical decision support use case in mind (87). Studies designed to maximize the likelihood of a statistically significant finding at the expense of clinical generalizability ignore that practical implementation is a greater obstacle than

discovery. To illustrate, one study that discriminated triple-negative from other breast cancer types using radiomics features ostensibly aspires to be a diagnostic alternative to biopsy (88), but would need to be less expensive yet no less accurate—a steep challenge.

Radiomics feature stability should be benchmarked on public, multi-institutional datasets (85, 89). For example, Bakas et al. (90) publicly provided radiomics features manually extracted from neuroradiologist segmentations of glioblastomas and low-grade gliomas for benchmarking future studies of these cancers. Stability should be benchmarked per anatomic site, since features that are repeatable and reproduceable at one site may degrade in the context of another.

INITIATIVES FOR FAIRER DATA

Inter-institutional findable, accessible, interoperable, reusable (FAIR) (91) and high-quality data is essential for establishing the clinical value of MRIgRT. However, political, financial, and legal obstacles silo data within institutions (92) and ethical questions surrounding health data analytics, particularly by tech institutions currently not subject to the same patient privacy laws as healthcare institutions, are unresolved (93, 94). The need for FAIR data is not exclusive to MRIgRT: FAIR data are critical for achieving the vision for machine learning in healthcare widely (95–97).

A recent AAPM council observed that RT data increase as cancer patients survive longer and genomic data move toward mainstream clinical use (98). The council predicted, "Whereas success in medical research in the past has favored very large single institutions that can develop a critical mass of knowledge and resources in close physical proximity, *diffuse networks of institutions able to generate and share information will have an advantage in the future*" (emphasis added). We now conclude with a discussion of two emerging initiatives for FAIRer data: "distributed learning," a method for inter-institutional machine learning, and Fast Healthcare Interoperability Resources (FHIR, pronounced "fire") a healthcare data standard.

Distributed Learning
Distributed learning refers to training machine learning models on multi-institutional data without sharing the data (99–101). The key is that the statistical weights and parameters of the machine learning model travel between institutions, not the data. Distributed learning is an option method for generating statistical models for emerging technologies, such as MRLs (7, 8). Distributed learning is possible between horizontally-partitioned data (same features, different patients) or vertically-partitioned data (different features, same patients) (101, 102).

FHIR
Conceived in 2014, FHIR is a specification for health data formatting (i.e., XML and JSON) and messaging (i.e., RESTful application programming interfaces). FHIR-conforming data are retrievable between health information technology softwares (103), and FHIR may soon be a mandated EHR specification (104). FHIR provides a standard for storing and querying

radiotherapy data objects independent of vendor, such as total dose, or DICOM-RT structure sets. Conformity with FHIR also makes it possible to build applications that integrate health information technologies. For example, Substitutable Medical Applications and Reusable Technologies (SMART) is an EHR app platform built on FHIR (105). SMART delineates authorization, authentication, and user interface specifications for FHIR-conforming apps. Because RT treatment planning and information systems are usually separate from EHRs (98, 106), initiatives like SMART on FHIR envision a future where it is possible to build RT task-specific apps into EHRs (107). Open-source, FHIR-conforming applications may be one platform for scaling MRIgRT software.

AUTHOR CONTRIBUTIONS

FUNDING

REFERENCES

1. Chandarana H, Wang H, Tijssen RHN, Das IJ. Emerging role of MRI in radiation therapy. *J Magn Reson Imaging*. (2018) 48:1468–78. doi: 10.1002/jmri.26271

2. McWilliam A, Rowland B, van Herk M. The challenges of using MRI during Radiotherapy. *Clin Oncol*. (2018) 30:680–5. doi: 10.1016/j.clon.2018.08.004

3. Henke L, Kashani R, Robinson C, Curcuru A, DeWees T, Bradley J, et al. Phase I trial of stereotactic MR-guided online adaptive radiation therapy (SMART) for the treatment of oligometastatic or unresectable primary malignancies of the abdomen. *Radiother Oncol*. (2018) 126:519–26. doi: 10.1016/j.radonc.2017.11.032

4. Lim-Reinders S, Keller BM, Al-Ward S, Sahgal A, Kim A. Online adaptive radiation therapy. *Int J Radiat Oncol Biol Phys*. (2017) 99:994–1003. doi: 10.1016/j.ijrobp.2017.04.023

5. Heukelom J, Fuller CD. Head and neck cancer Adaptive Radiation Therapy (ART): conceptual considerations for the informed clinician. *Semin Radiat Oncol*. (2019) 29:258–73. doi: 10.1016/j.semradonc.2019.02.008

6. Mutic S, Dempsey JF. The ViewRay system: magnetic resonance–guided and controlled radiotherapy. *Semin Radiat Oncol*. (2014) 24:196–9. doi: 10.1016/j.semradonc.2014.02.008

7. Raaymakers BW, Jürgenliemk-Schulz IM, Bol GH, Glitzner M, Kotte ANTJ, van Asselen B, et al. First patients treated with a 1.5 T MRI-LINAC: clinical proof of concept of a high-precision, high-field MRI guided radiotherapy treatment. *Phys Med Biol*. (2017) 62:L41–50. doi: 10.1088/1361-6560/aa9517

8. Wendler R. *Cutting-Edge MRI-Guided Radiation Therapy Provides Real-Time View of Tumors*. Houston, TX: The University of Texas MD Anderson Cancer Center (2019). Available online at:https://www.mdanderson.org/publications/cancer-frontline/2019/collaboration-leads-to-fda-approval-of-cutting-edge-radiation-tr.html

9. Bayouth JE, Low DA, Zaidi H. MRI-linac systems will replace conventional IGRT systems within 15 years. *Med Phys*. (2019) 46:3753–6. doi: 10.1002/mp.13657

10. Henke LE, Contreras JA, Green OL, Cai B, Kim H, Roach MC, et al. Magnetic resonance image-guided radiotherapy (MRIgRT): A 4.5-year clinical experience. *Clin Oncol*. (2018) 30:720–7. doi: 10.1016/j.clon.2018.08.010

11. Nitz WR, Reimer P. Contrast mechanisms in MR imaging. *Eur Radiol*. (1999) 9:1032–46. doi: 10.1007/s003300050789

12. Weygand J, Fuller CD, Ibbott GS, Mohamed AS, Ding Y, Yang J, et al. Spatial precision in magnetic resonance imaging-guided radiation therapy: the role of geometric distortion. *Int J Radiat Oncol Biol Phys*. (2016) 95:1304–16. doi: 10.1016/j.ijrobp.2016.02.059

13. Mohamed ASR, Hansen C, Weygand J, Ding Y, Frank SJ, Rosenthal DI, et al. Prospective analysis of *in vivo* landmark point-based MRI geometric distortion in head and neck cancer patients scanned in immobilized radiation treatment position: Results of a prospective quality assurance protocol. *Clin Transl Radiat Oncol*. (2017) 7:13–9. doi: 10.1016/j.ctro.2017.09.003

14. Wang X, Li L, Hu C, Qiu J, Xu Z, Feng Y. A comparative study of three CT and MRI registration algorithms in nasopharyngeal carcinoma. *J Appl Clin Med Phys*. (2009) 10:2906. doi: 10.1120/jacmp.v10i2.2906

15. Fortunati V, Verhaart RF, Angeloni F, Van Der Lugt A, Niessen WJ, Veenland JF, et al. Feasibility of multimodal deformable registration for head and neck tumor treatment planning. *Int J Radiat Oncol Biol Phys*. (2014) 90:85–93. doi: 10.1016/j.ijrobp.2014.05.027

16. Webster GJ, Kilgallon JE, Ho KF, Rowbottom CG, Slevin NJ, Mackay RI. A novel imaging technique for fusion of high-quality immobilised MR images of the head and neck with CT scans for radiotherapy target delineation. *Br J Radiol*. (2009) 82:497–503. doi: 10.1259/bjr/50709041

17. Kiser KJ, Meheissen MAM, Mohamed ASR, Kamal M, Ng SP, Elhalawani H, et al. Prospective quantitative quality assurance and deformation estimation of MRI-CT image registration in simulation of head and neck radiotherapy patients. *Clin Transl Radiat Oncol*. (2019) 18:120–7. doi: 10.1016/j.ctro.2019.04.018

18. Johnstone E, Wyatt JJ, Henry AM, Short SC, Sebag-Montefiore D, Murray L, et al. Systematic review of synthetic computed tomography generation methodologies for use in magnetic resonance imaging-only

radiation therapy. *Int J Radiat Oncol Biol Phys.* (2018) 100:199–217. doi: 10.1016/j.ijrobp.2017.08.043

19. Edmund JM, Nyholm T. A review of substitute CT generation for MRI-only radiation therapy. *Radiat Oncol.* (2017) 12:28–42. doi: 10.1186/s13014-016-0747-y

20. Law MY, Liu B, Chan LW. Informatics in radiology: DICOM-RT-based electronic patient record information system for radiation therapy. *Radiographics.* (2009) 29:961–72. doi: 10.1148/rg.294085073

21. Thariat J, Marcy PY, Lacout A, Ramus L, Girinsky T, Pointreau Y, et al. Radiotherapy and radiology: joint efforts for modern radiation planning and practice. *Diagn Interv Imaging.* (2012) 93:342–50. doi: 10.1016/j.diii.2012.02.004

22. Kreimeyer K, Foster M, Pandey A, Arya N, Halford G, Jones SF, et al. Natural language processing systems for capturing and standardizing unstructured clinical information: a systematic review. *J Biomed Inform.* (2017) 73:14–29. doi: 10.1016/j.jbi.2017.07.012

23. Dreisbach C, Koleck TA, Bourne PE, Bakken S. A systematic review of natural language processing and text mining of symptoms from electronic patient-authored text data. *Int J Med Inform.* (2019) 125:37–46. doi: 10.1016/j.ijmedinf.2019.02.008

24. Mayo CS, Moran JM, Bosch W, Xiao Y, McNutt T, Popple R, et al. American association of physicists in medicine task group 263: standardizing nomenclatures in radiation oncology. *Int J Radiat Oncol Biol Phys.* (2018) 100:1057–66. doi: 10.1016/j.ijrobp.2017.12.013

25. Kulikowski CA, Shortliffe EH, Currie LM, Elkin PL, Hunter LE, Johnson TR, et al. AMIA Board white paper: definition of biomedical informatics and specification of core competencies for graduate education in the discipline. *J Am Med Inform Assoc.* (2012) 19:931–8. doi: 10.1136/amiajnl-2012-001053

26. Bernstam EV, Smith JW, Johnson TR. What is biomedical informatics? *J Biomed Inform.* (2010) 43:104–10. doi: 10.1016/j.jbi.2009.08.006

27. Rowley J. The wisdom hierarchy: representations of the DIKW hierarchy. *J Inform Sci.* (2007) 33:163–80. doi: 10.1177/0165551506070706

28. Shortliffe EH, Blois MS. The computer meets medicine and biology: emergence of a discipline. In: Shortliffe EH, Cimino JJ, editors. *Biomedical Informatics: Computer Applications in Health Care and Biomedicine, 3rd ed.* New York, NY: Springer Science + Business Media, LLC (2006). doi: 10.1007/0-387-36278-9-1

29. Anonymous. *ISO/IEC Guide 2: Standardization and Related Activities – General Vocabulary.* Geneva: International Organization for Standardization (2004).

30. Anonymous. *Standards Glossary: Interoperability: Institute of Electrical and Electronics Engineers.* (2016). Available online at: https://www.standardsuniversity.org/article/standards-glossary/ (accessed June 27, 2019).

31. Hersh WR. Standards and Interoperability. *Health Informatics: Practical Guide,* 7th ed. Informatics Education. (2018). Available online at: lulu.com (accessed June 27, 2019).

32. Fraass BA, Eisbruch A, Feng M. Intensity-modulated and image-guided radiation therapy. In: Gunderson LL, Tepper JE, editors. *Clinical Radiation Oncology, 4th ed.* Philadelphia, PA: Elsevier (2016). p. 294–324. doi: 10.1016/B978-0-323-24098-7.00016-2

33. Gkika E, Oehlke O, Bunea H, Wiedenmann N, Adebahr S, Nestle U, et al. Biological imaging for individualized therapy in radiation oncology: part II medical and clinical aspects. *Future Oncol.* (2018) 14:751–69. doi: 10.2217/fon-2017-0465

34. Pathmanathan AU, van As NJ, Kerkmeijer LGW, Christodouleas J, Lawton CAF, Vesprini D, et al. Magnetic resonance imaging-guided adaptive radiation therapy: a "Game Changer" for prostate treatment? *Int J Radiat Oncol Biol Phys.* (2018) 100:361–73. doi: 10.1016/j.ijrobp.2017.10.020

35. Weinreb JC, Barentsz JO, Choyke PL, Cornud F, Haider MA, Macura KJ, et al. PI-RADS prostate imaging - reporting and data system: 2015, version 2. *Eur Urol.* (2016) 69:16–40. doi: 10.1016/j.eururo.2015.08.052

36. Greer MD, Shih JH, Lay N, Barrett T, Bittencourt L, Borofsky S, et al. Interreader variability of prostate imaging reporting and data system version 2 in detecting and assessing prostate cancer lesions at prostate MRI. *AJR Am J Roentgenol.* (2019) 212:1–8. doi: 10.2214/AJR.18.20536

37. Menard C, Paulson E, Nyholm T, McLaughlin P, Liney G, Dirix P, et al. Role of prostate MR imaging in radiation oncology. *Radiol Clin North Am.* (2018) 56:319–25. doi: 10.1016/j.rcl.2017.10.012

38. Salembier C, Villeirs G, De Bari B, Hoskin P, Pieters BR, Van Vulpen M, et al. ESTRO ACROP consensus guideline on CT- and MRI-based target volume delineation for primary radiation therapy of localized prostate cancer. *Radiother Oncol.* (2018) 127:49–61. doi: 10.1016/j.radonc.2018.01.014

39. Contouring atlases: Radiation Therapy Oncology Group (RTOG). *RTOG Contouring Consensus Guidelines for Various Anatomic Sites.* (2019). Available online at: https://www.rtog.org/CoreLab/ContouringAtlases.aspx (accessed April 4, 2019).

40. Ali AN, Rossi PJ, Godette KD, Martin D, Liauw S, Vijayakumar S, et al. Impact of magnetic resonance imaging on computed tomography-based treatment planning and acute toxicity for prostate cancer patients treated with intensity modulated radiation therapy. *Pract Radiat Oncol.* (2013) 3:e1–9. doi: 10.1016/j.prro.2012.04.005

41. Sander L, Langkilde NC, Holmberg M, Carl J. MRI target delineation may reduce long-term toxicity after prostate radiotherapy. *Acta Oncol.* (2014) 53:809–14. doi: 10.3109/0284186X.2013.865077

42. Hamstra DA, Mariados N, Sylvester J, Shah D, Karsh L, Hudes R, et al. Continued benefit to rectal separation for prostate radiation therapy: final results of a phase III trial. *Int J Radiat Oncol Biol Phys.* (2017) 97:976–85. doi: 10.1016/j.ijrobp.2016.12.024

43. Karotki A, Mah K, Meijer G, Meltsner M. Comparison of bulk electron density and voxel-based electron density treatment planning. *J Appl Clin Med Phys.* (2011) 12:3522. doi: 10.1120/jacmp.v12i4.3522

44. Jonsson JH, Karlsson MG, Karlsson M, Nyholm T. Treatment planning using MRI data: an analysis of the dose calculation accuracy for different treatment regions. *Radiat Oncol.* (2010) 5:62. doi: 10.1186/1748-717X-5-62

45. Lambert J, Greer PB, Menk F, Patterson J, Parker J, Dahl K, et al. MRI-guided prostate radiation therapy planning: Investigation of dosimetric accuracy of MRI-based dose planning. *Radiother Oncol.* (2011) 98:330–4. doi: 10.1016/j.radonc.2011.01.012

46. Morris ED, Price RG, Kim J, Schultz L, Siddiqui MS, Chetty I, et al. Using synthetic CT for partial brain radiation therapy: impact on image guidance. *Pract Radiat Oncol.* (2018) 8:342–50. doi: 10.1016/j.prro.2018.04.001

47. Christiansen RL, Jensen HR, Brink C. Magnetic resonance only workflow and validation of dose calculations for radiotherapy of prostate cancer. *Acta Oncol.* (2017) 56:787–91. doi: 10.1080/0284186X.2017.1290275

48. Zaorsky NG, Showalter TN, Ezzell GA, Nguyen PL, Assimos DG, D'Amico AV, et al. ACR Appropriateness Criteria for external beam radiation therapy treatment planning for clinically localized prostate cancer, part II of II. *Adv Radiat Oncol.* (2017) 2:437–54. doi: 10.1016/j.adro.2017.03.003

49. Maspero M, Seevinck PR, Willems NJW, Sikkes GG, de Kogel GJ, de Boer HCJ, et al. Evaluation of gold fiducial marker manual localisation for magnetic resonance-only prostate radiotherapy. *Radiat Oncol.* (2018) 13:105. doi: 10.1186/s13014-018-1029-7

50. Mohamed AS, Ruangskul MN, Awan MJ, Baron CA, Kalpathy-Cramer J, Castillo R, et al. Quality assurance assessment of diagnostic and radiation therapy-simulation CT image registration for head and neck radiation therapy: anatomic region of interest-based comparison of rigid and deformable algorithms. *Radiology.* (2015) 274:752–63. doi: 10.1148/radiol.14132871

51. Guy CL, Weiss E, Che S, Jan N, Zhao S, Rosu-Bubulac M. Evaluation of image registration accuracy for tumor and organs at risk in the thorax for compliance with TG 132 recommendations. *Adv Radiat Oncol.* (2019) 4:177–85. doi: 10.1016/j.adro.2018.08.023

52. Sarudis S, Karlsson A, Bibac D, Nyman J, Back A. Evaluation of deformable image registration accuracy for CT images of the thorax region. *Phys Med.* (2019) 57:191–9. doi: 10.1016/j.ejmp.2018.12.030

53. Motegi K, Tachibana H, Motegi A, Hotta K, Baba H, Akimoto T. Usefulness of hybrid deformable image registration algorithms in prostate radiation therapy. *J Appl Clin Med Phys.* (2019) 20:229–36. doi: 10.1002/acm2.12515

54. Brock KK, Mutic S, McNutt TR, Li H, Kessler ML. Use of image registration and fusion algorithms and techniques in radiotherapy: report of the AAPM radiation therapy committee task Group No. 132. *Med Phys.* (2017) 44:e43–76. doi: 10.1002/mp.12256

55. Henke LE, Kashani R, Hilliard J, DeWees TA, Curcuru A, Przybysz D, et al. *In silico* trial of mr-guided midtreatment adaptive planning for hypofractionated stereotactic radiation therapy in centrally

located thoracic tumors. *Int J Radiat Oncol Biol Phys*. (2018) 102:987–95. doi: 10.1016/j.ijrobp.2018.06.022

56. Kashani R, Olsen JR. Magnetic resonance imaging for target delineation and daily treatment modification. *Semin Radiat Oncol*. (2018) 28:178–84. doi: 10.1016/j.semradonc.2018.02.002

57. Yun J, Yip E, Gabos Z, Wachowicz K, Rathee S, Fallone BG. Neural-network based autocontouring algorithm for intrafractional lung-tumor tracking using Linac-MR. *Med Phys*. (2015) 42:2296–310. doi: 10.1118/1.4916657

58. Tahmasebi N, Boulanger P, Yun J, Fallone BG, Punithakumar K. Tracking tumor boundary using point correspondence for adaptive radio therapy. *Comput Methods Programs Biomed*. (2018) 165:187–95. doi: 10.1016/j.cmpb.2018.08.002

59. Fu Y, Mazur TR, Wu X, Liu S, Chang X, Lu Y, et al. A novel MRI segmentation method using CNN -based correction network for MRI -guided adaptive radiotherapy. *Med Phys*. (2018) 45:5129–37. doi: 10.1002/mp.13221

60. Riegel AC, Antone JG, Zhang H, Jain P, Raince J, Rea A, et al. Deformable image registration and interobserver variation in contour propagation for radiation therapy planning. *J Appl Clin Med Phys*. (2016) 17:347–57. doi: 10.1120/jacmp.v17i3.6110

61. Hellebust TP, Tanderup K, Lervag C, Fidarova E, Berger D, Malinen E, et al. Dosimetric impact of interobserver variability in MRI-based delineation for cervical cancer brachytherapy. *Radiother Oncol*. (2013) 107:13–9. doi: 10.1016/j.radonc.2012.12.017

62. Persson GF, Nygaard DE, Hollensen C, Munck af Rosenschold P, Mouritsen LS, Due AK, et al. Interobserver delineation variation in lung tumour stereotactic body radiotherapy. *Br J Radiol*. (2012) 85:e654–60. doi: 10.1259/bjr/76424694

63. Yang J, Woodward WA, Reed VK, Strom EA, Perkins GH, Tereffe W, et al. Statistical modeling approach to quantitative analysis of interobserver variability in breast contouring. *Int J Radiat Oncol Biol Phys*. (2014) 89:214–21. doi: 10.1016/j.ijrobp.2014.01.010

64. Mukesh M, Benson R, Jena R, Hoole A, Roques T, Scrase C, et al. Interobserver variation in clinical target volume and organs at risk segmentation in post-parotidectomy radiotherapy: can segmentation protocols help? *Br J Radiol*. (2012) 85:e530–6. doi: 10.1259/bjr/66693547

65. Stanley J, Dunscombe P, Lau H, Burns P, Lim G, Liu HW, et al. The effect of contouring variability on dosimetric parameters for brain metastases treated with stereotactic radiosurgery. *Int J Radiat Oncol Biol Phys*. (2013) 87:924–31. doi: 10.1016/j.ijrobp.2013.09.013

66. Wee CW, An HJ, Kang HC, Kim HJ, Wu HG. Variability of gross tumor volume delineation for stereotactic body radiotherapy of the lung with Tri-(60)Co magnetic resonance image-guided radiotherapy system (viewray): a comparative study with magnetic resonance- and computed tomography-based target delineation. *Technol Cancer Res Treat*. (2018) 17:1533033818787383. doi: 10.1177/1533033818787383

67. Olberg S, Green O, Cai B, Yang D, Rodriguez V, Zhang H, et al. Optimization of treatment planning workflow and tumor coverage during daily adaptive magnetic resonance image guided radiation therapy (MR-IGRT) of pancreatic cancer. *Radiat Oncol*. (2018) 13:51. doi: 10.1186/s13014-018-1000-7

68. Han F, Zhou Z, Du D, Gao Y, Rashid S, Cao M, et al. Respiratory motion-resolved, self-gated 4D-MRI using Rotating Cartesian K-space (ROCK): initial clinical experience on an MRI-guided radiotherapy system. *Radiother Oncol*. (2018) 127:467–73. doi: 10.1016/j.radonc.2018.04.029

69. Cusumano D, Dhont J, Boldrini L, Chiloiro G, Teodoli S, Massaccesi M, et al. Predicting tumour motion during the whole radiotherapy treatment: a systematic approach for thoracic and abdominal lesions based on real time MR. *Radiother Oncol*. (2018) 129:456–62. doi: 10.1016/j.radonc.2018.07.025

70. Palacios MA, Bohoudi O, Bruynzeel AME, van Sorsen de Koste JR, Cobussen P, Slotman BJ, et al. Role of daily plan adaptation in MR-guided stereotactic ablative radiation therapy for adrenal metastases. *Int J Radiat Oncol Biol Phys*. (2018) 102:426–33. doi: 10.1016/j.ijrobp.2018.06.002

71. van Sornsen de Koste JR, Palacios MA, Bruynzeel AME, Slotman BJ, Senan S, Lagerwaard FJ. MR-guided gated stereotactic radiation therapy delivery for lung, adrenal, and pancreatic tumors: a geometric analysis. *Int J Radiat Oncol Biol Phys*. (2018) 102:858–66. doi: 10.1016/j.ijrobp.2018.05.048

72. Heerkens HD, van Vulpen M, Erickson B, Reerink O, Intven MP, van den Berg CA, et al. MRI guided stereotactic radiotherapy for locally advanced pancreatic cancer. *Br J Radiol*. (2018) 91:20170563. doi: 10.1259/bjr.20170563

73. 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/ncomms5644

74. Peeken JC, Nusslin F, Combs SE. "Radio-oncomics": the potential of radiomics in radiation oncology. *Strahlenther Onkol*. (2017) 193:767–79. doi: 10.1007/s00066-017-1175-0

75. Andreassen CN, Schack LM, Laursen LV, Alsner J. Radiogenomics - current status, challenges and future directions. *Cancer Lett*. (2016) 382:127–36. doi: 10.1016/j.canlet.2016.01.035

76. Dang M, Lysack JT, Wu T, Matthews TW, Chandarana SP, Brockton NT, et al. MRI texture analysis predicts p53 status in head and neck squamous cell carcinoma. *AJNR Am J Neuroradiol*. (2015) 36:166–70. doi: 10.3174/ajnr.A4110

77. Meyer HJ, Schob S, Hohn AK, Surov A. MRI texture analysis reflects histopathology parameters in thyroid cancer - a first preliminary study. *Transl Oncol*. (2017) 10:911–6. doi: 10.1016/j.tranon.2017.09.003

78. Yang X, Wu N, Cheng G, Zhou Z, Yu DS, Beitler JJ, et al. Automated segmentation of the parotid gland based on atlas registration and machine learning: a longitudinal MRI study in head-and-neck radiation therapy. *Int J Radiat Oncol Biol Phys*. (2014) 90:1225–33. doi: 10.1016/j.ijrobp.2014.08.350

79. Shiradkar R, Podder TK, Algohary A, Viswanath S, Ellis RJ, Madabhushi A. Radiomics based targeted radiotherapy planning (Rad-TRaP): a computational framework for prostate cancer treatment planning with MRI. *Radiat Oncol*. (2016) 11:148. doi: 10.1186/s13014-016-0718-3

80. Zhang B, Ouyang F, Gu D, Dong Y, Zhang L, Mo X, et al. Advanced nasopharyngeal carcinoma: pre-treatment prediction of progression based on multi-parametric MRI radiomics. *Oncotarget*. (2017) 8:72457–65. doi: 10.18632/oncotarget.19799

81. Zhang B, He X, Ouyang F, Gu D, Dong Y, Zhang L, et al. Radiomic machine-learning classifiers for prognostic biomarkers of advanced nasopharyngeal carcinoma. *Cancer Lett*. (2017) 403:21–7. doi: 10.1016/j.canlet.2017.06.004

82. Thor M, Tyagi N, Hatzoglou V, Apte A, Saleh Z, Riaz N, et al. A magnetic resonance Imaging-based approach to quantify radiation-induced normal tissue injuries applied to trismus in head and neck cancer. *Phys Imaging Radiat Oncol*. (2017) 1:34–40. doi: 10.1016/j.phro.2017.02.006

83. Gainey M, Carles M, Mix M, Meyer PT, Bock M, Grosu AL, et al. Biological imaging for individualized therapy in radiation oncology: part I physical and technical aspects. *Future Oncol*. (2018) 14:737–49. doi: 10.2217/fon-2017-0464

84. Jethanandani A, Lin TA, Volpe S, Elhalawani H, Mohamed ASR, Yang P, et al. Exploring applications of radiomics in magnetic resonance imaging of head and neck cancer: a systematic review. *Front Oncol*. (2018) 8:131. doi: 10.3389/fonc.2018.00131

85. Traverso A, Wee L, Dekker A, Gillies R. Repeatability and reproducibility of radiomic features: a systematic review. *Int J Radiat Oncol Biol Phys*. (2018) 102:1143–58. doi: 10.1016/j.ijrobp.2018.05.053

86. Guan Y, Li W, Jiang Z, Chen Y, Liu S, He J, et al. Whole-lesion apparent diffusion coefficient-based entropy-related parameters for characterizing cervical cancers: initial findings. *Acad Radiol*. (2016) 23:1559–67. doi: 10.1016/j.acra.2016.08.010

87. McNutt TR, Benedict SH, Low DA, Moore K, Shpitser I, Jiang W, et al. Using big data analytics to advance precision radiation oncology. *Int J Radiat Oncol Biol Phys*. (2018) 101:285–91. doi: 10.1016/j.ijrobp.2018.02.028

88. Wang J, Kato F, Oyama-Manabe N, Li R, Cui Y, Tha KK, et al. Identifying triple-negative breast cancer using background parenchymal enhancement heterogeneity on dynamic contrast-enhanced mri: a pilot radiomics study. *PLoS ONE*. (2015) 10:e0143308. doi: 10.1371/journal.pone.0143308

89. 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) 69:127–57. doi: 10.3322/caac.21552

90. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, et al. Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. *Sci Data.* (2017) 4:170117. doi: 10.1038/sdata.2017.117

91. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data.* (2016) 3:160018. doi: 10.1038/sdata.2016.18

92. Kiser KJ, Fuller CD, Reed VK. Artificial intelligence in radiation oncology treatment planning: a brief overview. *J Med Artif Intell.* (2019) 10:e2475. doi: 10.21037/jmai.2019.04.02

93. Parasidis E, Pike E, McGraw D. A belmont report for health data. *N Engl J Med.* (2019) 380:1493–5. doi: 10.1056/NEJMp1816373

94. Rajkomar A, Hardt M, Howell MD, Corrado G, Chin MH. Ensuring fairness in machine learning to advance health equity. *Ann Intern Med.* (2018) 169:866–72. doi: 10.7326/M18-1990

95. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. *N Engl J Med.* (2019) 380:1347–58. doi: 10.1056/NEJMra1814259

96. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, et al. A guide to deep learning in healthcare. *Nat Med.* (2019) 25:24–9. doi: 10.1038/s41591-018-0316-z

97. Hinton G. Deep learning-a technology with the potential to transform health care. *JAMA.* (2018) 320:1101–2. doi: 10.1001/jama.2018.11100

98. Mayo CS, Phillips M, McNutt TR, Palta J, Dekker A, Miller RC, et al. Treatment data and technical process challenges for practical big data efforts in radiation oncology. *Med Phys.* (2018) 45:e793–810. doi: 10.1002/mp.13114

99. Jochems A, Deist TM, van Soest J, Eble M, Bulens P, Coucke P, et al. Distributed learning: developing a predictive model based on data from multiple hospitals without data leaving the hospital – a real life proof of concept. *Radiother Oncol.* (2016) 121:459–67. doi: 10.1016/j.radonc.2016.10.002

100. Jochems A, Deist TM, El Naqa I, Kessler M, Mayo C, Reeves J, et al. Developing and validating a survival prediction model for NSCLC patients through distributed learning across 3 countries. *Int J Radiat Oncol Biol Phys.* (2017) 99:344–52. doi: 10.1016/j.ijrobp.2017.04.021

101. Wu Y, Jiang X, Kim J, Ohno-Machado L. Grid Binary LOgistic REgression (GLORE): building shared models without sharing data. *J Am Med Inform Assoc.* (2012) 19:758–64. doi: 10.1136/amiajnl-2012-000862

102. Li Y, Jiang X, Wang S, Xiong H, Ohno-Machado L. VERTIcal Grid lOgistic regression (VERTIGO). *J Am Med Inform Assoc.* (2016) 23:570–9. doi: 10.1093/jamia/ocv146

103. Welcome to FHIR [Web Page]. *HL7 International.* (2018). Available online at: http://hl7.org/fhir/ (accessed December 27, 2018).

104. HHS Proposes New Rules to Improve the Interoperability of Electronic Health Information. *New Innovations in Technology Promote Patient Access and Could Make No-Cost Health Data Exchange a Reality for Millions.* Washington, DC: US Department of Health & Human Services. (2019). Available online at: https://www.hhs.gov/about/news/2019/02/11/hhs-proposes-new-rules-improve-interoperability-electronic-health-information.html

105. Cannon R. *SMART on FHIR: The Basics [Video]: InterSystems Developer Community* (2018). Available online at: https://www.youtube.com/watch?v=MMfe181tmwU&t=6s/ (accessed April 20, 2019).

106. Pan HY, Mazur LM, Martin NE, Mayo CS, Santanam L, Pawlicki T, et al. Radiation oncology health information technology: is it working for or against us? *Int J Radiat Oncol Biol Phys.* (2017) 98:259–62. doi: 10.1016/j.ijrobp.2017.02.018

107. Phillips M, Halasz L. Radiation oncology needs to adopt a comprehensive standard for data transfer: the case for HL7 FHIR. *Int J Radiat Oncol Biol Phys.* (2017) 99:1073–5. doi: 10.1016/j.ijrobp.2017.08.007

Bulk Anatomical Density Based Dose Calculation for Patient-Specific Quality Assurance of MRI-Only Prostate Radiotherapy

Jae Hyuk Choi[1], Danny Lee[1], Laura O'Connor[1,2], Stephan Chalup[3], James S. Welsh[3], Jason Dowling[1,4,5] and Peter B. Greer[1,2]*

[1] School of Mathematical and Physical Sciences, University of Newcastle, Newcastle, NSW, Australia, [2] Department of Radiation Oncology, Calvary Mater Newcastle Hospital, Newcastle, NSW, Australia, [3] School of Electrical Engineering and Computing, University of Newcastle, Newcastle, NSW, Australia, [4] South Western Sydney Clinical School, University of New South Wales, Sydney, NSW, Australia, [5] The Australian E-Health Research Centre, Commonwealth Scientific and Industrial Research Organisation, Brisbane, QLD, Australia

Prostate cancer treatment planning can be performed using magnetic resonance imaging (MRI) only with sCT scans. However, sCT scans are computer generated from MRI data and therefore robust, efficient, and accurate patient-specific quality assurance methods for dosimetric verification are required. Bulk anatomical density (BAD) maps can be generated based on anatomical contours derived from the MRI image. This study investigates and optimizes the BAD map approach for sCT quality assurance with a large patient CT and MRI dataset. 3D T2-weighted MRI and full density CT images of 54 patients were used to create BAD maps with different tissue class combinations. Mean Hounsfield units (HU) of Fat (F: below −30 HU), the entire Tissue [T: excluding bone (B)], and Muscle (M: excluding bone and fat) were derived from the CT scans. CT based BAD maps ($BAD_{BT,CT}$ and $BAD_{BMF,CT}$) and a conventional bone and water bulk-density method ($BAD_{BW,CT}$) were compared to full CT calculations with bone assignments to 366 HU (measured) and 288 HU (obtained from literature). Optimal bulk densities of Tissue for $BAD_{BT,CT}$ and Bone for $BAD_{BMF,CT}$ were derived to provide zero mean isocenter dose agreement to the CT plan. Using the optimal densities, the dose agreement of $BAD_{BT,CT}$ and $BAD_{BMF,CT}$ to CT was redetermined. These maps were then created for the MRI dataset using auto-generated contours and dose calculations compared to CT. The average mean density of Bone, Fat, Muscle, and Tissue were 365.5 ± 62.2, −109.5 ± 12.9, 23.3 ± 9.7, and −46.3 ± 15.2 HU, respectively. Comparing to other bulk-density maps, $BAD_{BMF,CT}$ maps provided the closest dose to CT. Calculated optimal mean densities of Tissue and Bone were −32.7 and 323.7 HU, respectively. The isocenter dose agreement of the optimal density assigned $BAD_{BT,CT}$ and $BAD_{BMF,CT}$ to full density CT were 0.10 ± 0.65% and 0.01 ± 0.45%, respectively. The isocenter dose agreement of MRI generated $BAD_{BT,MR}$ and $BAD_{BMF,MR}$ to full density CT were −0.15 ± 0.90% and −0.16 ± 0.65%, respectively. The BAD method with optimal bulk densities can provide robust, accurate and efficient patient-specific quality assurance for dose calculations in MRI-only radiotherapy.

Keywords: MRI-only planning, synthetic CT, bulk density, anatomical structure, quality assurance, dosimetric verification

INTRODUCTION

Magnetic resonance imaging (MRI) only treatment planning is of current interest to reduce systematic registration errors between CT and MRI and improve workflows (1–4). MRI-only treatment planning involves generation of synthetic CT (sCT), since it is not straightforward to convert MRI to electron densities of different tissue classes which are necessary for photon dose calculation in treatment planning systems (TPS).

Different methods have been introduced to create sCT scans for prostate radiotherapy planning. Bulk-density planning was initially investigated as a method for sCT generation (5–10). These studies applied a density of water to the body with an additional separate density for bone. Atlas-based methods involve pair-wise image registration of CT and MRI scans based on anatomical structures to form the atlas, registration of atlas MRI scans to target MRI scan, and mapping the estimated Hounsfield unit (HU) values based on the atlas CT data (11–14). Patch based methods involve feature extraction and patch partitioning from interpatient group-wise affine registration (10, 15). The target feature patches are selected using the approximate nearest neighbor search from the training cohort and sCT patches are generated using the multipoint-wise aggregation scheme. Tissue-classification methods have been developed which assign a single density to each tissue class or assign the continuous HU value based on tissue class probabilities (2). Calibration-type voxel methods use mapping of the MRI signal to HU, however, these require initial identification of bone and surrounding tissue regions with application of separate mapping functions (16). More recently deep learning approaches show promise particularly for generation speed (17, 18). Information on sCT generation methods are available from recent review articles (19, 20).

However, sCT scans are computer generated from large-field-of-view MRI data which can contain artifacts due to image non-uniformities and magnetic field inhomogeneities which can be both scanner and patient dependent (21, 22). They must perform accurately for the variation in patient anatomy that is present in the population and this remains a challenge (23). A recent failure modes and effects analysis (FMEA) of MRI-only planning identified that generation of sCT propagated 46 unique failure modes with 15 failure modes having high risk priority numbers (24). This was significantly more failure modes than the conventional workflow. While CT scanning is a robust and consistent technique, the robustness of sCT is not as high or as well-understood and clinical implementation should proceed with appropriate verifications. Therefore, a quality assurance method that could validate sCT on a patient-specific basis would be desirable. Such a method would ideally fulfill the following criteria: be independent of the sCT method; robust to patient anatomical variations; insensitive to MRI scanner artifacts; efficient to perform; easy to automate; and accurate within clinically acceptable limits.

The bulk-density approach is potentially an ideal candidate to achieve the above criteria for quality assurance of sCT. Most studies that have been performed for bulk-density assignment however have had relatively small patient datasets and assigned

TABLE 1 | Patients and image acquisition parameters.

Patients		Group 1 (n = 39)	Group 2 (n = 15)
Age		58 ~ 78 year (69 ± 4.7 year)	58 ~ 83 year (72 ± 6.5 year)
Weight		54 ~ 115.4 kg (87.1 ± 13.2 kg)	62 ~ 122 kg (90.2 ± 17.3 kg)
Imaging year		2012 ~ 2014	2017 ~ 2018
Imaging parameters or sequence	CT	32 patients: GE LightSpeedRT (140 kVp; 2.5 mm slice thickness) 7 patients: Toshiba Aquilion (120 kVp; 2.0 mm slice thickness)	3 patients: GE LightSpeedRT (140 kVp; 2.5 mm slice thickness) 6 patients: Toshiba Aquilion (120 kVp; 2.0 mm slice thickness) 6 patients: Siemens SOMATOM (120 kVp; 2.0 mm slice thickness)
	MRI	Siemens Skyra 3.0 T 3D T2-weighted SPACE (Sampling perfection with application optimized contrasts using different flip angle evolution) sequence Echo time (TE) = 102 ms; Repetition time (TR) = 1,200 ms; Flip angle = 135°; Field of view (FOV) = 430 mm; Slice thickness = 1.6 mm	
Treatment plan		39 IMRT	11 IMRT, 4 VMAT

arbitrary or literature derived values for the densities (5, 6, 8, 9, 25). Improved agreement to CT dose has been demonstrated with calculation of bone density using effective path-length calculations suggesting that accurate dose calculations are achievable (8).

In this study, we investigate and optimize the bulk-density planning approach to develop a method for patient-specific quality assurance of sCT. Two separate bulk-density methods are investigated with two and three tissue classes, respectively. A large patient cohort of 54 prostate patients is used to measure and determine optimal bulk HU values for the tissue classes that minimize differences with full CT dose calculations. The method is tested using MRI scan assignment of the optimal bulk HU values for the 54 patients following automatic segmentation of bone and body contours. It is referred to here as the bulk anatomical density (BAD) map method.

METHODS

Patient Data

This study used CT and MR data of 54 prostate cancer patients measured in clinical studies. All data was acquired under ethics board approval with informed consent. Detailed patient data and imaging parameter settings are shown in **Table 1**. These patient MRI scans were previously acquired for development and validation of sCT generation for MRI-only planning or for a prospective study of MRI-only workflow implementation (26).

Bulk Anatomical Density Maps

The role of the BAD map in quality assurance of the MRI-only workflow is shown in **Figure 1**.

The BAD method is an extension of the conventional bulk-density map. Up to three tissue classes have been investigated

FIGURE 1 | Proposed MRI-only radiation therapy workflow with suggested quality assurance steps for sCT dose verification using the bulk anatomical density (BAD) method.
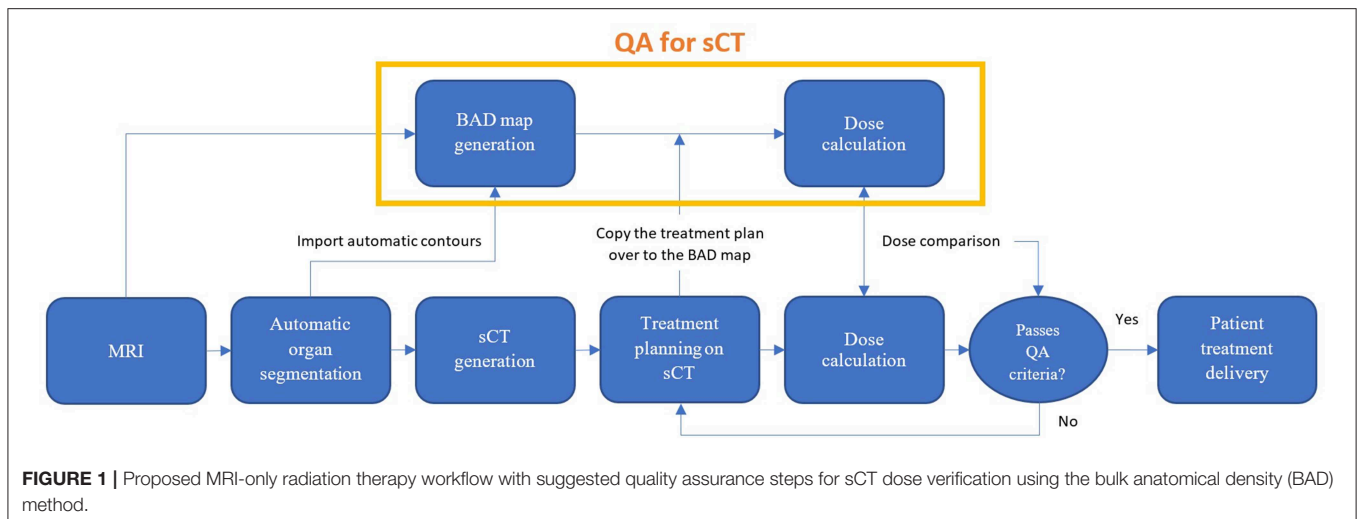
(Bone, Muscle, and Fat). BAD maps can be made with different tissue class combinations by assigning the bulk HU values to either CT or MRI patient scans; (1) BAD_{BW} (Bone and Water), (2) BAD_{BT} (Bone and the entire Tissue), and (3) BAD_{BMF} (Bone, Muscle, and Fat). The BAD methods are compared to the conventionally used methods and are shown in **Figure 2**. Derivation of the method has been performed using CT scan data.

Mean Density of Tissue Classes

Tissue segmentation was performed on CT scans to determine mean HU values. CT images were segmented into different tissue classes based on the HU values: Bone (B: >100 HU); Fat (F: HU below −30 HU), and the entire tissue (T: including Fat and Muscle but excluding Bone) areas. Muscle area (M: excluding Bone and Fat) was found using a Boolean operation of $[B^c \cap F^c]$ within body. These were performed using the automatic contouring tools of Varian Eclipse (Varian Medical Systems, Palo Alto, CA, USA). According to Kim et al., HU of adipose tissue, including both subcutaneous and visceral, is within a range of −140 to −30 HU (27). Note that, the Muscle volume includes other organ structures such as bladder and rectum (with gas/air).

The population average mean HU values (±1 standard deviation) of each structure were; Bone = 365.5 ± 62.2 HU, Fat = −109.5 ± 12.9 HU, Muscle = 23.3 ± 9.7 HU, the entire Tissue (T) = −46.3 ± 15.2 HU. Using the department's HU to electron-density conversion curve within the Eclipse TPS these HUs are corresponding to the relative electron densities of 1.17 (Bone), 0.91 (Fat), 1.03 (Muscle), and 0.97 (the entire Tissue) and physical densities of 1.23 g/cm^3 (Bone), 0.92 g/cm^3 (Fat), 1.04 g/cm^3 (Muscle), and 0.98 g/cm^3 (the entire Tissue).

Derivation of Optimal Densities

To determine the optimal bulk-density values a linear fitting method was employed with the assumption that the dose change is approximately proportional to tissue density change at least over small ranges. The $BAD_{BW,CT}$, $BAD_{BT,CT}$, and $BAD_{BMF,CT}$ were created using the mean values for the tissue classes as

determined above. Additionally they were also generated with a separate bone value of 288 HU that was derived using effective path lengths by Lambert et al. (8) and the equation presented by Thomas (28). Assigned densities were rounded up since fractional values are not accepted on planning system (Varian Eclipse).

Here, 6 separate BAD maps were created as follows: $BAD_{BW,CT}$ [Bone = 288 HU; Tissue = 0 HU (Water)], $BAD_{BT,CT}$ (Bone = 288 HU; Tissue = −46 HU), $BAD_{BMF,CT}$ (Bone = 288 HU; Muscle = 23 HU; Fat = −109 HU), and $BAD_{BW,CT}$ [Bone = 366 HU; Tissue = 0 HU (Water)], $BAD_{BT,CT}$ (Bone = 366 HU; Tissue = −46 HU), $BAD_{BMF,CT}$ (Bone = 366 HU; Muscle = 23 HU; Fat = −109 HU). IMRT treatment plans previously developed on the corresponding patient CT or sCT scans were then copied to the BAD maps and dose was recalculated on the BAD maps using the same monitor units and fluences. The same plan dose was also calculated on the gold-standard CT scan.

The mean differences in dose to isocenter for the BAD maps and the CT scan of all 54 patients were determined and plotted. **Figure 3A** illustrates a linear plot for each BAD map method with bone density as the x-axis. This plot can be used to determine an approximately optimal bone density for the methods using the intercept for zero mean dose difference to CT. The calculated optimal bone densities for $BAD_{BW,CT}$, $BAD_{BT,CT}$, and $BAD_{BMF,CT}$ were approximately 127.2, 463.8, and 323.7 HU, respectively. The averaged measured mean density of bone was 365.5 ± 62.2 HU, and therefore the derived optimal bulk-density of bone value of 324 HU (rounded up) was subsequently only used for the BAD_{BMF} maps, and the measured value of 366 HU (rounded up) was retained for the BAD_{BW} and the BAD_{BT} maps.

Figure 3B shows similar plots for the $BAD_{BW,CT}$ and $BAD_{BT,CT}$ maps but with the x-axis changed to the Tissue values. These maps are analogous in that the anatomical regions used are identical. This allows for determination of the optimal density for Tissue for $BAD_{BT,CT}$ maps from the intercept for zero mean dose difference. The calculated optimal densities of the Tissue were approximately −22.0 and −32.7 HU with bone densities of 288 and 366 HU, respectively. For subsequent BAD_{BT} maps the
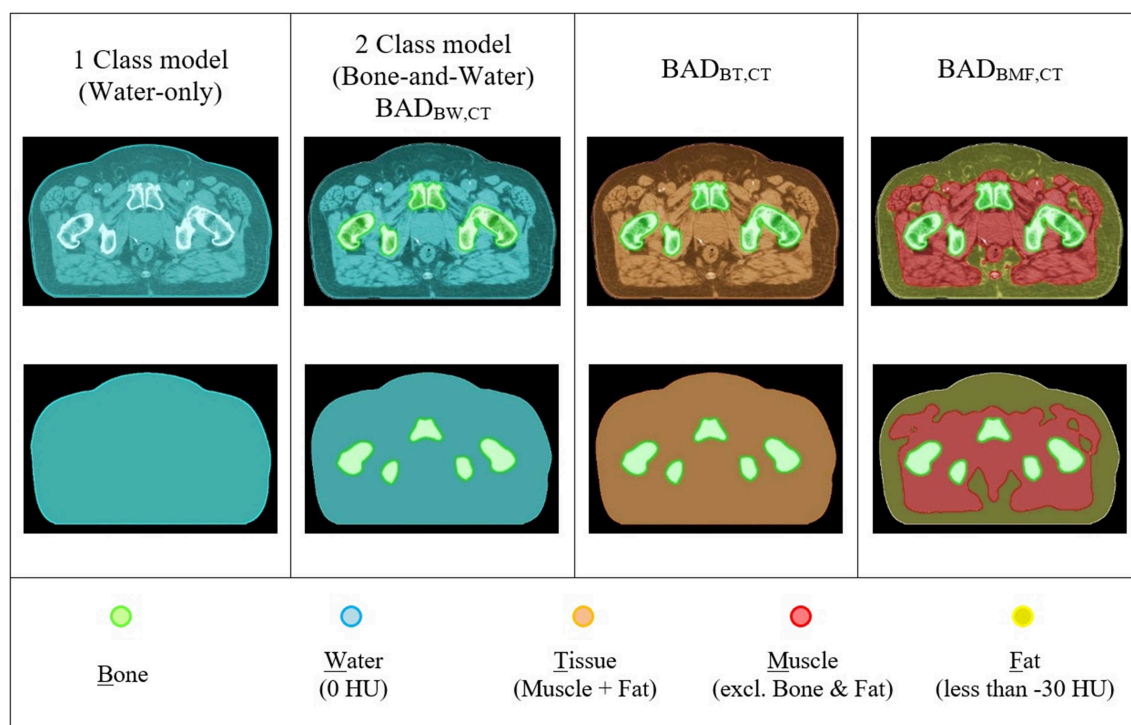
FIGURE 2 | CT based BAD maps. Colors represent the bulk-density assignments on different structures.

measured bone value of 366 HU was used and therefore the tissue density of −33 HU (rounded up) was adopted i.e., BAD_{BT} (Bone = 366 HU; Tissue = −33 HU). HU of 324 is corresponding to relative electron densities of 1.16 and physical densities of 1.20 g/cm^3 while HU of −33 is corresponding to relative electron densities of 0.97 and physical densities of 0.99 g/cm^3.

Dosimetric Accuracy for Optimal BAD$_{CT}$ Maps

Using the optimal densities and CT derived anatomical contours, two BAD_{CT} maps [$BAD_{BT,CT}$ (Bone = 366 HU; Tissue = −33 HU) and $BAD_{BMF,CT}$ (Bone = 324 HU; Muscle = 23 HU; Fat = −109 HU)] were created and tested for all 54 patients. As described above, doses were recalculated on these BAD maps and compared to CT calculation. Isocenter doses of each plan were compared to the corresponding CT dose.

Dosimetric Accuracy for BAD$_{MR}$ Maps

The method was then applied to the large-field-of-view T2-weighted MRI scans for the patients. Two BAD_{MR} maps [$BAD_{BT,MR}$ (Bone = 366 HU; Tissue = −33 HU) and $BAD_{BMF,MR}$ (Bone = 324 HU; Muscle = 23 HU; Fat = −109 HU)] were created analogous to CT above (**Figure 4**). To derive the anatomical contours for density assignment, the automatic MRI body and bone contouring method that was developed in a previous sCT study was utilized (13). For the $BAD_{BMF,MR}$ the fat contour created on the anatomically (rigid) registered CT was used for density assignment. This will require replacement with an MRI based method in the future, for example using

DIXON scans, however these were not available at this time. It is assumed that the segmentation of fat in MRI will correspond to the fat utilized here. The treatment plan on CT was copied directly over to the BAD_{MR} maps and isocenter doses of each map were compared with the corresponding CT plan. The 3D Gamma comparison metric was used for dose comparisons in all voxels (29). This used varying gamma criteria 3%, 3 mm, 2%, 2 mm, 1%, 1 mm, a low dose threshold of 20% for inclusion, and the CT dose was used as the reference. A 15 mm erosion operator was used to remove the region close to the skin border from the calculation as this gives large gamma discrepancies due to contour differences.

RESULTS

Dosimetric Accuracy for BAD$_{CT}$ Maps Using Measured Densities

The isocenter point dose results are shown in **Table 2** and **Figure 5** for the six BAD_{CT} maps for measured (non-optimal) mean HU values and the same maps but with bone density of 288. The BAD map with three tissue classifications (BMF) provided the closest matching to the full density CT plans with smaller variations compared to other two bulk-density maps (BW and BT). However, significant systematic differences to CT are still evident particularly for the first two methods. The literature bone density performs slightly better for the BW map while the measured mean bone density performs better for the BT map while there is no clear winner for the BMF map.
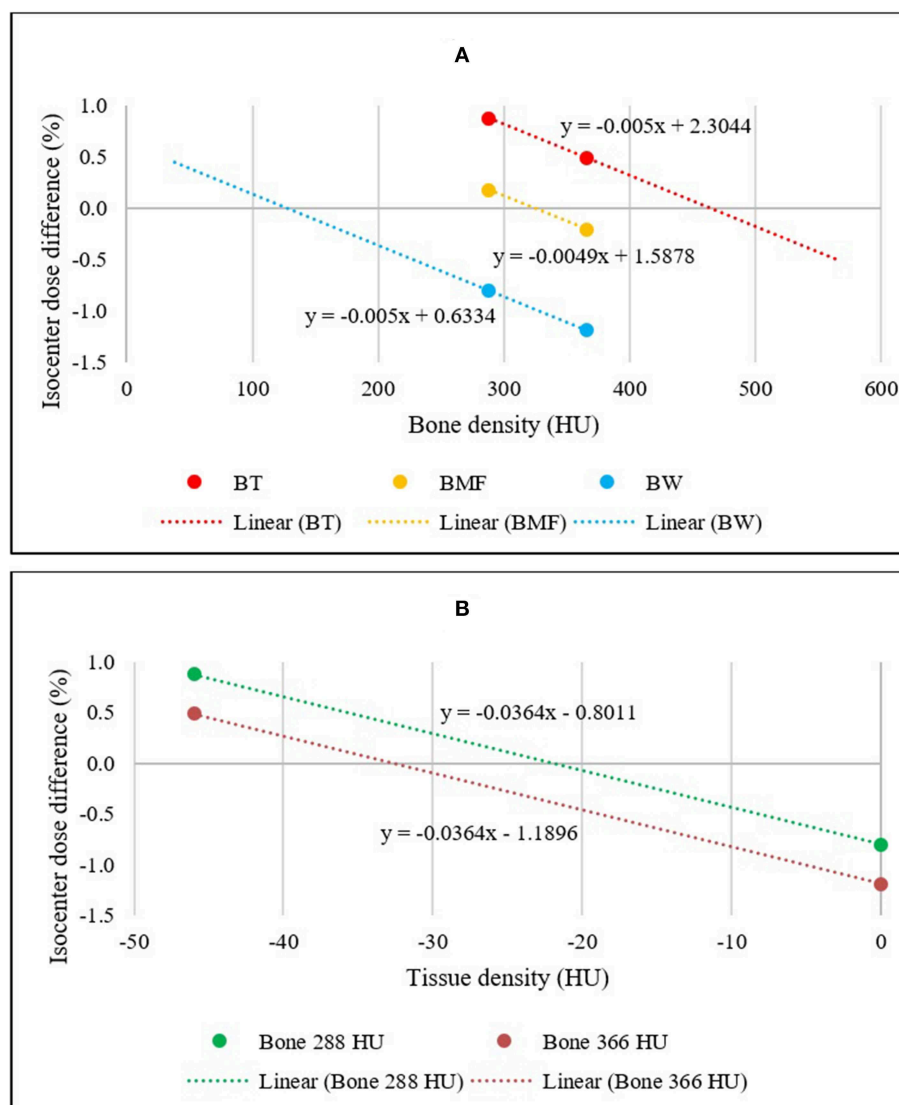
FIGURE 3 | (A) Linear relationships between the mean isocenter dose differences of BAD$_{BW,CT}$, BAD$_{BT,CT}$, and BAD$_{BMF,CT}$ maps to full density CT for two bone density values. The x-intercepts represent the optimal bulk bone density of each method. **(B)** Linear relationships between the mean isocenter dose differences of BAD$_{BW,CT}$ and BAD$_{BT,CT}$ maps to full density CT. The x-intercepts represent the optimal bulk Tissue density for BAD$_{BT,CT}$.

Dosimetric Accuracy for Optimal BAD$_{CT}$ Maps

Figure 6 shows the results for two optimal density BAD$_{CT}$ maps [BAD$_{BT,CT}$ (Bone, B = 366 HU; Tissue, T = −33 HU) and BAD$_{BMF,CT}$ (Bone, B = 324 HU; Muscle, M = 23 HU; Fat, F = −109 HU)]. Significant improvements were observed for both optimal BAD maps and the isocenter dose differences to CT were observed as 0.10 ± 0.65 and $0.01 \pm 0.45\%$, respectively. The interquartile ranges (IQR) were from −0.34% to 0.65% and −0.39% to 0.41%, respectively.

Dosimetric Accuracy for BAD$_{MR}$ Maps

With the optimal density assignment, the isodose differences to full density CT plan were observed to be $-0.15 \pm 0.90\%$ on

BAD$_{BT,MR}$ and $-0.16 \pm 0.65\%$ on BAD$_{BMF,MR}$ (**Figure 7**). The IQR of both BAD$_{MR}$ maps were within ±0.7%; BAD$_{BT,MR}$ was from −0.65 to 0.31%, while BAD$_{BMF,MR}$ was from −0.60 to 0.22%. The mean differences and standard deviations are slightly larger than the CT derived BAD maps which would be expected due to the different anatomical contours used for MRI. Results of gamma analysis pass-rate are shown in **Table 3**.

DISCUSSION

In this study, the BAD method was developed to map bulk densities to anatomical structures using data measured from 54 prostate cancer patients. A linear interpolation method was used to determine the optimal HU values to give approximately
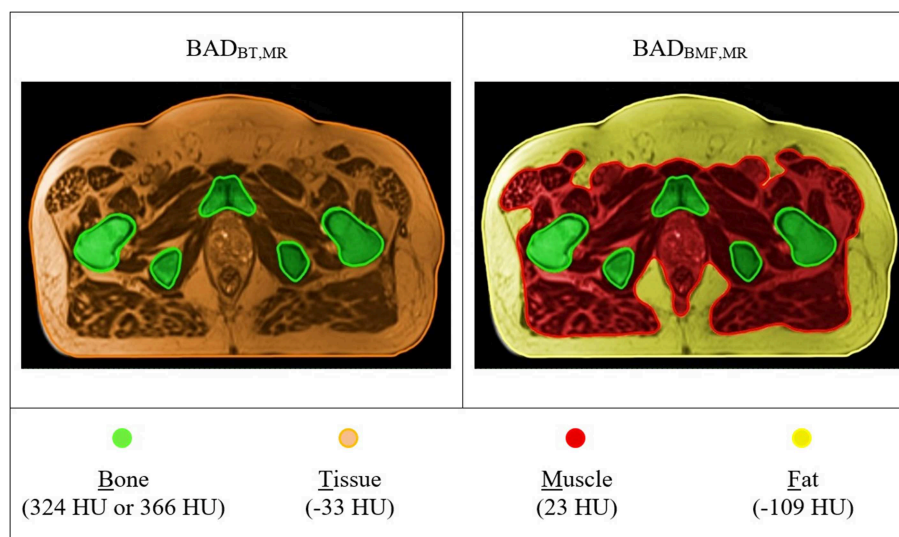
FIGURE 4 | T2-weighted MR based BAD maps with optimal density values. Colors represent the bulk-density assignments on different structures. $BAD_{BT,MR}$ (Bone, B = 366 HU; Tissue, T = −33 HU) and $BAD_{BMF,MR}$ (Bone, B = 324 HU; Muscle, M = 23 HU; Fat, F = −109 HU).

TABLE 2 | Mean (±1 standard deviation) isocenter dose difference for BAD maps from full density CT plan using measured (non-optimal) densities (Water, W = 0 HU; Tissue, T = 46 HU; Muscle, M = 23 HU; and Fat, F = −109 HU) and two bone densities, measured (B = 366 HU) and literature (B = 288 HU).

BAD_{CT} plan	Isocenter dose difference to CT (%)	
	B = 288 HU	B = 366 HU
BW	−0.80 ± 0.69	−1.19 ± 0.69
BT	0.87 ± 0.63	0.49 ± 0.63
BMF	0.18 ± 0.45	−0.21 ± 0.45

zero mean isocenter dose agreement to CT. Using the optimal HU values, an improvement in isocenter dose agreement was observed when compared with using measured mean HU values. Particularly, the three tissue class BAD_{BMF} method provided the closest dose agreement to conventional CT but the two class BAD_{BT} still gave acceptable agreement. Both optimal bulk-density assigned BAD_{MR} maps provided mean dosimetric differences within 0.2% to conventional CT with standard deviations within 0.9%. Therefore, both methods can be considered for dose verification and patient-specific quality assurance of sCT scans. To use the BAD_{BT} with MRI-only workflows only requires that the body and bone contours on a large-field-of-view MRI scan are segmented and this can be automated with atlas or similar methods. The generation of the BAD map could potentially be fully automated. It meets the criteria outlined above for a method for quality assurance of sCT (independence; robustness; insensitive to scan artifacts; efficient and easy to automate; and accurate within clinically acceptable limits). In principle, the results here also suggest that this method may be adequate in accuracy for sCT generation for dose calculation in MRI-only workflows however these scans may not be suitable for image-guidance.

The linear interpolation method used makes the assumption that the isocenter dose is a linear function of the particular tissue density that is modified. However, this is clearly an oversimplification of dose deposition processes and this can be seen in the results. The optimal densities do not result in exactly zero dose difference to CT dose for the optimal BAD_{CT} maps although the difference is within 0.1%. To obtain a mean of zero an iterative process could be conducted although this would be extremely time consuming for negligible benefit. Furthermore, when applied to MRI scans where the bone contours are derived using an entirely different method, the mean isocenter dose to CT is modified further with a decrease in the BAD map doses to isocenter for the same patients when compared to CT. This could be due to the MRI bone contour being larger than the CT bone contour which is often observed in clinical practice although other factors could also contribute including the body contour.

The methodology was derived using HU which corresponds to the sCT literature. The Eclipse TPS converts these to relative electron density (RED) using our in-house conversion and therefore the RED values as well as physical density are stated in the manuscript.

These results can be compared with other bulk-density methods reported on previous studies. The early reported works in MRI-only planning used bulk-density assignment to CT or MRI but with limited datasets, and in some cases homogeneous CT calculations (5–7). Kim et al. used bulk-density assignment to water and bone for 15 prostate patients with 300 HU assigned to bone that was measured as the mean value within femoral head contours on CT. The PTV ($D_{95\%}$) differences were 1.9% (9). Lambert et al. studied 39 prostate patients and for their CT dataset with a density of water and a density of 1.19 g/cm³ for bone found a mean dose isocenter difference to CT of 0.2% which is lower than that found here using the same density value (8). The reason for this is not entirely clear, it could be related to
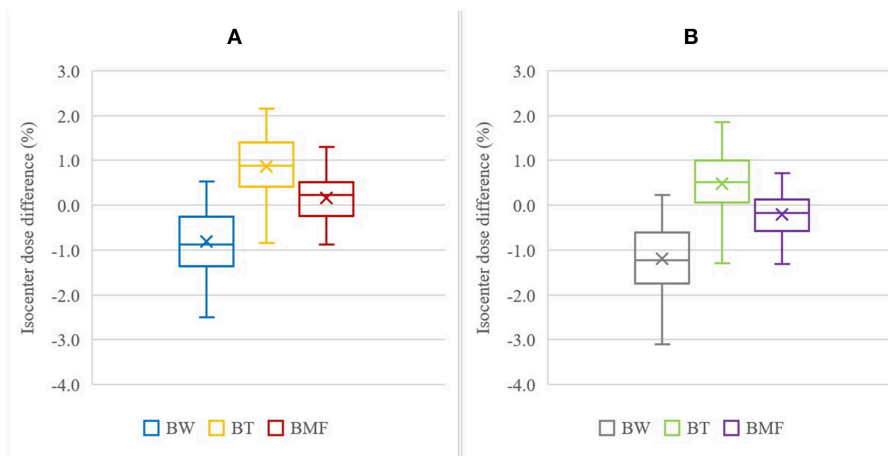
FIGURE 5 | Isocenter dose difference to CT for BAD$_{CT}$ methods using measured (non-optimal) densities with bone, B, as 288 HU **(A)** and 366 HU **(B)**. The other tissue densities used were: Water, W = 0 HU; Tissue, T = 46 HU; Muscle, M = 23 HU; and Fat, F = −109 HU. The cross mark "×" represent the mean of the results, the horizontal bar inside the box is the median, the extent of the boxes represents the interquartile range (IQR) between the first quartile (Q1) and the third quartile (Q3), and the ends of whiskers represent the minimum and maximum range.
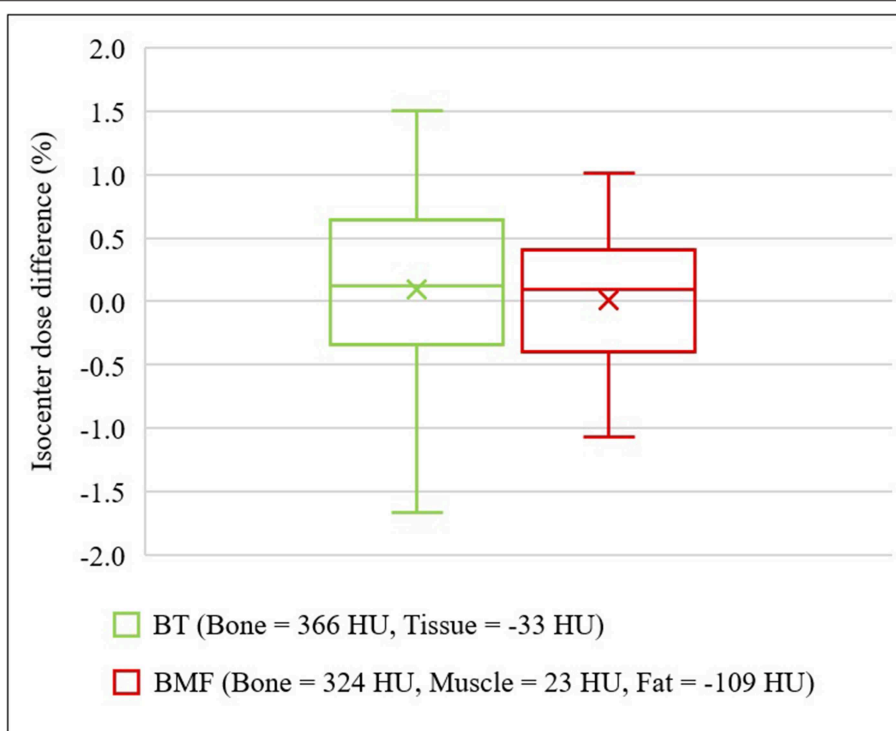
FIGURE 6 | Isocenter dose differences to CT for optimal density BAD$_{CT}$ maps. BAD$_{BT,CT}$ (Bone, B = 366 HU; Tissue, T = −33 HU) and BAD$_{BMF,CT}$ (Bone, B = 324 HU; Muscle, M = 23 HU; Fat, F = −109 HU).

changes in TPS calculation algorithms. Their paper used two separate planning systems and dose calculation algorithms have since improved.

There is general lack of consensus on the density to assign to bone and as we have shown here consideration of three tissue classes yields better results. The optimal density is likely to be influenced by the method used to determine the anatomical contour, the method used to determine the optimal density (mean or path length), the volume and location of the anatomical structure, i.e., all bone or just femur, and the TPS algorithm. In many cases these details are not currently given in the relevant literature. An assessment of the sensitivity of bulk-density
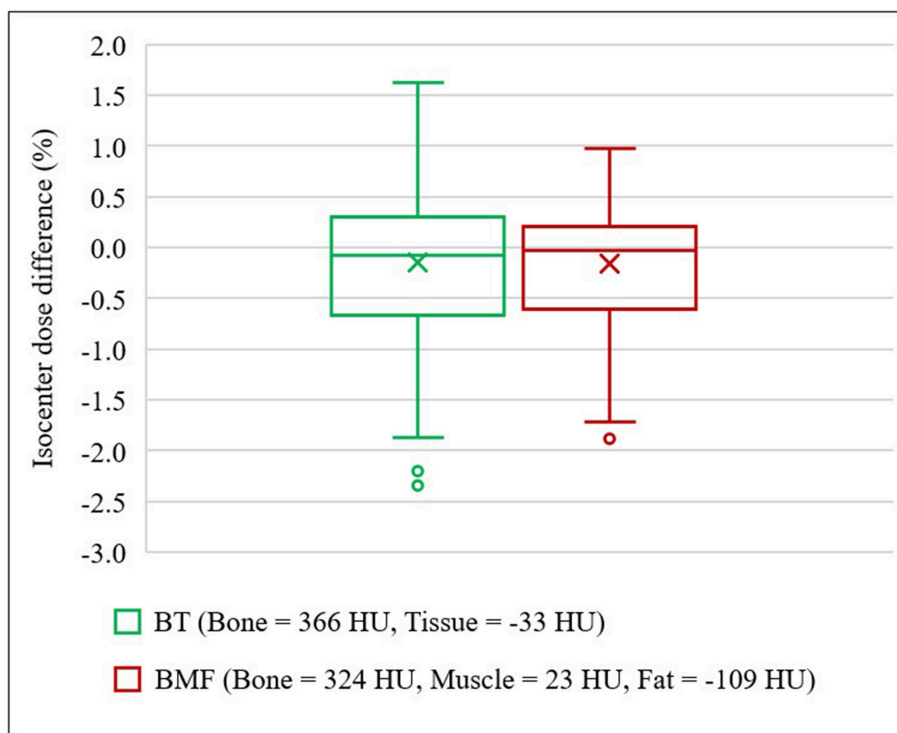
FIGURE 7 | Isocenter dose differences to CT for optimal density BAD_{MR} maps. $BAD_{BT,MR}$ (Bone, B = 366 HU; Tissue, T = −33 HU) and $BAD_{BMF,MR}$ (Bone, B = 324 HU; Muscle, M = 23 HU; Fat, F = −109 HU). In the presence of outliers, the ends of whiskers indicate the lower (Q1 − 1.5 × IQR) and upper (Q3 + 1.5 × IQR) limits.

TABLE 3 | Gamma analysis pass-rate results for comparison between BAD_{MR} maps and CT dose calculations.

	$BAD_{BT,MR}$			$BAD_{BMF,MR}$		
	3%/3 mm	2%/2 mm	1%/1 mm	3%/3 mm	2%/2 mm	1%/1 mm
Mean (%)	99.9 ± 0.2	99.3 ± 1.1	90.7 ± 10.2	99.9 ± 0.2	99.6 ± 0.6	93.8 ± 8.6
Min (%)	99.1	94.6	58.8	99.3	97.2	58.5

$BAD_{BT,MR}$ (Bone, B = 366 HU; Tissue, T = −33 HU) and $BAD_{BMF,MR}$ (Bone, B = 324 HU; Muscle, M = 23 HU; Fat, F = −109 HU).

planning to these factors would be of interest. When the method derived here was applied to MRI data using a different bone contouring algorithm the results were similar which suggest that it is relatively robust to the segmentation method and variations in the bone contour. The method is simple to perform using standard anatomical contouring techniques and TPS system reporting of mean densities to these contours. Therefore, the method can be used to determine an optimal density for any particular center's practice if necessary.

This study used a large patient dataset of 54 CT scans to determine the optimal HU values and validation with 54 patient large-field-of-view MRI scans. Both optimal bulk-density values for bone and the entire tissue are within the variations of the average mean bulk-density calculated from the cohort. Thus, these would be valid for future applications particularly for those male patients weighing between 54 and 122 kg. A consideration would be to apply these values to an external patient CT/MRI dataset for further validation.

One limitation of the study is that the fat contour from the registered CT scan was used for generation of the $BAD_{BMF,MR}$ map. For fat class segmentation on MRI, fast DIXON scans could be incorporated into the MRI acquisition protocol for future study to generate three-class BAD_{MR} maps to improve the dose calculation accuracy, in particular to reduce the standard deviation of the results when compared to CT dose.

Application of automated bone segmentation on MRI may cause dosimetric inaccuracy for the BAD_{MR} maps. More accurate segmentation can potentially be achieved via manual contouring however this is time consuming and the level of accuracy can vary depending on the level of expertise and experience of individuals (10, 30).

Previous studies have demonstrated the accuracy of the atlas-based automatic segmentation method that was used for this study. The automated bone contours had mean Dice similarity coefficient scores of 0.91 ± 0.03 and the mean absolute surface distance of 1.45 ± 0.47 mm when compared to expert drawn

manual contours (14). Automatic MRI bone segmentation has become an important component of many sCT generation methods due to its efficiency and accuracy. Korhonen et al. and Koivula et al. also used an atlas-based method for their dual model method for sCT generation and the average PTV mean dose differences of their sCTs to CT were 0.3 ± 0.2% and −0.6 ± 0.4%, respectively (16, 31). Currently, commercially available sCT generation products, for example the FDA-approved Philips Magnetic Resonance for Calculating Attenuation (MRCAT) software package, use a model-based segmentation method for delineating bony structures from the patient's body outline from mDIXON water, fat, and in-phase images (23, 32).

Dose calculation on BAD map could be improved if the density of gas within the rectum was considered. However, most centers routinely control the magnitude of rectal gas through patient preparation prior to scanning or voiding and rescanning. In many cases. the gas may be atypical of treatment and therefore the dose calculation may not reflect treatment dose. Some centers override the density of gas in the rectum for dose calculations. This is a general problem for radiotherapy planning and can be managed in the same way as conventional CT based planning with the advantage that for MRI-only treatment planning rescanning does not require additional patient dose.

In summary, the BAD map is a technique that utilizes anatomical structures for generating BAD maps for patient-specific dose calculations to compare to sCT. With the optimal density assignments, it provides clinically acceptable dose agreement to the conventional full density CT based plans. The three-class BAD model (Bone, Muscle, and Fat) performs best however the two-class BAD model (Bone, Tissue) is also acceptable. The BAD method can provide accurate dose calculations for verifying sCT for clinical use in MRI-only workflows. It has currently been implemented as a quality assurance method in a multi-center trial of prostate stereotactic radiation therapy (NINJA) that includes an MRI-only sub-study.

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 Hunter New England Human Research Ethics Committee. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

JC performed overall data analysis, drafting and editing the manuscript, and manuscript submission. DL assisted with data analysis, contributed to review draft and editing. LO'C assisted with treatment planning and dose calculation. JD calculated MRI automatic bone and body contouring. SC and JW assisted with project design, data analysis, manuscript review and editing. PG assisted with project development, ethical submission, methodology, data analysis, literature review, manuscript review and editing.

FUNDING

REFERENCES

1. Tenhunen M, Korhonen J, Kapanen M, Seppälä T, Koivula L, Collan J, et al. MRI-only based radiation therapy of prostate cancer: workflow and early clinical experience. *Acta Oncol.* (2018) 57:902–7. doi: 10.1080/0284186X.2018.14 45284

2. Tyagi N, Fontenla S, Zhang J, Cloutier M, Kadbi M, Mechalakos J, et al. Dosimetric and workflow evaluation of first commercial synthetic CT software for clinical use in pelvis. *Phys Med Biol.* (2017) 62:2961. doi: 10.1088/1361-6560/aa5452

3. Owrangi AM, Greer PB, Glide-Hurst CK. MRI-only treatment planning: benefits and challenges. *Phys Med Biol.* (2018) 63:05TR1. doi: 10.1088/1361-6560/aaaca4

4. Kerkmeijer LGW, Maspero M, Meijer GJ, van der Voort van Zyp JRN, de Boer HCJ, van den Berg CAT. Magnetic resonance imaging only workflow for radiotherapy simulation and planning in prostate cancer. *Clin Oncol.* (2018) 30:692–701. doi: 10.1016/j.clon.2018.08.009

5. Lee YK, Bollet M, Charles-Edwards G, Flower MA, Leach MO, McNair H, et al. Radiotherapy treatment planning of prostate cancer using magnetic resonance imaging alone. *Radiother Oncol.* (2003) 66:203–16. doi: 10.1016/S0167-8140(02)00440-1

6. Chen L, Price RA Jr, Wang L, Li J, Qin L, McNeeley S, et al. MRI-based treatment planning for radiotherapy: dosimetric verification for prostate IMRT. *Int J Radiat Oncol Biol Phys.* (2004) 60:636–47. doi: 10.1016/S0360-3016(04)00960-5

7. Eilertsen K, Vestad LN, Geier O, Skretting A. A simulation of MRI based dose calculations on the basis of radiotherapy planning CT images. *Acta Oncol.* (2008) 47:1294–302. doi: 10.1080/02841860802256426

8. Lambert J, Greer PB, Menk F, Patterson J, Parker J, Dahl K, et al. MRI-guided prostate radiation therapy planning: investigation of dosimetric accuracy of MRI-based dose planning. *Radiother Oncol.* (2011) 98:330–4. doi: 10.1016/j.radonc.2011.01.012

9. Kim J, Garbarino K, Schultz L, Levin K, Movsas B, Siddiqui MS, et al. Dosimetric evaluation of synthetic CT relative to bulk density assignment-based magnetic resonance-only approaches for prostate radiotherapy. *Radiat Oncol.* (2015) 10:239. doi: 10.1186/s13014-015-0549-7

10. Largent A, Barateau A, Nunes J-C, Lafond C, Greer PB, Dowling JA, et al. Pseudo-CT generation for MRI-only radiation therapy treatment planning: comparison among patch-based, atlas-based, and bulk density methods. *Int J Radiat Oncol Biol Phys.* (2019) 103:479–90. doi: 10.1016/j.ijrobp.2018.10.002

11. Siversson C, Nordström F, Nilsson T, Nyholm T, Jonsson J, Gunnlaugsson A, et al. MRI only prostate radiotherapy planning using the statistical decomposition algorithm. *Med Phys.* (2015) 42:6090–7. doi: 10.1118/1.4931417

12. Persson E, Gustafsson C, Nordström F, Sohlin M, Gunnlaugsson A, Petruson K, et al. MR-OPERA: a multicenter/multivendor validation of magnetic resonance imaging–only prostate treatment planning using

synthetic computed tomography images. *Int J Radiat Oncol Biol Phys.* (2017) 99:692–700. doi: 10.1016/j.ijrobp.2017.06.006

13. Dowling JA, Lambert J, Parker J, Salvado O, Fripp J, Capp A, et al. An atlas-based electron density mapping method for magnetic resonance imaging (MRI)-alone treatment planning and adaptive MRI-based prostate radiation therapy. *Int J Radiat Oncol Biol Phys.* (2012) 83:e5–11. doi: 10.1016/j.ijrobp.2011.11.056

14. Dowling JA, Sun J, Pichler P, Rivest-Hénault D, Ghose S, Richardson H, et al. Automatic substitute computed tomography generation and contouring for magnetic resonance imaging (MRI)-alone external beam radiation therapy from standard MRI sequences. *Int J Radiat Oncol Biol Phys.* (2015) 93:1144–53. doi: 10.1016/j.ijrobp.2015.08.045

15. Andreasen D, Van Leemput K, Edmund JM. A patch-based pseudo-CT approach for MRI-only radiotherapy in the pelvis. *Med Phys.* (2016) 43:4742–52. doi: 10.1118/1.4958676

16. Korhonen J, Kapanen M, Keyriläinen J, Seppälä T, Tenhunen M. A dual model HU conversion from MRI intensity values within and outside of bone segment for MRI-based radiotherapy treatment planning of prostate cancer. *Med Phys.* (2014) 41:011704. doi: 10.1118/1.4842575

17. Maspero M, Savenije MHF, Dinkla AM, Seevinck PR, Intven MPW, Jurgenliemk-Schulz IM, et al. Dose evaluation of fast synthetic-CT generation using a generative adversarial network for general pelvis MR-only radiotherapy. *Phys Med Biol.* (2018) 63:185001. doi: 10.1088/1361-6560/aada6d

18. Arabi H, Dowling JA, Burgos N, Han X, Greer PB, Koutsouvelis N, et al. Comparative study of algorithms for synthetic CT generation from MRI: consequences for MRI-guided radiation planning in the pelvic region. *Med Phys.* (2018) 45:5218–33. doi: 10.1002/mp.13187

19. Johnstone E, Wyatt JJ, Henry AM, Short SC, Sebag-Montefiore D, Murray L, et al. Systematic review of synthetic computed tomography generation methodologies for use in magnetic resonance imaging–only radiation therapy. *Int J Radiat Oncol Biol Phys.* (2018) 100:199–217. doi: 10.1016/j.ijrobp.2017.08.043

20. Edmund JM, Nyholm T. A review of substitute CT generation for MRI-only radiation therapy. *Radiat Oncol.* (2017) 12:28. doi: 10.1186/s13014-016-0747-y

21. Wang D, Doddrell DM. Geometric distortion in structural magnetic resonance imaging. *Curr Med Imaging Rev.* (2005) 1:49–60. doi: 10.2174/1573405052953029

22. Wang H, Balter J, Cao Y. Patient-induced susceptibility effect on geometric distortion of clinical brain MRI for radiation treatment planning on a 3T scanner. *Phys Med Biol.* (2013) 58:465–77. doi: 10.1088/0031-9155/58/3/465

23. Nejad-Davarani SP, Sevak P, Moncion M, Garbarino K, Weiss S, Kim J, et al. Geometric and dosimetric impact of anatomical changes for MR-only radiation therapy for the prostate. *J Appl Clin Med Phys.* (2019) 20:10–7. doi: 10.1002/acm2.12551

24. Kim J, Miller B, Siddiqui MS, Movsas B, Glide-Hurst C. FMEA of MR-only treatment planning in the pelvis. *Adv Radiat Oncol.* (2019) 4:168–76. doi: 10.1016/j.adro.2018.08.024

25. Korsholm ME, Waring LW, Edmund JM. A criterion for the reliable use of MRI-only radiotherapy. *Radiat Oncol.* (2014) 9:16. doi: 10.1186/1748-717X-9-16

26. Greer P, Pichler P, Young T, Martin J, Hunter P, Wratten C, et al. PV-0428 A multi-centre study for implementation of MRI-only prostate planning. *Radiother Oncol.* (2019) 133:S224–S5. doi: 10.1016/S0167-8140(19)30848-5

27. Kim WH, Kim CG, Kim D-W. Optimal CT number range for adipose tissue when determining lean body mass in whole-body F-18 FDG PET/CT studies. *Nucl Med Mol Imaging.* (2012) 46:294–9. doi: 10.1142/8245

28. Thomas S. Relative electron density calibration of CT scanners for radiotherapy treatment planning. *Br J Radiol.* (1999) 72:781–6. doi: 10.1259/bjr.72.860.10624344

29. Low DA, Harms WB, Mutic S, Purdy JA. A technique for the quantitative evaluation of dose distributions. *Med Phys.* (1998) 25:656–61. doi: 10.1118/1.598248

30. Vinod SK, Jameson MG, Min M, Holloway LC. Uncertainties in volume delineation in radiation oncology: a systematic review and recommendations for future studies. *Radiother Oncol.* (2016) 121:169–79. doi: 10.1016/j.radonc.2016.09.009

31. Koivula L, Kapanen M, Seppälä T, Collan J, Dowling JA, Greer PB, et al. Intensity-based dual model method for generation of synthetic CT images from standard T2-weighted MR images–Generalized technique for four different MR scanners. *Radiother Oncol.* (2017) 125:411–9. doi: 10.1016/j.radonc.2017.10.011

32. Köhler M, Vaara T, Van Grootel M, Hoogeveen R, Kemppainen R, Renisch S. MR-only simulation for radiotherapy planning. In: *White Paper: Philips MRCAT for Prostate Dose Calculations Using Only MRI Data.* Amsterdam: Koninklijke Philips N.V. (2016), 1–16.

Super-Resolution [1]H Magnetic Resonance Spectroscopic Imaging Utilizing Deep Learning

Zohaib Iqbal[1], Dan Nguyen[1], Gilbert Hangel[2], Stanislav Motyka[2], Wolfgang Bogner[2] and Steve Jiang[1]*

[1] Medical Artificial Intelligence and Automation Laboratory, Department of Radiation Oncology, University of Texas Southwestern Medical Center, Dallas, TX, United States, [2] Christian Doppler Laboratory for Clinical Molecular MR Imaging, Department of Biomedical Imaging and Image-guided Therapy, High Field MR Center, Medical University of Vienna, Vienna, Austria

Magnetic resonance spectroscopic imaging (SI) is a unique imaging technique that provides biochemical information from *in vivo* tissues. The ^{1}H spectra acquired from several spatial regions are quantified to yield metabolite concentrations reflective of tissue metabolism. However, since these metabolites are found in tissues at very low concentrations, SI is often acquired with limited spatial resolution. In this work, we test the hypothesis that deep learning is able to upscale low resolution SI, together with the T1-weighted (T1w) image, to reconstruct high resolution SI. We report on a novel densely connected UNet (D-UNet) architecture capable of producing super-resolution spectroscopic images. The inputs for the D-UNet are the T1w image and the low resolution SI image while the output is the high resolution SI. The results of the D-UNet are compared both qualitatively and quantitatively to simulated and *in vivo* high resolution SI. It is found that this deep learning approach can produce high quality spectroscopic images and reconstruct entire ^{1}H spectra from low resolution acquisitions, which can greatly advance the current SI workflow.

Keywords: super-resolution, magnetic resonance spectroscopic imaging (SI), deep learning (DL), magnetic resonance spectroscopy (^{1}H MRS), artificial intelligence

1. INTRODUCTION

Magnetic resonance imaging (MRI) continues to be a versatile modality capable of providing anatomical, metabolic, and functional information from various regions of the body *in vivo*. In particular, magnetic resonance spectroscopic imaging (SI) (1) is able to yield important data regarding the metabolism of different tissues, and has been especially useful for studying the metabolism of the human brain (2). Some important biochemicals, or metabolites, in the brain include N-acetyl aspartate (NAA), glutamate (Glu), glutamine (Gln), creatine (Cr), choline (Ch), and myo-Inositol (mI) (3). Each metabolite plays an important role in regulating energy consumption in the brain, and some metabolites also play critical functional roles, including roles as neurotransmitters (4). It is well-known that metabolic changes occur in parallel with anatomical changes for a myriad of pathologies (2), and these metabolic changes may even occur before structural changes are detected. While SI has continued to be an active area of research over the past several decades, there are still major roadblocks into standardizing this technique and including it into clinical protocols.

One of the major disadvantages of SI is the long acquisition duration associated with obtaining spectra from several voxels of interest. This is primarily due to the fact that many of the important metabolites are found in the brain at low concentrations; these metabolites are typically present in the brain at 1–12 mM concentrations (3). Therefore, in order to accurately detect these biochemicals, several signal averages have to be obtained or larger voxel volumes have to be acquired to improve the signal to noise ratio (SNR) for the experiment. As a result, spatial resolution tends to be coarse for many SI sequences. This low resolution, coupled with other technical problems such as partial volume effects, hinders the overall diagnostic capabilities of the SI technique.

There have been many advances in the technological implementation of SI that allow for faster acquisition and better spatial resolution. One of the primary acceleration methods is echo planar spectroscopic imaging (EPSI) (5, 6), which collects spectral data from an entire line of k-space in a single repetition time (TR) utilizing an echo planar readout. This spatio-spectral acquisition approach has also been applied in non-cartesian SI methods, such as spiral acquisitions (7), concentric circular acquisitions (8), and rosette acquisitions (9). In addition, parallel imaging (10–12) can also be used to accelerate the collection of SI data. Sensitivity encoding (SENSE) has been applied in combination with EPSI (13) to facilitate even faster acquisition times. Recently, research has also focused on the application of various sampling schemes that allow for reduced scan time (14–18). Some studies (19, 20) have even demonstrated protocols capable of obtaining spectroscopic images at 64x64 or 128x128 resolution in less than 20 min. Although these advances have improved the field significantly, SI is still understandably seen as a low SNR, low resolution technique.

In order to combat the limits of the experimentally acquired resolution, many post-processing methods have been developed for super-resolution SI (21–27). These methods have mainly focused on model-based reconstruction methods and regularized reconstruction approaches. While many super-resolution methods are independent of the acquisition protocols, there are some techniques, such as the spectroscopic imaging by exploiting spatio-spectral correlation (SPICE) method (18), that show reconstruction benefits by employing inter-dependent sequences. Unfortunately, the majority of super-resolution methods either tend to be very complicated to implement, or generally show poor reconstruction results. Since experimental acquisitions have many technical challenges, there is also a large concern over the true gold standard for these super-resolution techniques. Without a true standard of comparison, which is a large problem in the spectroscopic imaging field, many studies qualitatively and quantitatively compare their methods with less ideal standards such as bicubic interpolation.

Deep learning is an advancing field that has shown extraordinary results for image processing (28–30). Convolutional layers and networks are capable of extracting valuable features from images, and can further process these features into labels or other images for classification, segmentation, and other uses. One network that has been extremely beneficial for the field of automated medical imaging

segmentation is the UNet (31), which allows for a pixel-wise transformation of an input image into an output image. Essentially, deep learning excels at computing an unknown transformation by using a large example dataset, often referred to as a training set. We hypothesize that UNet, or some other deep neural networks are able to upscale low resolution SI (LRSI), together with the T1-weighted (T1w) image, to produce high resolution SI (HRSI). To test this hypothesis the biggest challenge is that a large, publicly available SI dataset is unavailable and difficult to acquire experimentally. In order to create this data set, HRSI (128x128 pixels) and LRSI (16x16 pixels or some other low resolution) experiments would have to be performed on thousands of diverse patients with different pathologies, which is not feasible. Thus, it is seemingly impossible to perform deep learning for super-resolution SI.

In this paper, we report a novel work on the development of a deep learning technology capable of producing super-resolution spectroscopic images. An SI generator is used to produce LRSI and HRSI data in order to train and test a deep learning model. Using this data, a UNet taking advantage of densely connected layers (D-UNet) is built and trained. The inputs for the D-UNet are the T1w image and the low resolution SI image while the output is the high resolution SI. The results of the D-UNet reconstruction are compared both qualitatively and quantitatively to simulated and *in vivo* high resolution SI data.

2. METHODS

2.1. Spectroscopic Imaging Dataset

Two different MRI data sets were utilized to produce synthetic SI data for developing the deep learning model. The first MRI data set comprised of 27 axial slices from the MATLAB MRI dataset. MR images as well as white matter (WM) and gray matter (GM) masks from the open access series of imaging studies (OASIS) project (32), which contained 416 axial images from subjects ranging in age from 18 to 96 years old, were also used. From these limited data, 102,169 SI datasets were synthesized using an SI generator, the details of which are found below.

2.2. Spectroscopic Imaging Generator

The SI generator was designed to address the lack of T1w images, as well as the lack of paired LRSI and HRSI data. First, the generator created augmented T1w ($aT1w$), white matter (WM), and gray matter (GM) images from an input T1w image. Then, the generator would produce a matched LRSI and HRSI for the $aT1w$ image.

2.2.1. Augmenting T1w Images

An input T1w image is first segmented into WM and GM images via an intensity based approach. First, the maximum WM intensity (WM_{max}), and the minimum GM intensity (GM_{min}) are determined from the image. Then, WM and GM images are made by applying the following:

$$WM = \left(\frac{S - GM_{min}}{WM_{max}} \right) \cdot M \tag{1}$$

$$GM = \left(1 - WM \right) \cdot M \tag{2}$$

Above, S is the original signal intensity of the input T1w image, and M is a mask for the brain region only, and is applied through an element-wise multiplication. The above equations ensure that the elements of both the WM and GM images range from zero to one, and are representative of the percentage of WM or GM present in any voxel.

Then, the SI generator modifies the input T1w image to produce an $aT1w$ image. The contrast of the T1w image is altered by the following:

$$aT1w = R\left(S_n^{r_1} + L\right) \tag{3}$$

Here, S_n is the normalized input T1w signal and r_1 is a random number between 0.5 and 2.5. $R()$ is a rotation and field of view (FOV) truncation transformation that rotates the image randomly in the range of $-15°$ to $15°$ and randomly truncates the image in the range of 0 to 40 pixels in any direction. L is a matrix that represents up to 6 lesions of varying intensity, location, and size. Since this lesion matrix is random, the $aT1w$ image may or may not contain any hyper-intense or hypo-intense regions. The same transformation used in Equation (3) is also applied to the WM and GM images.

2.2.2. Production of Matched LRSI and HRSI Maps

In order to produce data useful for clinical applications, the SI generator operated under an assumption that is biologically valid: WM and GM regions of the brain have metabolism associated with biochemical concentrations (33). With this assumption, a given metabolite could be more concentrated in WM vs. GM, less concentrated in WM vs. GM, or equally concentrated in WM and GM regions.

Working with this biological assumption, a high resolution metabolite map is generated by adding a random ratio of the WM and GM images together:

$$HRSI = r_2 * WM + (1 - r_2) * GM + B + r_3 * L \tag{4}$$

In Equation (4), r_2 is a random number between 0 and 1. B is a matrix that adds a random signal bias into the metabolite map, which helps to simulate the presence of more metabolite signal from the anterior or posterior, as well as the left or right brain regions. L is the same lesion matrix used in Equation (3), and r_3 is a random number between -1 and 1.

Finally, the HRSI is downsampled to the desired low resolution via k-space truncation. Random noise is also added to this low resolution k-space data before a Fourier transformation is used to bring this data back to the spatial domain. Next, the low resolution image is upscaled to the same resolution as the HRSI using nearest-neighbor interpolation to yield the final low resolution SI.

It is important to note that because of the variables r_1, r_2, r_3, and L, it is possible to produce several different matched $aT1w$ images, HRSI, and LRSI from the same input T1w image. In addition, the same $aT1w$ image can give rise to a large number of matched HRSI and LRSI, and thus this transformation is a one to many transformation. Therefore, a single input T1w image can produce hundreds of unique datasets for training a deep learning model.

2.3. Densely Connected UNet (D-UNet) Architecture and Training

The UNet architecture (31) is typically implemented for segmentation purposes, however it primarily operates by performing pixel-wise transformations on input images, which is applicable to the SI super-resolution problem. Using standard convolutional and max pooling layers, the UNet first continuously convolves and pools the input image until the image reaches a small size, which aids in extracting valuable global features. Next the image is scaled up through a combination of up-pooling, transpose convolutions, and feature concatenations. This second process helps to identify vital local features so that the UNet can refine the image at a finer resolution. However, due to the number of features necessary for this process, the classical UNet suffers from extremely long training times, overfitting issues, and potential inefficiencies when tuning the weights. Therefore, this study utilized densely connected convolutional layers (34) to develop the novel densely connected UNet (D-UNet) architecture, and the workflow for training is shown in **Figure 1**. Densely connected networks carry over features from layer to layer, allowing for all previous information to be used for determining important features. The general architecture of the D-UNet used in this study is shown in **Figure 2**. The D-UNet utilized 32 feature maps at every max pooling layer. In addition, all convolutional layers made use of the ReLU activation function (30) and used a dropout (35) of 0.1. Certain features, shown in green and orange in **Figure 2**, were copied over to the following layers, and were also concatenated later on in the network. In total, three max pooling layers were used for the D-UNet. Since low resolution SI experiments can have diverse resolutions, three identical D-UNets were made to upscale low resolution spectroscopic images for acquisitions with 16x16, 24x24, and 32x32 spatial points.

The D-UNet required two inputs: a rescaled (128x128 points) T1w image and the corresponding LRSI image (16x16, 24x24, or 32x32 points) upscaled using nearest-neighbor interpolation (128x128 points). The predicted output of the D-UNet was a denoised HRSI image (128x128 points). For training, $aT1w$, HRSI, and LRSI were created from the SI generator, as described above. The Adam optimizer (36) was used with a learning rate set to 1×10^{-3}, and mean squared error (MSE) was used as the cost function, which determined the difference between the D-UNet output and the desired output:

$$MSE = \sum \sum \frac{(O - HRSI)^2}{m^2} \tag{5}$$

Above, O is the output of the D-UNet, $HRSI$ is the true simulated high resolution SI, and m is the output dimension of the network, which in this case is 128. The summations are performed over both dimensions to yield a single value. The network was trained on an 8GB Quadro K5200 graphical processing unit (GPU) using the Keras (37) and Tensorflow (38) packages in Python 3.6.

Two datasets were made for the development and evaluation of the three D-UNets: a training dataset and a testing dataset. The training dataset comprised of 102,000 data from the SI generator using 135 axial images. The testing dataset used 169 different
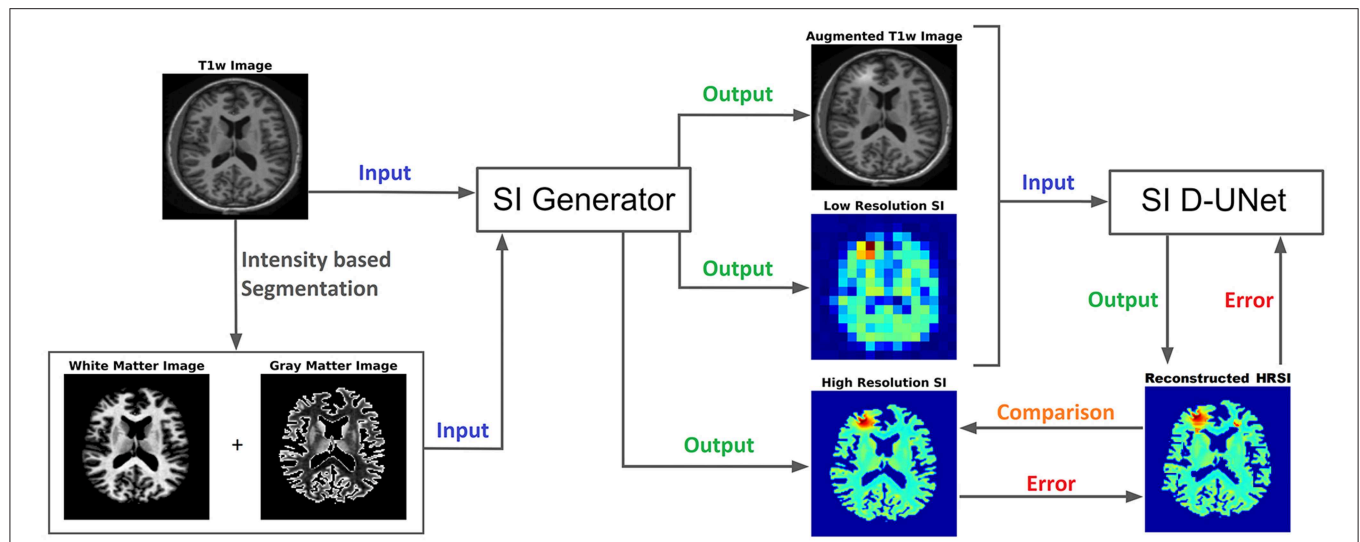
FIGURE 1 | The workflow for training the D-UNet model is shown. The SI generator provides a dataset consisting of an augmented T1w image, a low resolution spectroscopic image, and a ground truth high resolution spectroscopic image. The spectroscopic images already show the distribution of a particular metabolite (or the distribution of a particular spectral point), such as choline, and therefore do not contain a spectral dimension. Then, the network transforms the $aT1w$ (128x128 pixels) and LRSI (128x128 pixels after nearest-neighbor interpolation) into an initial HRSI reconstruction (128x128 pixels). In the example above, the LRSI and HRSI reconstruction have in-plane spatial resolutions of 1.4×1.4 cm^2 and 1.7×1.7 mm^2, respectively. This reconstruction is compared to the ground truth, and the mean squared error is calculated. Utilizing this error, the model changes the weighting parameters for the features, and continues training by using a different dataset. After training on 102,000 datasets, the model weights are refined and the reconstruction errors are minimized.

FIGURE 2 | The general D-UNet architecture is displayed. Each forward convolution consisted of a convolutional layer and a concatenation process. This concatenation carries over important features which can be used to make the next layer more intelligent. In addition to local concatenations, certain features were concatenated to deeper layers in the network. More specifically, every feature map that is produced from a convolution is carried over to the end. Maxpooled features are not, since a higher resolution of the feature already exists. This allows for prior information to improve the overall reconstruction quality. In order to use the most information possible, the last convolutional layer contains all of the carried over features.

axial images (independent from the training set) from the OASIS project, and 169 matched $aT1w$, HRSI, and LRSI images were produced via the SI generator. Each of the three D-UNets were trained for a total of 102 epochs. For this study, an epoch was defined as a new set of 1,000 matched HRSI and LRSI data. The first two epochs were trained using a batch size of one to ensure that the network would not fall into a local minimum. The remaining 100 epochs were trained with a batch size of 10. Varying batch size in this manner has been shown to help reduce the number of epochs necessary for training, while also reducing the need for hyper-parameter tuning (39).

2.4. D-UNet Evaluation and Comparison Metrics

2.4.1. Testing Set Evaluation

The three D-UNets evaluated all 169 matched images ($aT1w$ and LRSI) to produce reconstructed high resolution spectroscopic images ($Recon_{16x16}$, $Recon_{24x24}$, and $Recon_{32x32}$). These reconstructed images were compared to the ground truth HRSI using mean squared error. This process was repeated with varying noise levels inserted into the input LRSI in order to determine the role of noise on the reconstruction process. Example low resolution spectroscopic images can be seen in **Figure 3**. The reconstructed images were also compared to zero-filling and bicubic interpolation to assess the improvement of the D-UNet results over standard methods. For this comparison, both zero-filling and bicubic interpolation were applied to an LRSI of 32x32 points to generate the 128x128 interpolated images.

2.4.2. Spectral Reconstruction Evaluation

In addition, the three D-UNets were used to reconstruct magnitude spectra point-by-point from low spatial resolution to high spatial resolution. Magnitude spectra were used because the model was not trained for evaluating real and imaginary numbers simultaneously. From the test set, a single subject was used to generate high resolution chemical maps of the major metabolites, including NAA, Glu, Gln, Cr, Ch, and mI. GAMMA simulation (40) was used to simulate the spectra for these metabolites using an echo time (TE) = 30 ms, spectral bandwidth of 2,000 Hz, and time points = 512 for a magnetic field strength (B_0) of 3T. Also, the spectra were exponentially line broadened to roughly 8 Hz. These spectra were then distributed spatially based on their respective high resolution maps, and were transformed to produce LRSI. The T1w image and LRSI were input into the three D-UNets to produce $Recon_{16x16}$, $Recon_{24x24}$, and $Recon_{32x32}$ spectral data. Two example spectra were extracted from these reconstructed images and compared to the simulated ground truth using mean squared error.

2.4.3. *In vivo* Evaluation

Finally, high resolution spectroscopic images were acquired on a 7T whole-body MR scanner (Magnetom, Siemens Healthcare, Erlangen, Germany) using a previously published protocol (20). The Institutional Review Board (IRB) at the Medical University of Vienna approved the study and ten healthy volunteers (mean age = 31.7 years old) signed written and informed consent

forms. All experiments were performed in accordance with relevant guidelines and regulations. The protocol utilized free induction decay based MR spectroscopic imaging (41) with TR = 200 ms for a total scan time of 21 min. After acquisition, residual lipids were removed using ℓ_2 regularization (42) and the spectra were quantified using the LCModel (43) package to yield concentrations for several metabolites. Therefore, high resolution (128x128 pixels, 1.7×1.7 mm^2) metabolite maps for NAA, Cr, Ch, Glu, Gln, and mI were obtained. These metabolite maps were down-sampled to 32x32 resolution images and were input into the 32x32 D-UNet along with corresponding T1w images to yield $Recon_{32x32}$ for all datasets. These reconstructed images were then compared to the experimentally acquired HRSI using mean squared error as described in Equation (5). In addition, Glu/Cr and Ch/Cr ratios for both the reconstructed and experimentally acquired images were measured over all ten subjects. These ratios were investigated as a function of T1w intensity, which directly corresponds to the ratio of WM and GM in the brain. Finally, correlations between the reconstructed and experimental results were performed to yield the correlation coefficients (r) for the Glu/Cr and Ch/Cr ratios.

3. RESULTS

3.1. Training Results

Due to the novel D-UNet architecture, the mean squared error loss rapidly converged close to a reasonable value after only 2 epochs for all three networks, and the loss functions are shown in **Figure 4**. The loss continued to decrease with more epochs when a larger batch size was used for the remaining 100 epochs. From **Figure 4**, it is clear that the final loss was better for the 32x32 D-UNet than the 24x24 or 16x16 D-UNets. This is theoretically expected because higher initial resolution should aid in the estimation of unknown points, and this is true for conventional resolution enhancement techniques as well. While a low dropout was used in the architecture, overfitting was not a primary concern for the D-UNet training framework because of the reduced number of weighting parameters in the model. The results from the testing dataset also highlight the fact that the D-UNet training was generalized and applicable to never before seen data.

3.2. Test Set Results

Figure 5 displays the results from the three different D-UNet reconstructions, as well as the results of the standard zero-filling and bicubic interpolation methods. In order to provide a more stringent comparison, both zero-filling and bicubic interpolation were applied to the 32x32 low resolution metabolite maps instead of the lower resolution 16x16 or 24x24 metabolite maps. All of the D-UNet reconstructions are able to determine the abnormally high signal from the lesion shown in the T1w image. While zero-filling outperforms both bicubic interpolation and the 16x16 D-UNet, both the 24x24 and 32x32 D-UNets yield better results than zero-filling.

To demonstrate the capability of the SI generator, **Figure 6** shows a sample of the possible images produced from the same $aT1w$ image. The $Recon_{32x32}$ images are also shown, as well as
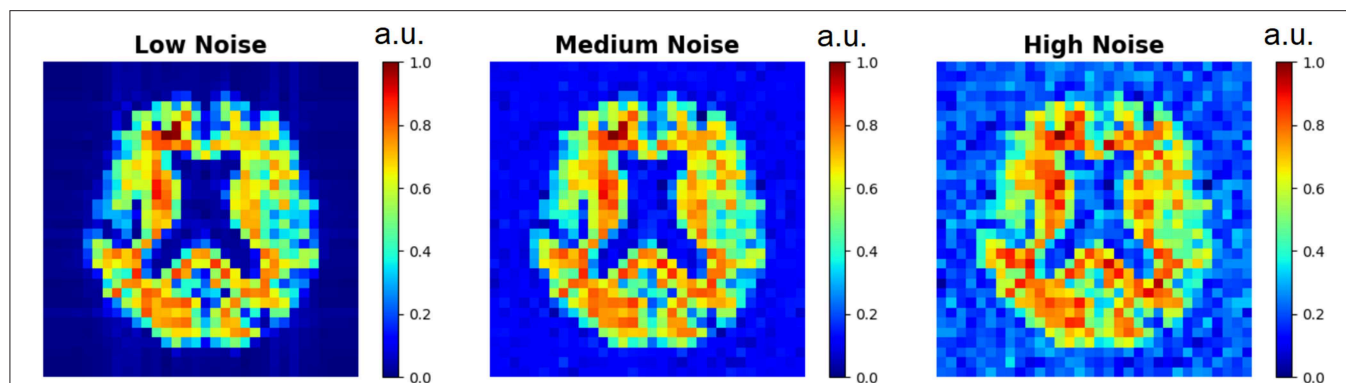
FIGURE 3 | Low resolution spectroscopic images generated using the SI generator are shown. To show the effect of different random noise levels, all other random parameters were the same between the three images. Low noise level, medium noise level, and high noise level were classified as 2–5, 15–20, and 30–40% of the maximum signal intensity, respectively.
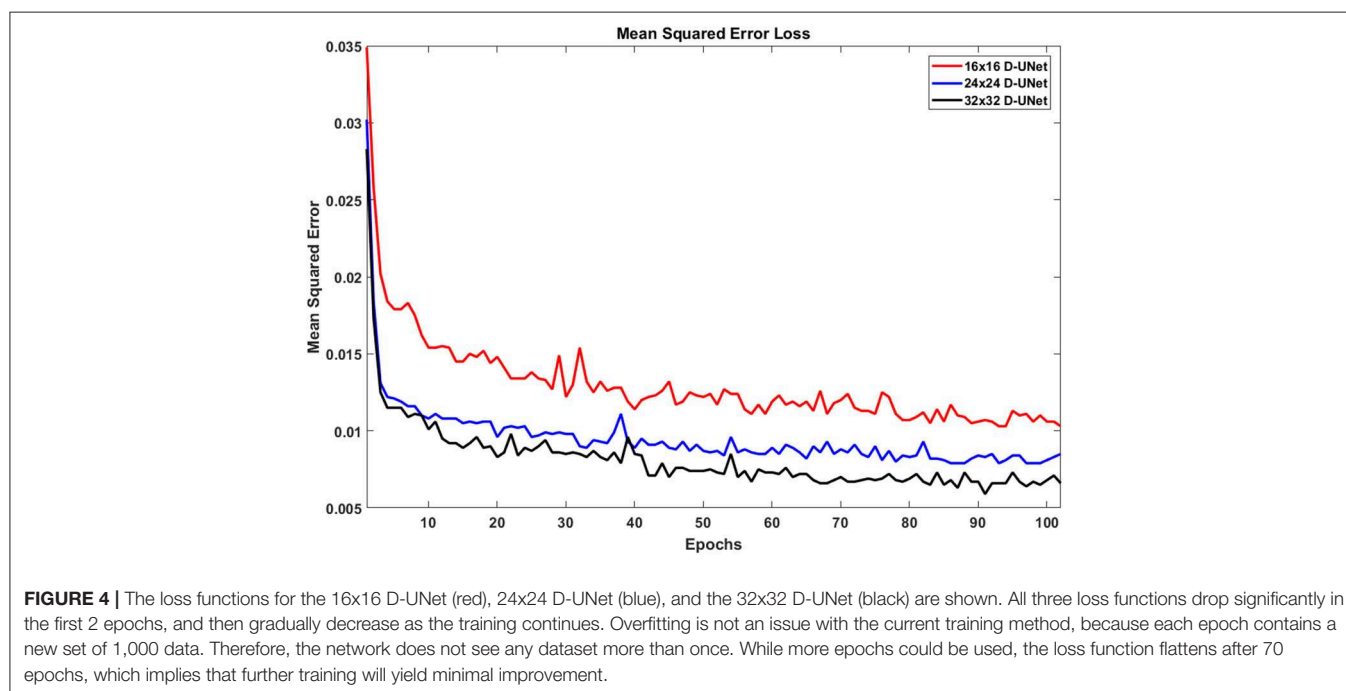
FIGURE 4 | The loss functions for the 16x16 D-UNet (red), 24x24 D-UNet (blue), and the 32x32 D-UNet (black) are shown. All three loss functions drop significantly in the first 2 epochs, and then gradually decrease as the training continues. Overfitting is not an issue with the current training method, because each epoch contains a new set of 1,000 data. Therefore, the network does not see any dataset more than once. While more epochs could be used, the loss function flattens after 70 epochs, which implies that further training will yield minimal improvement.

difference maps between the HRSI and $Recon_{32x32}$. It is clear that the SI generator is capable of producing a wide variety of *SI* images that mimic biochemicals that are more prominent in GM, more prominent in WM, or equally prominent in both tissue types.

In addition, a quantitative comparison between these methods is shown in **Table 1**. Noise level was varied to determine the effect of noise on the super-resolution methods. Low noise level, medium noise level, and high noise level were classified as 2–5, 15–20, and 30–40% of the maximum signal intensity, respectively. From **Table 1**, the 32x32 D-UNet demonstrated the best performance at every noise level. At medium noise levels, the 24x24 D-UNet outperformed zero-filling, and at high noise levels both the 16x16 D-UNet and 24x24 D-UNet outperformed both zero-filling and bicubic interpolation.

3.3. Spectral Reconstruction Results

The ability of the D-UNets to reconstruct spectra at high spatial resolutions are highlighted in **Figure 7**. The 32x32 D-UNet reconstructs the lesion and contra-lateral white matter spectra reliably. In contrast, the 16x16 D-UNet underestimates the white matter spectrum. The 24x24 D-UNet performs very similarly to the 32x32 D-UNet, however it overestimates the Ch and mI signals in the lesion spectrum by roughly 20%. Overall, the mean squared error for the healthy white matter spectrum was 0.34, 0.030, and 0.0085 for the 16x16 D-UNet, 24x24 D-UNet, and 32x32 D-UNet, respectively. For the lesion spectrum, the mean squared error was 0.051, 0.36, and 0.13 for the 16x16 D-UNet, 24x24 D-UNet, and 32x32 D-UNet, respectively. From a quantitative standpoint, all three D-UNets would be able to

FIGURE 5 | The results of the three D-UNets are shown for an example test subject. The augmented T1w image is used in conjunction with the three low resolution images as inputs for the three D-UNets. The reconstructed HRSI (16x16 High Res, 24x24 High Res, and 32x32 High Res) are shown below their respective low resolution images. In addition, zero-filling and bicubic interpolation were applied to the 32x32 LRSI to produce 128x128 interpolated images. Error maps are produced by subtracting the reconstructed images and the ground truth high resolution image (High Res). The 16x16 High Res displays much more error than the 24x24 and 32x32 High Res images. This is mostly due to better local signal refinement at the location of the lesion for the 24x24 and 32x32 reconstructions.

determine the abnormally elevated Ch, as demonstrated from the metabolite maps.

3.4. *In vivo* Results

The ability of the 32x32 D-UNet to reconstruct the LRSI of Cr, NAA, Glu, Gln, Ch, and mI for the *in vivo* data is shown in **Figure 8**. This figure shows the reconstructed images, experimental HRSI, and difference maps between the two for each metabolite for one healthy volunteer. All reconstructed images retain the metabolite signals from the low resolution maps, and also show regional changes similar to the HRSI. For example, Glu is more concentrated in the GM and less concentrated in the WM, which is a well-known regional difference in the brain (33). Another well-known regional difference is that Ch is more concentrated in WM regions, which is apparent in both the reconstructed and experimental images. **Figure 9** shows reconstructions with low, average, and large MSE values. In general, lower SNR metabolites appeared to have a larger MSE value compared to higher SNR metabolites. From a quantitative standpoint, the average MSE values over the ten volunteers for Cr, NAA, Glu, Gln, Ch, and mI were 0.0048, 0.0042, 0.0060, 0.0079, 0.0059, and 0.0056 respectively. These errors are displayed in **Figure 10D** and plotted against the average MSE values obtained for the testing set using different

noise levels (low, medium, high). It is clear that the MSE values are in most cases comparable to simulated test images with 2–20% noise, with the exception of Gln which is most comparable to test images with 35% noise.

Figure 10 also shows the Glu/Cr and Ch/Cr ratios as a function of the T1w intensity averaged over the ten volunteers. The ratios are taken after normalization of the metabolites as part of the super-resolution reconstruction, which is why Ch/Cr appears larger than Glu/Cr in the figure. The trend shows that with higher WM content, Glu/Cr decreases while Ch/Cr increases. The correlation between the experimental HRSI and Recon results are shown in **Figure 10C**. Quantitatively, both Glu/Cr and Ch/Cr ratios have high squared correlation coefficients, $r^2 > 0.99$. This highlights the fact that important biological relationships are preserved in the reconstructed images.

4. DISCUSSION

Although SI provides invaluable information regarding the biomolecular processes of tissues *in vivo*, experimental limitations have greatly hindered the integration of this method into standard clinical protocols. This study demonstrates a technique capable of overcoming one of the greatest challenges

FIGURE 6 | From one augmented T1w image, the generator is capable of producing multiple ground truth high resolution (High Resolution SI) images and low resolution (Low Resolution SI) images, a small sample of which are shown. In this example, the top row shows images where metabolite signal is higher in WM. In the middle row, the metabolite signal is equal in WM and GM, whereas the metabolite signal is higher in the GM in the bottom row. Since a single input T1w image can produce many augmented T1w images, the generator allows for an exponentially large number of unique training data. The reconstruction for each aT1w image and LRSI is performed with the 32x32 D-UNet to yield the reconstructed HRSI images (Reconstructed). The difference maps are produced by subtracting the reconstructed and ground truth images.

TABLE 1 | The mean squared error between the high resolution ground truth (HRSI) and several methods are tabulated.

Method	Noise Level		
	Low	**Medium**	**High**
Zero-Fill from 32x32	1.109	1.652	4.505
Bicubic from 32x32	2.794	3.129	3.820
16x16 D-UNet	1.863	2.420	2.761
24x24 D-UNet	1.139	1.316	1.745
32x32 D-UNet	**0.7460**	**0.9722**	**1.599**

These values are the total sum of the mean squared error over 169 test subjects. The 32x32 D-UNet reconstruction outperforms all of the other methods. With higher random noise present in the LRSI, the 16x16 and 24x24 D-UNets outperform both zero-filling and bicubic interpolation. It is important to note that this is true even though the zero-filling and bicubic interpolation methods are applied to a 32x32 image. Bold values indicate the method with the lowest mean squared error for each comparison.

in SI, which is poor spatial resolution. By utilizing a deep learning framework, it is shown in **Figures 5–9** that high resolution spectroscopic images can be produced from the combination of low resolution spectroscopic images and T1w images. In addition, as seen in **Figure 7**, it is possible to reconstruct spectra at higher spatial resolutions. The reconstruction method also preserves important regional metabolic differences and shows low errors for *in vivo* reconstructions, as shown in **Figure 10**. This deep learning super-resolution method was compared to both zero-filling and bicubic interpolation, and proved to be better than these methods for all noise levels.

Deep learning requires large datasets, which are not readily available for SI. Unfortunately, there is also a lack of ground truth for high resolution spectroscopic imaging due to the fact that experimental results may contain chemical shift displacement

FIGURE 7 | Magnitude spectral reconstructions using the three D-UNets are shown for two voxels. The voxel locations for the white matter and lesion spectra are displayed in the T1w image as red points. The spectra generated from the ground truth (red), the 16x16 (blue), 24x24 (green), and 32x32 D-UNets (black) for the 1.5–4.3 ppm range are displayed. For spatial comparison, the choline metabolite maps for each method are also shown. All metabolite maps are scaled from 0 to 1. The 24x24 and 32x32 D-UNet reconstructions over-estimate the amount of choline in the lesion. However, the 16x16 reconstruction under-estimates the amount of metabolite signal in the healthy white matter region.

artifacts, B_0 inhomogeneity issues, partial volume effects, low signal to noise ratios, water contamination, or other forms of signal contamination. It is also prohibitively long to scan at high resolution (128x128) without using several acceleration methods, making a ground truth impossible to obtain from the human brain with current technology. Therefore, an SI generator was developed to simulate training and testing data from a publicly available dataset. By including various probabilistic transformations, such as contrast variations, metabolic signal changes, and FOV variations, the SI generator was capable of providing a diverse and large dataset for the training of the three D-UNets. These data may not be entirely realistic, and this generator must be validated more rigorously in the future. For this study, the dataset does seem to be representative of real acquisitions, as seen from the *in vivo* results.

The $Recon_{32x32}$ and HRSI experimental images are very similar, as seen from **Figures 8–10**. The reconstructed images show better resemblance to the anatomical T1w images, including cerebral spinal fluid localization. However, both the $Recon_{32x32}$ and HRSI experimental images provide similar quantitative results, as seen in **Figure 10**. Theoretically, the $Recon_{32x32}$ images would require $\frac{1}{16}th$ to $\frac{1}{4}th$ the time to acquire, depending on the acceleration methods implemented. Therefore, it is important to note that aside from super-resolution, the D-UNet may also be used as a means to accelerate a spectroscopic

imaging protocol in the future. Additionally, the reconstructed *in vivo* images are denoised while retaining essential metabolic information for different tissues of the brain, which may be desirable for certain applications. While the simulated and *in vivo* data demonstrate that the reconstruction method is accurate, one of the main disadvantages of this work is that it has not been validated *in vitro*. This is due to the fact that a high resolution SI phantom similar to the human brain is not available. Since the D-UNet model is trained using *in vivo* anatomy, it is not capable of reconstructing high resolution images from unrealistic geometries. Therefore, future work will focus on the development of a realistic, high resolution SI phantom for validation.

Even though the D-UNets outperformed zero-filling and bicubic interpolation, these models may not be perfect for HRSI reconstruction primarily due to experimental imperfections. As seen from **Table 1**, error increases as a function of noise. Intuitively, chemicals that are found in the body at lower concentration may have larger reconstruction errors than chemicals with higher SNR, which is also supported by the *in vivo* results shown in **Figure 9** where the Gln reconstructed images have higher error than the other metabolite images. Therefore, prediction accuracy is limited by the quality of the original LRSI. Also, while the *in vivo* results have low mean-squared errors, it is important to note that down-sampling from a high resolution acquisition decreases potential acquisition problems
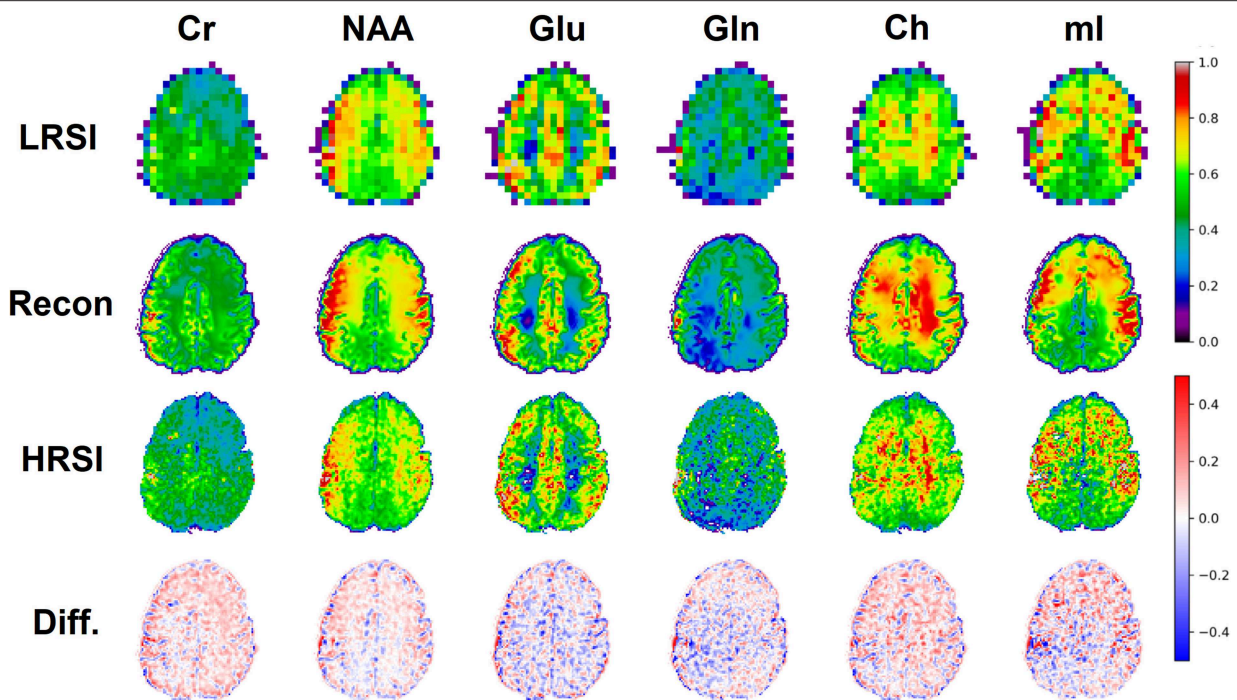
FIGURE 8 | An *in vivo* example of a healthy volunteer is used to demonstrate the potential application for the D-UNet. The experimental high resolution SI (HRSI) data was acquired at 128x128 resolution using an accelerated acquisition protocol (20). This data was then down-sampled to produce 32x32 low resolution SI (LRSI) metabolite maps for Cr, NAA, Glu, Gln, Ch, and ml. Together with the T1w image, the low resolution metabolite images were used to reconstruct high resolution spectroscopic images (Recon) using the 32x32 D-UNet model. The difference maps between the Recon and HRSI images (Diff) are also shown.
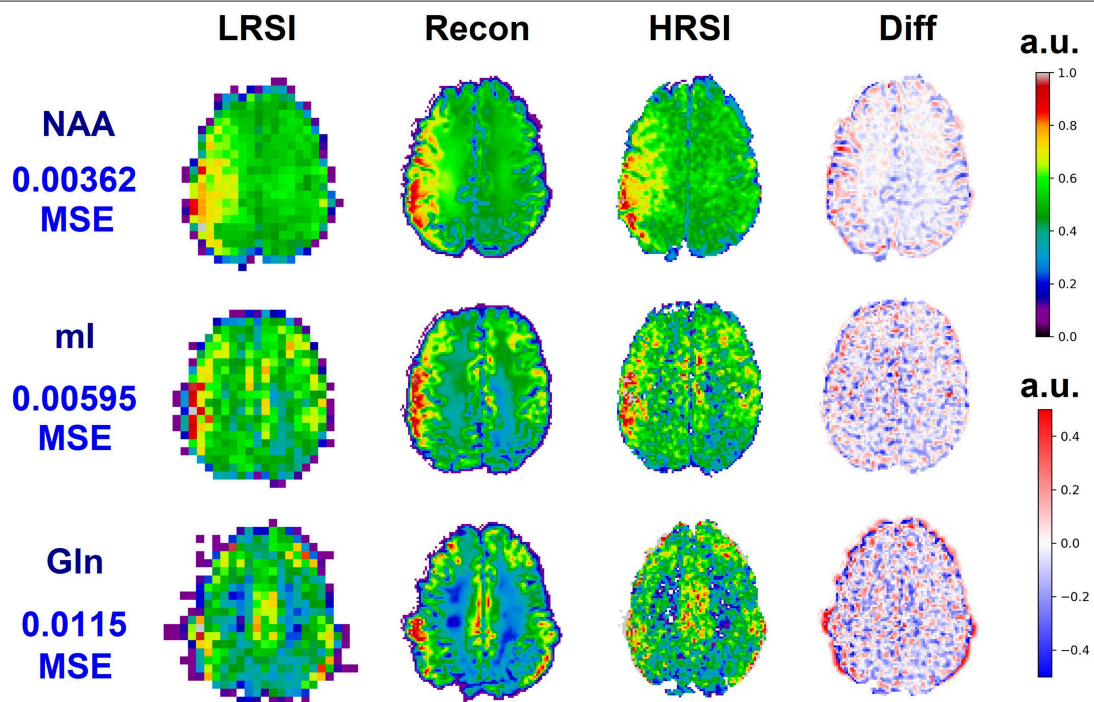
FIGURE 9 | Three reconstructions showing low MSE (top), average MSE (middle), and large MSE (bottom) are shown for three different volunteers. In general, lower MSE were observed for metabolites with higher signal-to-noise ratios such as NAA, whereas larger MSE values were calculated for low SNR metabolites such as Gln. The different MSE values are highlighted by the errors seen in the difference maps.
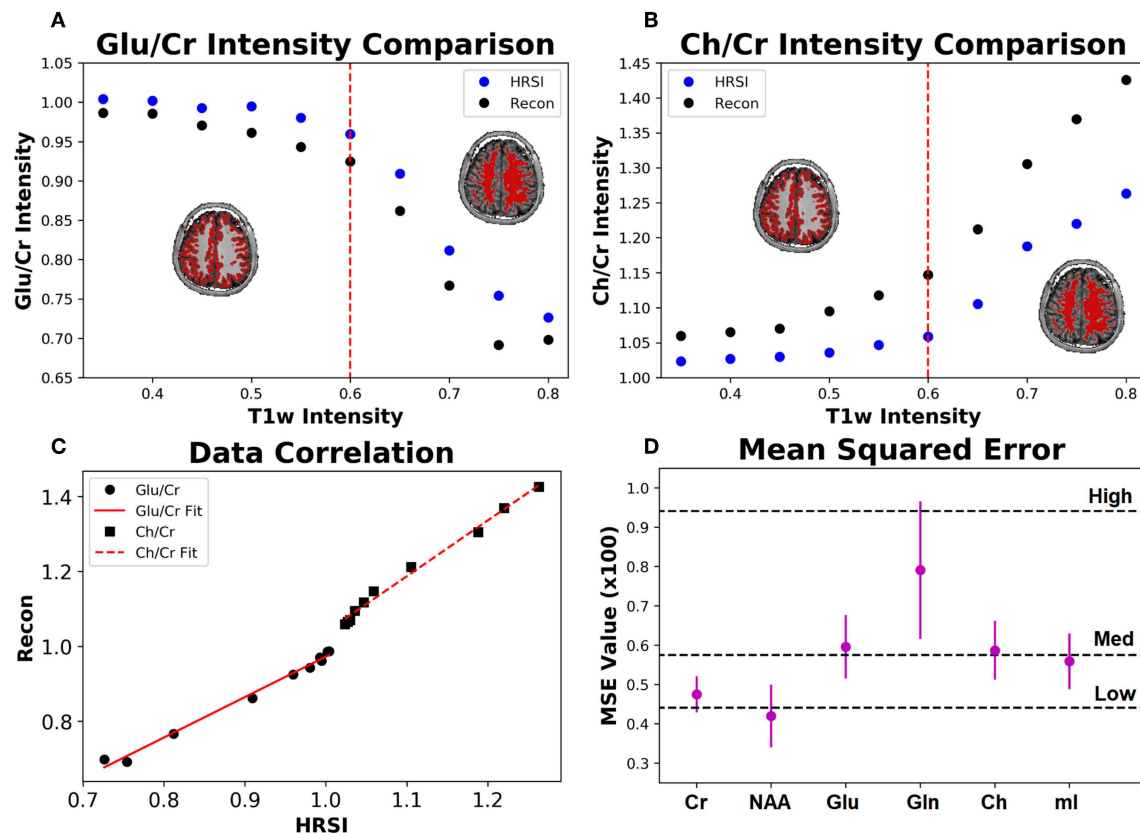
FIGURE 10 | The Glu/Cr **(A)** and Ch/Cr **(B)** signals averaged over ten subjects are shown for the experimentally acquired images (HRSI) and the images reconstructed from low resolution 32x32 images (Recon). The signals are shown as a function of the T1w intensity, which is representative of the gray and white matter content of the voxel. The red dotted line represents the point at which the gray matter content equals the white matter content in a voxel. The correlation between the HRSI and Recon values are plotted **(C)** with linear fits. For both Glu/Cr and Ch/Cr, the r^2 values of the fits are above 0.99. Finally, the mean squared error for the ten subjects calculated between the HRSI and Recon for each metabolite map **(D)** is displayed. The dotted black lines reflect the MSE from the testing set for different noise values (low, medium, and high).

such as lipid contamination and partial volume effects. Therefore, it is expected that a prospectively acquired low resolution data set will yield higher errors when reconstructed using the D-UNet. This must be evaluated in a more rigorous study where both low resolution and high resolution experimental SI data are acquired.

Of course, the original resolution of the experimental SI plays a large role in the reconstruction process. While 24x24 and 32x32 matrices provide relatively accurate high resolution reconstructions, the 16x16 resolution does not perform as well. This suggests that there is a lower bound necessary to accurately upscale high resolution SI. This might be true for other super-resolution techniques (21), so a more thorough comparison between this deep learning method and other methods may aid in identifying this lower bound. Furthermore, results may be biased by the quantitative methods implemented to produce the LRSI before the super-resolution process is performed. This bias could be removed in the future by developing a deep learning based approach to metabolite quantitation (44). However, it may be worthwhile to explore the differences between common one dimensional spectral quantitation

programs, such as LCModel (43) or TARQUIN (45), on the upscaling process.

From the spectral reconstruction results shown in **Figure 7**, it is apparent that some metabolites are over- and under- estimated during the reconstruction process. Therefore, clinical diagnosis based on the D-UNet reconstruction must be made with caution, as results from this method could lead to false positives or false negatives. Before basing diagnosis on the D-UNet reconstruction, the process should be evaluated *in vivo* in a well-known brain cancer pathology to assess the rates of false positives or false negatives detected by experienced radiologists in the field.

The deep learning method presented in this study may be useful for other super-resolution transformations in the field of medical imaging. This is especially true for spectroscopic imaging of other nuclei, such as ^{13}C and ^{31}P, where lower SNR results in low spatial resolution acquisitions. Recently, accelerated hyper-polarized ^{13}C spectroscopic imaging has shown to be promising for imaging prostate cancer (46, 47), and this technique could benefit by using the D-UNet model. In addition, ^{31}P spectroscopic imaging has also been used to image

cancer (48, 49). The main drawback, again, is the lack of SNR to adequately acquire high spatial resolution data. High resolution acquisition schemes have been proposed for ^{31}P spectroscopic imaging (50), and the D-UNet model could provide an alternative for improving spatial resolution. The same SI generation process could be used for training for these other nuclei, however different anatomical sites must be included (breast, prostate, etc.) to yield accurate results depending on the desired application.

The same principles discussed in this work may also apply to positron emission tomography (PET) (51). It is well-known that the radioactive tracer is more prominent in certain tissues and lesions, and positrons from this tracer travel some distance before annihilating to produce the PET signal. The distance between the source and the annihilation can be thought of as a partial volume effect. This model can potentially be used to learn how to remove this partial volume effect artifact, and this would be applicable for CT-PET or MR-PET acquisitions. Ultimately, this deep learning model allows for the acquisition of high quality images without increasing the scan time or improving the hardware of the imaging system.

5. CONCLUSION

The D-UNet model presented in this study allows for the reconstruction of accurate super-resolution magnetic resonance spectroscopic images from the human brain. Utilizing this method, we demonstrate that a simulated, low resolution chemical map can be transformed together with the T1w image to produce a high resolution chemical map. This method demonstrates better accuracy than typical zero-filling and bicubic interpolation methods. Furthermore, we demonstrate that the accuracy of this model holds when evaluating our method on retrospective *in vivo* data. This model still needs to be validated

on prospective *in vivo* data in the future. After further *in vitro* and *in vivo* validation, this method may be utilized for denoising, scan acceleration, and improved tissue delineation.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The Institutional Review Board (IRB) at the Medical University of Vienna approved the study and ten healthy, adult volunteers signed written and informed consent forms prior to imaging studies. All experiments were performed in accordance with relevant guidelines and regulations of the IRB.

AUTHOR CONTRIBUTIONS

ZI and SJ conceived the experiments. DN designed the deep learning architecture. ZI and DN conducted the deep learning experiments. GH, SM, and WB acquired and processed the *in vivo* data. ZI and SJ analyzed the results. All authors reviewed the manuscript.

FUNDING

REFERENCES

1. Brown T, Kincaid B, Ugurbil K. NMR chemical shift imaging in three dimensions. *Proc Natl Acad Sci USA.* (1982) 79:3523–6. doi: 10.1073/pnas.79.11.3523
2. Soares D, Law M. Magnetic resonance spectroscopy of the brain: review of metabolites and clinical applications. *Clin Radiol.* (2009) 64:12–21. doi: 10.1016/j.crad.2008.07.002
3. Govindaraju V, Young K, Maudsley AA. Proton NMR chemical shifts and coupling constants for brain metabolites. *NMR Biomed.* (2000) 13:129–53. doi: 10.1002/1099-1492(200005)13:3<129::AID-NBM619>3.0.CO;2-V
4. Ramadan S, Lin A, Stanwell P. Glutamate and glutamine: a review of *in vivo* mrs in the human brain. *NMR Biomed.* (2013) 26:1630–46. doi: 10.1002/nbm.3045
5. Mansfield P. Spatial mapping of the chemical shift in NMR. *Magn Reson Med.* (1984) 1:370–86. doi: 10.1002/mrm.1910010308
6. Posse S, Tedeschi G, Risinger R, Ogg R, Bihan DL. High speed ^{1}H spectroscopic imaging in human brain by echo planar spatial-spectral encoding. *Magn Reson Med.* (1995) 33:34–40. doi: 10.1002/mrm.1910330106
7. Adalsteinsson E, Irarrazabal P, Topp S, Meyer C, Macovski A, Spielman DM. Volumetric spectroscopic imaging with spiral-based k-space trajectories. *Magn Reson Med.*(1998) 39:889–98.
8. Furuyama JK, Wilson NE, Thomas MA. Spectroscopic imaging using concentrically circular echo-planar trajectories *in vivo*. *Magn Reson Med.* (2012) 67:1515–22. doi: 10.1002/mrm.23184

9. Schirda CV, Tanase C, Boada FE. Rosette spectroscopic imaging: optimal parameters for alias-free, high sensitivity spectroscopic imaging. *J Magn Reson Imaging.* (2009) 29:1375–85. doi: 10.1002/jmri.21760
10. Pruessmann KP, Weiger M, Scheidegger MB, Boesiger P. Sense: sensitivity encoding for fast MRI. *Magn Reson Med.* (1999) 42:952–62. doi: 10.1002/(SICI)1522-2594(199911)42:5<952::AID-MRM16>3.0.CO;2-S
11. Dydak U, Weiger M, Pruessmann KP, Meier D, Boesiger P. Sensitivity-encoded spectroscopic imaging. *Magn Reson Med.* (2001) 46:713–22. doi: 10.1002/mrm.1250
12. Griswold MA, Jakob PM, Heidemann RM, Nittka M, Jellus V, Wang J, et al. Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magn Reson Med.* (2002) 47:1202–10. doi: 10.1002/mrm.10171
13. Otazo R, Tsai SY, Lin FH, Posse S. Accelerated short-te 3d proton echo-planar spectroscopic imaging using 2d-sense with a 32-channel array coil. *Magn Reson Med.* (2007) 58:1107–16. doi: 10.1002/mrm.21426
14. Strasser B, Považan M, Hangel G, Hingerl L, Chmelik M, Gruber S, et al. (2+ 1) d-caipirinha accelerated MR spectroscopic imaging of the brain at 7t. *Magn Reson Med.* (2017) 78:429–40. doi: 10.1002/mrm.26386
15. Wilson NE, Iqbal Z, Burns BL, Keller M, Thomas MA. Accelerated five-dimensional echo planar J-resolved spectroscopic imaging: implementation and pilot validation in human brain. *Magn Reson Med.* (2016) 75:42–51. doi: 10.1002/mrm.25605
16. Iqbal Z, Wilson NE, Thomas MA. 3d spatially encoded and accelerated te-averaged echo planar spectroscopic imaging in healthy human brain. *NMR Biomed.* (2016) 29:329–39. doi: 10.1002/nbm.3469

17. Posse S, Otazo R, Dager SR, Alger J. MR spectroscopic imaging: principles and recent advances. *J. Magn. Reson. Imaging.* (2013) 37:1301–25. doi: 10.1002/jmri.23945

18. Lam F, Liang ZP. A subspace approach to high-resolution spectroscopic imaging. *Magn Reson Med.* (2014) 71:1349–57. doi: 10.1002/mrm.25168

19. Hingerl L, Bogner W, Moser P, Považan M, Hangel G, Heckova E, et al. Density-weighted concentric circle trajectories for high resolution brain magnetic resonance spectroscopic imaging at 7t. *Magn Reson Med.* (2017) 79:2874–85. doi: 10.1002/mrm.26987

20. Hangel G, Strasser B, Považan M, Heckova E, Hingerl L, Boubela R, et al. Ultra-high resolution brain metabolite mapping at 7 t by short-TR hadamard-encoded FID-MRSI. *Neuroimage.* (2016) 168:199–210. doi: 10.1016/j.neuroimage.2016.10.043

21. Jain S, Sima DM, Sanaei Nezhad F, Hangel G, Bogner W, Williams S, et al. Patch-based super-resolution of MR spectroscopic images: application to multiple sclerosis. *Front Neurosci.* (2017) 11:13. doi: 10.3389/fnins.2017.00013

22. Haldar JP, Hernando D, Song SK, Liang ZP. Anatomically constrained reconstruction from noisy data. *Magn Reson Med.* (2008) 59:810–8. doi: 10.1002/mrm.21536

23. Hu X, Levin DN, Lauterbur PC, Spraggins T. Slim: spectral localization by imaging. *Magn Reson Med.* (1988) 8:314–22. doi: 10.1002/mrm.1910080308

24. Cengiz S, Valdes-Hernandez MdC, Ozturk-Isik E. Super resolution convolutional neural networks for increasing spatial resolution of ^{1} h magnetic resonance spectroscopic imaging. In: *Annual Conference on Medical Image Understanding and Analysis.* Cham: Springer (2017). p. 641–50.

25. Jacob M, Zhu X, Ebel A, Schuff N, Liang ZP. Improved model-based magnetic resonance spectroscopic imaging. *IEEE Trans Med Imaging.* (2007) 26:1305–18. doi: 10.1109/TMI.2007.898583

26. Liang ZP, Lauterbur PC. A generalized series approach to MR spectroscopic imaging. *IEEE Trans Med Imaging.* (1991) 10:132–7. doi: 10.1109/42.79470

27. Kasten J, Klauser A, Lazeyras F, Van De Ville D. Magnetic resonance spectroscopic imaging at superresolution: overview and perspectives. *J Magn Reson.* (2016) 263:193–208. doi: 10.1016/j.jmr.2015.11.003

28. LeCun Y, Boser B, Denker JS, Henderson D, Howard RE, Hubbard W, et al. Backpropagation applied to handwritten zip code recognition. *Neural Comput.* (1989) 1:541–51. doi: 10.1162/neco.1989.1.4.541

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

30. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In: *Advances in Neural Information Processing Systems.* (2012). p. 1097–105.

31. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention.* Cham: Springer (2015). p. 234–41.

32. Marcus DS, Wang TH, Parker J, Csernansky JG, Morris JC, Buckner RL. Open access series of imaging studies (oasis): cross-sectional mri data in young, middle aged, nondemented, and demented older adults. *J Cogn Neurosci.* (2007) 19:1498–507.doi: 10.1162/jocn.2007.19.9.1498

33. Pouwels PJ, Frahm J. Regional metabolite concentrations in human brain as determined by quantitative localized proton MRS. *Magn Reson Med.* (1998) 39:53–60. doi: 10.1002/mrm.1910390110

34. Huang G, Liu Z, Weinberger KQ, van der Maaten L. Densely connected convolutional networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Vol. 1.* (2017). p. 3.

35. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res.* (2014) 15:1929–58.

36. Kingma DP, Ba J. Adam: a method for stochastic optimization. *arXiv:1412.6980* (2014).

37. Chollet F. *Keras.* (2015). Available online at: https://github.com/fchollet/keras

38. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, et al. Tensorflow: large-scale machine learning on heterogeneous distributed systems. *arXiv:1603.04467* (2016).

39. Smith SL, Kindermans PJ, Ying C, Le QV. Don't decay the learning rate, increase the batch size. *arXiv:1711.00489* (2017).

40. Smith S, Levante T, Meier BH, Ernst RR. Computer simulations in magnetic resonance. An object-oriented programming approach. *J Magn Reson Ser A.* (1994) 106:75–105. doi: 10.1006/jmra.1994.1008

41. Bogner W, Gruber S, Trattnig S, Chmelik M. High-resolution mapping of human brain metabolites by free induction decay ^{1}H MRSI at 7 T. *NMR Biomed.* (2012) 25:873–82. doi: 10.1002/nbm.1805

42. Bilgic B, Chatnuntawech I, Fan AP, Setsompop K, Cauley SF, Wald LL, et al. Fast image reconstruction with l2-regularization. *J Magn Reson Imaging.* (2014) 40:181–91. doi: 10.1002/jmri.24365

43. Provencher SW. Estimation of metabolite concentrations from localized *in vivo* proton nmr spectra. *Magn Reson Med.* (1993) 30:672–9. doi: 10.1002/mrm.1910300604

44. Hatami N, Sdika M, Ratiney H. Magnetic resonance spectroscopy quantification using deep learning. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention.* Cham: Springer (2018). p. 467–75.

45. Wilson M, Reynolds G, Kauppinen RA, Arvanitis TN, Peet AC. A constrained least-squares approach to the automated quantitation of *in vivo* ^{1}H magnetic resonance spectroscopy data. *Magn Reson Med.* (2011) 65:1–12. doi: 10.1002/mrm.22579

46. Hu S, Lustig M, Chen AP, Crane J, Kerr A, Kelley DA, et al. Compressed sensing for resolution enhancement of hyperpolarized ^{13}C flyback 3D-MRSI. *J Magn Reson.* (2008) 192:258–64. doi: 10.1016/j.jmr.2008.03.003

47. Nelson SJ, Kurhanewicz J, Vigneron DB, Larson PE, Harzstark AL, Ferrone M, et al. Metabolic imaging of patients with prostate cancer using hyperpolarized [1-13c] pyruvate. *Sci Transl Med.* (2013) 5:198ra108. doi: 10.1126/scitranslmed.3006070

48. Hugg JW, Matson GB, Twieg DB, Maudsley AA, Sappey-Marinier D, Weiner MW. Phosphorus-31 mr spectroscopic imaging (MRSI) of normal and pathological human brains. *Magn Reson Imaging.* (1992) 10:227–43. doi: 10.1016/0730-725X(92)90483-G

49. Klomp DW, van de Bank BL, Raaijmakers A, Korteweg MA, Possanzini C, Boer VO, et al. 31p MRSI and ^{1}H MRS at 7 T: initial results in human breast cancer. *NMR Biomed.* (2011) 24:1337–42. doi: 10.1002/nbm.1696

50. Ma C, Clifford B, Liu Y, Gu Y, Lam F, Yu X, et al. High-resolution dynamic 31p-MRSI using a low-rank tensor model. *Magn Reson Med.* (2017) 78:419–28. doi: 10.1002/mrm.26762

51. Kennedy JA, Israel O, Frenkel A, Bar-Shalom R, Azhari H. Super-resolution in pet imaging. *IEEE Trans Med Imaging.* (2006) 25:137–47. doi: 10.1109/TMI.2005.861705

52. Iqbal Z, Nguyen D, Hangel G, Motyka S, Bogner W, Jiang S. Super-resolution ^{1}H magnetic resonance spectroscopic imaging utilizing deep learning. *arXiv:1802.07909* (2018).

Supervised Machine-Learning Enables Segmentation and Evaluation of Heterogeneous Post-treatment Changes in Multi-Parametric MRI of Soft-Tissue Sarcoma

Matthew D. Blackledge [1*†], Jessica M. Winfield [1,2*†], Aisha Miah [3,4], Dirk Strauss [5], Khin Thway [3,6], Veronica A. Morgan [1,2], David J. Collins [1,2], Dow-Mu Koh [1,2], Martin O. Leach [1,2] and Christina Messiou [1,2*]

[1] Cancer Research UK Cancer Imaging Centre, Division of Radiotherapy and Imaging, The Institute of Cancer Research, London, United Kingdom, [2] Department of Radiology, The Royal Marsden NHS Foundation Trust, Sutton, United Kingdom, [3] Sarcoma Unit, Department of Radiotherapy and Physics, The Royal Marsden NHS Foundation Trust, London, United Kingdom, [4] Division of Radiotherapy and Imaging, The Institute of Cancer Research, London, United Kingdom, [5] Department of Surgery, The Royal Marsden NHS Foundation Trust, London, United Kingdom, [6] Department of Histopathology, The Royal Marsden NHS Foundation Trust, London, United Kingdom

Background: Multi-parametric MRI provides non-invasive methods for response assessment of soft-tissue sarcoma (STS) from non-surgical treatments. However, evaluation of MRI parameters over the whole tumor volume may not reveal the full extent of post-treatment changes as STS tumors are often highly heterogeneous, including cellular tumor, fat, necrosis, and cystic tissue compartments. In this pilot study, we investigate the use of machine-learning approaches to automatically delineate tissue compartments in STS, and use this approach to monitor post-radiotherapy changes.

Methods: Eighteen patients with retroperitoneal sarcoma were imaged using multi-parametric MRI; 8/18 received a follow-up imaging study 2–4 weeks after pre-operative radiotherapy. Eight commonly-used supervised machine-learning techniques were optimized for classifying pixels into one of five tissue sub-types using an exhaustive cross-validation approach and expert-defined regions of interest as a gold standard. Final pixel classification was smoothed using a Markov Random Field (MRF) prior distribution on the final machine-learning models.

Findings: 5/8 machine-learning techniques demonstrated high median cross-validation accuracies (82.2%, range 80.5–82.5%) with no significant difference between these five methods. One technique was selected (Naïve-Bayes) due to its relatively short training and class-prediction times (median 0.73 and 0.69 ms, respectively on a 3.5 GHz personal machine). When combined with the MRF-prior, this approach was successfully applied in all eight post-radiotherapy imaging studies and provided visualization and quantification of changes to independent STS sub-regions following radiotherapy for heterogeneous response assessment.

Interpretation: Supervised machine-learning approaches to tissue classification in multi-parametric MRI of soft-tissue sarcomas provide quantitative evaluation of heterogeneous tissue changes following radiotherapy.

Keywords: magnetic resonance imaging, soft-tissue sarcoma, artificial intelligence, cancer heterogeneity, radiotherapy, imaging biomarkers

INTRODUCTION

Soft-tissue sarcoma (STS) is a rare form of cancer that develops in connective tissues. Approximately 3,300 new cases are diagnosed every year in the UK and the 5-years survival rate is ∼53% (1). STS tumors are often highly heterogeneous with variable tissue components that include cellular tumor, fat, necrosis, and cystic change. In patients undergoing non-surgical treatments, such as radiotherapy and systemic drug treatments, conventional imaging methods of assessing treatment response are limited as responding tumors may not change in size, or may even grow (pseudoprogression), after treatment (2–4). Hence, more effective and non-invasive methods for assessing treatment response are desired in trials of non-surgical treatments, such as combined radiotherapy with systemic agents. This is particularly difficult since the response of any tumor can be heterogeneous, with different components of a tumor responding differently to the same treatment.

Magnetic resonance imaging (MRI) is widely used in soft-tissue sarcoma, owing to its excellent soft-tissue contrast. Quantitative MRI techniques enable non-invasive investigation of the entire tumor and can provide information about the biological properties of tumors through functional measurements. For example, maps of apparent diffusion coefficient (ADC) derived from diffusion-weighted MRI inform on tissue cellularity, with lower ADC values observed in highly cellular or more aggressive regions within tumor (5). Using contrast enhanced MRI, the time course of T1 signal enhancement after intravenous injection of gadolinium-based contrast agent provides estimates of tumor perfusion and permeability (6). By applying the Dixon MRI techniques, the presence of fat in sarcomas can also be measured and quantified (7).

However, evaluation of multi-parametric quantitative MRI averaged over the entire tumor may not reveal the extent of heterogeneous changes following treatment. By combining quantitative MRI techniques that inform on different aspects of tumor properties (e.g., diffusion-weighted MRI, contrast enhanced MRI and Dixon MRI), it is possible identify sub-components of tumors demonstrating cellular, vascular or fatty phenotypes before and after treatment, thereby enabling tracking and monitoring of the heterogeneity of tumors in response to treatment.

The aim of this pilot study is to evaluate supervised machine learning methods for tissue classification of multi-parametric MRI measurements in soft-tissue sarcomas, and use these methods to quantify post-treatment changes in a cohort of patients treated with radiotherapy.

MATERIALS AND METHODS

Patient Cohort

Eighteen patients with retroperitoneal sarcomas were included in this prospective single-center study (11 male patients and seven female patients; age range 43–76). The study was approved by a national Research Ethics Committee, and all patients gave their written informed consent to participate. Tumors included 14 liposarcomas, two leiomyosarcomas, one spindle cell sarcoma, and one synovial sarcoma. All patients underwent an MRI examination at baseline. In eight patients who were treated with pre-operative radiotherapy (50.4 Gy in 28 fractions) another MRI examination was performed 2–4 weeks after the final fraction of radiotherapy and prior to surgery; 10 patients were treated by surgery alone.

Imaging Protocol

Patients were scanned on a 1.5 T Siemens MAGNETOM Aera MRI scanner (Siemens Healthcare AG, Erlangen, Germany). Anterior body matrix and posterior spine matrix receive coils were used for image acquisition. Following axial and coronal anatomical T_1-weighted and T_2-weighted imaging sequences, functional imaging was performed and consisted of diffusion-weighted imaging (DWI), Dixon imaging, and pre- and post-contrast T_1-weighted imaging sequences. Images were acquired with a field of view that fully covered the tumor volume; parameters are described in **Appendix A** and further detailed by Winfield et al. (8) (a second imaging station was used if necessary for large tumors). Post-Gadolinium (Gd) T_1-weighted images were acquired 4 min after injection of a Gd-based contrast agent (Dotarem, 0.2 ml/kg body weight, administered at 2 ml/s using a power injector).

Image Analysis

Maps of apparent diffusion coefficient (ADC) were calculated from the DWI and fat-fraction (FF) from Dixon images:

$$FF = \frac{S_{fat}}{S_{fat} + S_{water}} \times 100\% \qquad (1)$$

where S_{fat} and S_{water} represent the fat and water signals, respectively. Maps of fractional enhancement (EF) were calculated from the pre- and post-Gd T_1-weighted images using the following equation:

$$EF = \frac{S_{post} - S_{pre}}{S_{post} + S_{pre}} \times 100\% \qquad (2)$$

where S_{pre} and S_{post} are the signal intensities in pre- and post-Gd T_1-weighted images, respectively (9). Volumes of interest (VOIs)

FIGURE 1 | (Left) An illustration of our decision tree used to define habitats within our sarcoma population. Classes 3 and 4 were not further divided as cystic/necrotic regions and fat do not enhance in post-Gd images. ADC is not evaluable in fat- suppressed DWI. **(Right)** Images from one patient with a dedifferentiated liposarcoma showing examples of training ROIs positioned in regions corresponding to habitat 1 (red) and habitat 3 (blue). Training ROIs (2 cm^2) were drawn on either the fat fraction (FF), apparent diffusion-coefficient (ADC), or enhancing fraction (EF) maps, and then transposed onto other maps.

were defined for each tumor by an expert radiologist with 16 years of experience, outlining the whole tumor on every slice on which the tumor appeared on axial T$_2$-weighted images; VOIs were transferred to ADC, FF, and EF maps. All parameter maps were rescaled to ensure values were in the range [0, 1] using the following linear transformations: ADC $\rightarrow$ ADC/3 $\times$ 10^{-3} mm^2/s, EF $\rightarrow$ (EF + 100)/200%, and FF $\rightarrow$ FF/100%. No spatial registration was performed between parameter maps as adequate spatial alignment was verified by a consultant radiologist with experience in STS MRI.

Tissue Classification

We defined four possible tissue classes for the STS volumes as illustrated in **Figure 1**, reflecting the aim of segmenting cellular tumor (low ADC, classes 1 and 2) from necrotic/cystic regions (high ADC, class 3), fat (class 4). The cellular tumor was further separated into enhancing (class 1) and non-enhancing (class 2), which may have different biological behavior (2). In addition, we defined a further class to represent the combinations of MRI parameters that were not part of the training data, called "novelties" (10) (class 5). Training data for building the machine-learning classifiers were defined by placing square regions of interest (ROIs) with area 1–2 cm^2 (45–100 voxels) in regions that exemplified each class, at locations far from visible boundaries

(**Figure 1**). Training ROIs were drawn by a clinical scientist with more than 6 years of experience in tumor analysis and confirmed by a consultant radiologist with 16 years of experience. Between 1 and 4 ROIs were placed in each tumor depending on the classes present, providing a total of 36 ROIs across all 18 patients' baseline scans.

Eights machine-learning (ML) techniques were evaluated for classifying the tissue type for each voxel in this pilot supervised classification exercise using the Scikit-Learn software package (11): Logistic Regression (LR), Support Vector Machine (SVM with a radial basis function), Random Forest (RF), k-Nearest Neighbor (kNN), Kernel Density Estimation (KDE), Naïve-Bayes (NB), and a 20-node, three-layer, fully-connected Neural Network (NN). We also tested a variant of the KDE method where the hyperparameter (bandwidth) was automatically selected using Silverman's approximation (12). To ensure that techniques were sensitive to novelties (voxels that do not represent any of the classes defined in this study), data for an additional 15 ROIs were synthesized by randomly sampling from a uniform distribution covering the intrinsic range of the parameters: EF $\in$ [−100, 100] (%), FF $\in$ [0, 100] (%), ADC $\in$ [0, 3] ($\times$10^{-3} s/mm^2). All data were normalized to the range [0, 1] prior to training of algorithms. An exhaustive cross-validation approach was used for evaluating classification performance of

TABLE 1 | Median training and prediction times for each of the machine-learning techniques used in this study over the range of hyper-parameters tested (5th and 9th percentiles provided in parentheses).

ML technique	Median training time in ms (5th−95th perc.)	Median prediction time in ms (5th−95th perc.)	Hyper-parameter (range considered for optimization)
Logistic regression (LR)	9.51 (5.50, 12.58)	0.06 (0.05, 0.11)	**C:** Inverse of the regularization strength (10^{-3}–10^{10})
Support vector machine (SVM)	208.69 (96.01, 1,745.5)	12.94 (5.19, 29.81)	**C:** Penalty parameter that favors smoother decision boundaries when set to a smaller value (10^{-3}–10^{5})
Neural network (NN)	412.23 (47.72, 465.35)	0.24 (0.22, 0.33)	**α:** A L2-regularization parameter that attempts to reduce over-fitting. Smoother decision boundaries with larger values (10^{-8}–10^{5})
Naïve-Bayes (NB)	0.73 (0.70, 1.26)	0.69 (0.66, 1.23)	**None**
Random forest (RF)	399.98 (7.35, 7,304.26)	4.74 (0.23, 86.40)	**N Estimators**: The number of trees being used in the forest (**10–1,000**)
k-Nearest neighbor (kNN)	1.73 (1.64, 2.39)	6.00 (1.20, 67.10)	**N:** The number of closest training data (Euclidean distance) considered to be neighbors of the data being predicted (**10–1,000**)
Kernel density estimation (KDE)	1.57 (1.44, 2.55)	11.33 (6.15, 55.62)	**Bandwidth:** The standard-deviation of the Gaussian kernel used for fitting a KDE model (10^{-4}–10^{1})
Automatic KDE (aKDE)	1.81 (1.54, 2.82)	30.84 (26.83, 36.94)	**None**: Bandwidth is automatically calculated using Silverman's approximation

Training times are estimated for 1,350–3,000 samples in each case, whilst prediction times are for 135–300 samples (validation step). A brief description of the hyper-parameter used in each case is provided (if applicable), with range test provided in parentheses. Computation times are from a 3.5 GHz personal machine with 16 GB of memory and an Intel Iris Plus graphics card.

each of the machine-learning techniques: For each training cycle voxels from one ROI of each class were selected as a validation set, and the ML method under investigation was trained on voxel values from the remaining ROIs. This process was repeated for each unique combination of validation ROIs providing a total of 2,240 training/validation cycles. This process was repeated over the range of hyper-parameters considered for each ML method (see **Table 1** for a list of the hyper-parameters considered and their range, along with training/prediction times for each model), and the hyper-parameter that provided the highest median accuracy, defined as the percentage of voxels correctly classified in the validation ROI set, was chosen for further investigation. The data for one patient, for whom 3 different ROI classes had been drawn, was left out of this training/validation phase in order to evaluate the accuracy of these ML methods in an unseen case; this left a total of 33 ROIs for cross-validation analysis. Comparison between methods was achieved using a two-tailed Student's *t*-test ($p < 0.05$ for significance).

Once the optimum hyper-parameter was selected and models had been trained, they were used to classify the entire tumor volume in all patients, providing a map of the suspected STS tissue sub-type at each voxel location (13) for radiological review. Results were visualized using (i) 3D surface rendering and (ii) color-coded masks overlain on Multi-Planar Reformats (MPRs) of the anatomical images acquired (T2-HASTE). To reduce the level of classification noise observed in the derived habitat maps, a classification de-noising algorithm was used by applying a Markov Random Field (MRF) model to the machine-learned classifications (see **Appendix B** for the theoretical justification underlying this model, with Python code provided as supplementary file "ml_utilities.py").

RESULTS

The cross-validation accuracy for the ranges of hyper-parameters tested in each of the machine-learning methods is demonstrated in **Figure 2**. For both kNN and KDE methods, optimum hyper-parameters can be established (number of neighbors = 34 and bandwidth = 0.75, respectively). For the remaining ML methods, a plateau is reached in the cross-validation accuracy indicating relative insensitivity to the choice of hyper-parameter after some threshold. **Figure 2** also demonstrates the accuracy of each machine learning method on the test ROIs ignored during training: RF classification scored the highest in this case with a test accuracy of 98.1%, and SVM, NN and kNN methods demonstrating slightly lower accuracies of 96.3, 93.2, and 89.4% respectively.

Figure 3 demonstrates the cross-validation accuracy for each of the classes independently and for all classes combined, using the optimal hyper-parameters in each case. The results are sorted in order of ascending median accuracy. NB scored the highest in two out of five tissue classes: (3) high ADC, and (4) fatty tissue, whilst kNN scored highest for discriminating enhancing, well-vascularised (1) from non-enhancing, poorly vascularised (2) tumor tissue. The performance for all tissue types combined demonstrates that in general there is little to choose between 5/8 of these classification methods (NB, NN, KDE, NN, SVM), whilst logistic regression (LR) and random forest classifiers perform poorly in comparison

FIGURE 2 | Demonstration of cross-validation accuracies over the range of hyper-parameters tested in this study. For the Kernel Density Estimation and k-Nearest Neighbor methods, an optimum hyper-parameter can be identified. For the remaining techniques, a hyper-parameter limit is identified by the presence of a plateau in the validation accuracy curve. Solid curves represent median values, shaded areas demonstrate the interquartile range and dashed lines represent the 5th and 95th percentiles of the validation accuracy measurements. The optimum hyper parameter is annotated on each sub-plot with the corresponding validation accuracy shown in the top-left.

across the tissue sub-types considered. Of the five methods, the Naïve-Bayes (NB) classifier was chosen for further investigation due to its relatively short training and prediction times (**Table 1**).

Figure 4 compares the classification results of the NB classifier with and without MRF correction on the test-patient that was

not included in the initial training of our machine-learning approaches. It is evident that the application of a MRF reduces the classification noise induced when the classifier is applied on a voxel-wise basis without taking into consideration the correlations that are likely to occur between neighboring voxels. This figure also demonstrates the convergent properties of the

FIGURE 3 | Comparison of the validation accuracy for the different machine learning (ML) techniques applied to our labeled training-set data. Methods are compared for each tissue label separately (colored as per **Figure 1**) and for all tissue types combined (white). Boxplots demonstrate the distribution of validation accuracies (derived using a randomized cross-validation approach) following optimization of hyper-parameters (bold-lines represent median, shaded areas indicate the inter-quartile range and whiskers the 5th/95th percentiles). Methods are ordered from left to right in order of increasing median accuracy (**$p < 0.005$, ***$p < 0.0005$).

MRF algorithm, which converged after a median of 27 iterations in this case.

We used the NB classifier, in combination with our MRF class-label de-noising algorithm, to investigate the changes occurring to each of the tissue habitats in three patients who received a post-treatment MR exam following radiotherapy (**Figure 5**). Patient 1 demonstrated STS consisting of mostly viable tumor with high vascularity (class 1 in red), with a necrotic core (class 3 in blue). Following treatment, there was no clear change in the volume of either of these tissue types, nor any change in the ADC (as depicted through a pie-chart in the figure), indicating that the patient did not respond well to treatment. Patient 2 demonstrated with a highly heterogeneous STS, with a mix of tissue classes (1), (2), and (3). Following treatment, there is a clear increase in the proportion of non-enhancing tissue, suggestive of disruption to the vascular supply of the tumor following radiotherapy. When combined with an observed increase in ADC for the remaining well-vascularized tissue, this may provide evidence of tumor response to radiotherapy, regardless of the absence of any significant change in tumor volume (5.7% reduction

following treatment). Patient 3, however, demonstrated highly fatty, well-differentiated liposarcoma, which has been well-described through our approach; no change is found following radiotherapy. Results for all eight patients are provided in **Appendix C**.

DISCUSSION

Soft-tissue sarcoma is a highly heterogeneous disease, and there remains a lack of appropriate imaging biomarkers for monitoring the success of therapy. Novel therapeutic agents or radiotherapy may not result in a significant change in tumor size, but in a heterogeneous change in the tumor composition. In this technical development study, we have investigated the use of a number of machine-learning approaches for automatically segmenting the heterogeneous tissue compartments within STS, thereby providing a map that aims to characterize the tumor microenvironment for radiological review. This approach facilitates the quantification of changes in ADC, fat-fraction and

FIGURE 4 | Demonstration of the improvement to tissue sub-region classification following Markov Random Field (MRF) correction of the Naïve-Bayes classifier. This figure demonstrates results for the patient that was not included in the training of our machine-learning approaches (test data). Spie-charts (14) demonstrate the proportion of each tissue sub-compartment within the entire volume as the angle of each segment, whilst the mean ADC of each tissue sub-type is represented by the radius of each segment (note that the ADC of the fat/yellow tissue sub-type from fat-suppressed diffusion-weighted imaging studies should not be interpreted as it will be heavily noise-corrupted; only the proportion/angle of this tissue sub-type is informative). The far-right plot demonstrates the number of voxels that change classification following each iteration through the MRF fitting algorithm across all axial images in this patient: it is evident that the algorithm converges after a finite number of iterations.

enhancement-fraction estimated through co-registered, multi-parametric MRI occurring in each of the segmented tissue classes, and may provide a novel response biomarker in STS.

Out of the eight machine-learning approaches we investigated, we found that 5/8 methods did not outperform each other in terms of segmentation accuracy. This is likely due to the fact that our data is intrinsically low-dimensional (only three parameters per-voxel: ADC, enhancement-fraction and fat-fraction), and most of the techniques provide enough degrees of freedom to account for the variation of these parameters for the different classes investigated for STS. This is supported by the relatively poor performance of logistic regression, which was unable to model the full complexity of the data space.

In addition, we have investigated inclusion of the estimated class probabilities from machine-learning classification methods into a Markov Random Field framework, which allows for de-noising of the estimated habitat maps by introducing a spatial prior distribution on the segmented regions. This technique provided smoother classification maps when compared to classification based purely on the trained ML architectures alone. This approach could well be extended to any other machine-learning task where the classifications of a group of input data are not expected to be independent (15–17).

Although previous authors have investigated the role of machine learning for the segmentation of sarcoma using MRI data, these reports focused on the utility of dynamic contrast-enhanced MRI alone, and did not exploit the multi-parametric capabilities of MRI for determining a more complete habitat image of the tumor, as explored here (18, 19). Moving forward,

there is a clear need to explore a larger patient population for further validation of the methods described in this article. This should include multi-center studies to determine the sensitivity of the technique to images acquired from multiple vendors and at different institutions (20). Another important consideration is when MRI studies should be performed following neoadjuvant radiotherapy in order to observe a measureable treatment-induced change; the effects of treatment may not manifest immediately after the final radiation dose. However, the timing of imaging after neoadjuvant radiotherapy is limited by surgery, which is typically performed at 4–6 weeks post-treatment. Imaging following radiotherapy to non-resectable disease may enable insight into later effects. The segmentation methodology would also benefit from repeatability testing to determine its sensitivity as a radiotherapy response biomarker (21). A limitation of this study is that one expert radiologist generated training data samples in the patients investigated, and so further work may investigate the user-repeatability for generating gold-standard training data. The regions chosen for training data would ideally be validated through post-operative histopathological confirmation of the tissue type in that region. There may also be scope for including more complex deep-learning approaches for producing habitat maps for soft-tissue sarcoma, including methods, such as U-Net convolutional networks (22), but these techniques would require a much larger cohort size, which may be unfeasible in a population with a rare cancer type. Lastly, the cohort of eight patients who received radiotherapy had STS tumors that were predominantly well-vascularized, and future randomized studies should include

FIGURE 5 | A demonstration of our proposed habitat classification scheme on three patients applied before and after radiation therapy. Spie charts are presented with a radius equal to the mean ADC of the given tissue sub-compartment; dotted lines on the post-treatment Spie charts show the ADC of that tissue sub-type in the pre-treatment data. Multi-planar reformat habitat maps are overlain on T2 HASTE MR-images acquired within the same patient study. **Patient 1** demonstrates a patient with a liposarcoma where a necrotic core is clearly identified (blue) within a majority of strongly enhancing solid tumor (red) prior to treatment. Although there is a marginal increase in the volume of the necrotic core, there little overall change is observed following treatment. **Patient 2** demonstrates data from a pleomorphic
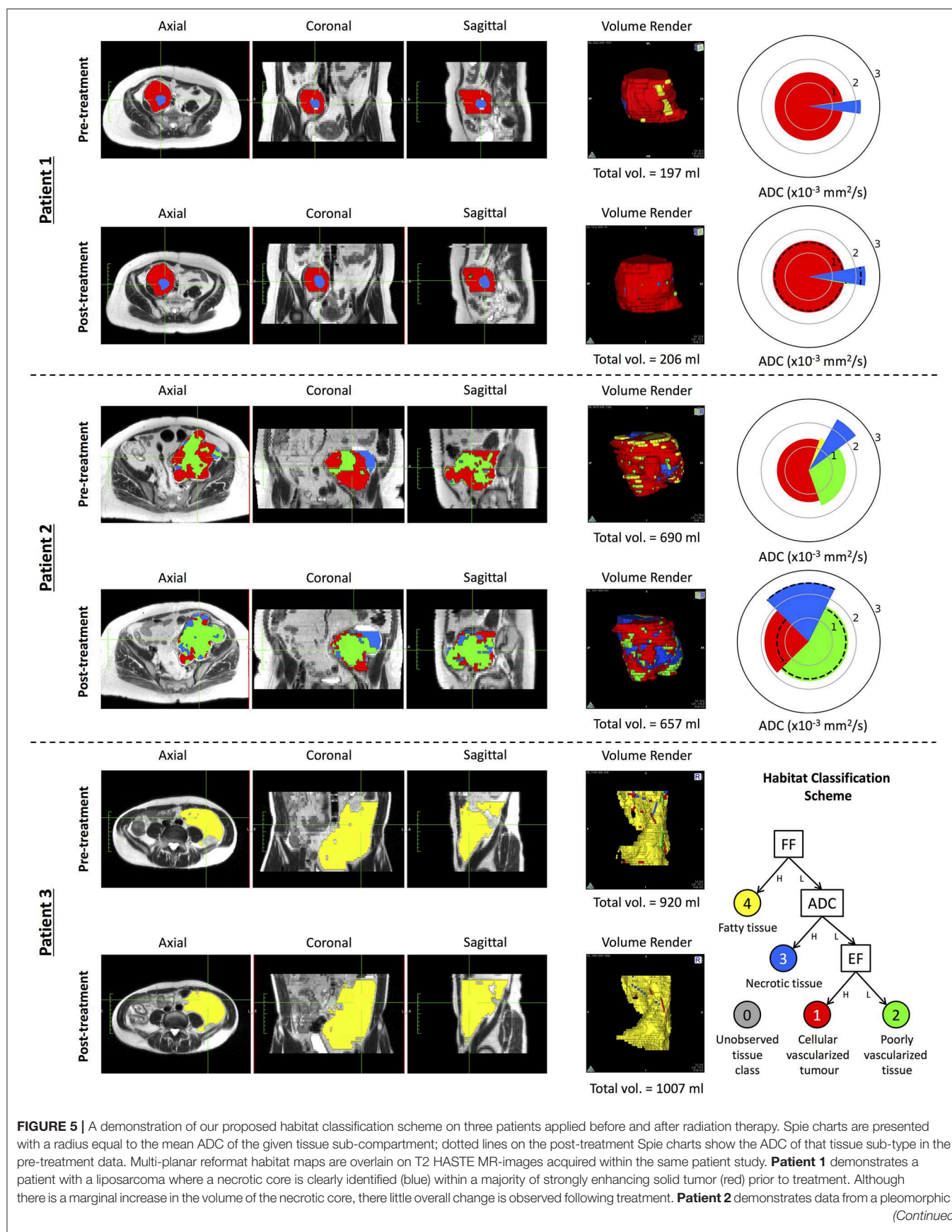
(Continued)

FIGURE 5 | sarcoma where there is a clear heterogeneous pattern observed with the majority of the disease consisting of strongly enhancing tumor. Following treatment, there is a marked increase in the proportion of poorly vascularized (green) and necrotic tissue. Within the remaining strongly enhancing tumor after radiotherapy, an increase in mean ADC is observed indicative of treatment response. **Patient 3** demonstrates a well-differentiated liposarcoma with the majority of the tumor consisting of fatty tissue before and after treatment. Results for all eight patients (including these three exemplary patients) with pre-/post-radiotherapy imaging are provided as supplementary information in **Appendix C**.

patients with more heterogeneous tumor phenotypes. However, the full cohort of this study, which included patients for whom no radiotherapy was delivered, provided sufficient examples of each tissue class to evaluate this technological development.

Modern advances in artificial intelligence and machine-learning are anticipated to improve automatic segmentation accuracies in the next few years, and supersede conventional image-processing methods for extracting regions of interest in medical imaging datasets. We have demonstrated that a variety of simple machine-learning approaches can be used to automatically extract sub-regions in a highly heterogeneous tumor phenotype, and that quantification of the volume and ADC within these regions may provide a radiotherapy response biomarker in soft-tissue sarcoma. Tools, such as these will facilitate clinical decision making for a disease that can be difficult to manage, and thus may promote personalized treatment regimens and improve patient outcome. Intra-tumoural heterogeneity confounds the interpretation of treatment response in many other, more common cancers; provided sufficient data is acquired, we envisage that these methods will be highly applicable in many prospective cancer studies investigating tumor response to targeted therapeutics.

DATA AVAILABILITY STATEMENT

Datasets are available upon request from the authors. Code developed is provided in **Appendix D**.

ETHICS STATEMENT

This study was carried out in accordance with the Declaration of Helsinki (1996) and conditions of the ethical approval obtained from the National Research Ethics Service (NRES) committee Cambridge East REC: 13/EE/1086.

AUTHOR CONTRIBUTIONS

MB, JW, AM, DS, KT, VM, DC, D-MK, ML, and CM: literature search. MB, JW, and CM: figures. MB, JW, AM, DS, KT, DC, and CM: study design. JW, AM, DS, KT, VM, DC, and CM: data acquisition. MB, JW, DC, and CM: data analysis. MB, JW, DC, D-MK, and CM: data interpretation. MB and JW: software development. MB, JW, DC, D-MK, ML, and CM: article writing.

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2019.00941/full#supplementary-material

REFERENCES

1. Cancer Research UK. *Soft Tissue Sarcoma Statistics*. Available online at: http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/soft-tissue-sarcoma

2. Messiou C, Bonvalot S, Gronchi A, Vanel D, Meyer M, Robinson P, et al. Evaluation of response after pre-operative radiotherapy in soft tissue sarcomas; The European Organisation for Research and Treatment of Cancer–Soft Tissue and Bone Sarcoma Group (EORTC–STBSG) and Imaging Group recommendations for radiological examina. *Eur J Cancer*. (2016) 56:37–44. doi: 10.1016/j.ejca.2015.12.008

3. Roberge D, Skamene T, Nahal A, Turcotte RE, Powell T, Freeman C. Radiological and pathological response following pre-operative radiotherapy for soft-tissue sarcoma. *Radiother Oncol*. (2010) 97:404–7. doi: 10.1016/j.radonc.2010.10.007

4. Canter RJ, Martinez SR, Tamurian RM, Wilton M, Li CS, Ryu J, et al. Radiographic and histologic response to neoadjuvant radiotherapy in patients with soft tissue sarcoma. *Ann Surg Oncol*. (2010) 17:2578–84. doi: 10.1245/s10434-010-1156-3

5. Koh DM, Collins DJ. Diffusion-weighted MRI in the body: applications and challenges in oncology. *Am J Roentgenol*. (2007) 188:1622–35. doi: 10.2214/AJR.06.1403

6. Messiou C, Orton M, Ang JE, Collins DJ, Morgan VA, Mears D, et al. Advanced solid tumors treated with cediranib: comparison of dynamic contrast-enhanced MR imaging and CT as markers of vascular activity. *Radiology*. (2012) 265:426–36. doi: 10.1148/radiol.12112565

7. Ma J. Dixon techniques for water and fat imaging. *J Magn Reson Imaging*. (2008) 28:543–58. doi: 10.1002/jmri.21492

8. Winfield JM, Miah AB, Strauss D, Thway K, Collins DJ, deSouza NM, et al. Utility of multi-parametric quantitative magnetic resonance imaging for characterization and radiotherapy response assessment in soft-tissue sarcomas and correlation with histopathology. *Front Oncol*. (2019) 9:280. doi: 10.3389/fonc.2019.00280

9. Blackledge MD, Rata M, Tunariu N, Koh D-M, George A, Zivi A, et al. Visualizing whole-body treatment response heterogeneity using multi-parametric magnetic resonance imaging. *J Algorithms Comput Technol*. (2016) 10:290–301. doi: 10.1177/1748301816668024.

10. Pimentel MAF, Clifton DA, Clifton L, Tarassenko L. A review of novelty detection. *Signal Process*. (2014) 99:215–49. doi: 10.1016/j.sigpro.2013.12.026

11. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in python. *J Mach Learn Res*. (2012) 12:2825–30. doi: 10.1007/s13398-014-0173-7.2

12. Silverman B. Density estimation for statistics and data analysis. *Chapman Hall*. (1986) 37:1–22. doi: 10.2307/2347507

13. Sala E, Mema E, Himoto Y, Veeraraghavan H, Brenton JD, Snyder A, et al. Unravelling tumour heterogeneity using next-generation imaging: radiomics, radiogenomics, and habitat imaging. *Clin Radiol*. (2017) 72:3–10. doi: 10.1016/j.crad.2016.09.013

14. Feitelson DG. Comparing partitions with spie charts. *Sch Comput Sci*. (2003):1–7. Available online at: http://www.cs.huji.ac.il/%7B~%7Dfeit/papers/Spie03TR.pdf

15. Do T-M-T, Artieres T. Neural conditional random fields. *Aistats*. (2010) 9:1–9. Available online at: http://books.nips.cc/papers/files/nips22/NIPS2009_0935.pdf

16. Zheng S, Jayasumana S, Romera-Paredes B, Vineet V, Su Z, Du D, et al. Conditional random fields as recurrent neural networks. *arXiv*. (2015) cs.CV:2015. doi: 10.1109/ICCV.2015.179

17. Chen L-C, Papandreou G, Kokkinos I, Murphy K, Yuille AL. Semantic image segmentation with deep convolutional nets and fully connected CRFs. *Iclr*. (2014):1–14. Available online at: http://arxiv.org/abs/1412.7062

18. Monsky WL, Jin B, Molloy C, Canter RJ, Li CS, Lin TC, et al. Semi-automated volumetric quantification of tumor necrosis in soft tissue sarcoma using contrast-enhanced MRI. *Anticancer Res*. (2012) 32:4951–62. doi: 10.1126/scisignal.2001449

19. Glass JO, Reddick WE. Hybrid artificial neural network segmentation and classification of dynamic contrast-enhanced MR imaging (DEMRI) of osteosarcoma. *Magn Reson Imaging*. (1998) 16:1075–83.

20. 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. *Radiology*. (2017) 284:88–99. doi: 10.1148/radiol.2017161965

21. Blackledge MD, Tunariu N, Orton MR, Padhani AR, Collins DJ, Leach MO, et al. Inter- and intra-observer repeatability of quantitative whole-body, diffusion-weighted imaging (WBDWI) in metastatic bone disease. *PLoS ONE*. (2016) 11:e0153840. doi: 10.1371/journal.pone.0153840

22. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. *Miccai*. (2015):234–41. doi: 10.1007/978-3-319-24574-4_28

Respiratory-Correlated (RC) vs. Time-Resolved (TR) Four-Dimensional Magnetic Resonance Imaging (4DMRI) for Radiotherapy of Thoracic and Abdominal Cancer

Guang Li*, Yilin Liu and Xingyu Nie

Department of Medical Physics, Memorial Sloan Kettering Cancer Center, New York, NY, United States

Recent technological and clinical advancements of both respiratory-correlated (RC) and time-resolved (TR) four-dimensional magnetic resonance imaging (4DMRI) techniques are reviewed in light of tumor/organ motion simulation, monitoring, and assessment in radiotherapy. For radiotherapy of thoracic and abdominal cancer, respiratory-induced tumor motion, and motion variation due to breathing irregularities are the major uncertainties in treatment. RC-4DMRI is developed to assess tumor motion for treatment planning, whereas TR-4DMRI is developed to assess both motion and motion variation for treatment planning, delivery and assessment. RC-4DMRI is reconstructed to provide one-breathing-cycle motion, similar to 4D computed tomography (4DCT), the current clinical standard, but with higher soft-tissue contrast, no ionizing radiation, and less binning artifacts due to the use of an internal respiratory surrogate. Recent studies have shown that its spatial resolution has reached or exceeded that of 4DCT and scanning time becomes clinically acceptable. TR-4DMRI is recently developed with an adequate spatiotemporal resolution to assess tumor motion and motion variations for treatment simulation, delivery and assessment. The super-resolution approach is most promising since it can image any organ/body motion, whereas RC-4D MRI are limited to resolve only respiration-induced motion and some TR-4DMRI approaches may more or less depend on RC-4DMRI. TR-4DMRI provides multi-breath motion data that are useful not only in MR-guided radiotherapy but also for building a patient-specific motion model to guide radiotherapy treatment using an non-MR-equipped linear accelerator. Based on 4DMRI motion data, motion-corrected dynamic contrast imaging and diffusion-weighted imaging have also been reported, aiming to facilitate tumor delineation for more accurate radiotherapy targeting. Both RC- and TR-4DMRI have been evaluated for potential clinical applications, such as delineation of tumor volumes, where sufficiently high spatial resolution and large field-of-view are required. The 4DMRI techniques are promising to play a role in motion assessment in radiotherapy treatment planning, delivery, assessment, and adaptation.

Keywords: 4DMRI, radiation therapy (radiotherapy), tumor motion assessment, treatment planning and delivery, respiratory motion and motion variation

INTRODUCTION

Respiratory motion management is a critical component for radiotherapy of malignant tumors in the thorax and abdomen, including lung, liver, pancreatic, and adrenal cancer. Clinical strategies for motion management include breath-hold, abdominal compression, as well as 4D imaging techniques for targeting internal tumor volume (ITV), respiratory gating, and tumor tracking (1–3). For motion assessment of a mobile tumor, 4D imaging is necessary to provide a patient-specific motion margin for treatment. In image-based treatment planning, respiratory-correlated (RC) four-dimensional computed tomography (4DCT) is the current clinical standard to assess respiratory-induced tumor motion. In image-guided radiotherapy (IGRT) using a conventional medical linear accelerator (Linac), 4D cone-beam CT (4D CBCT) may be applied for patient setup, while periodic MV/kV imaging, intrafractional motion review (IMR), and video-based optical surface imaging may be applied for intrafractional motion monitoring (3). Recently, magnetic resonance imaging (MRI) has been increasingly applied for MR-based planning and MR-guided radiotherapy (MRgRT), and various 4DMRI techniques have been developed, including respiratory-correlated (RC) and time-resolved (TR) 4DMRI.

Similar to 4DCT, RC-4DMRI is reconstructed by binning partial volumetric images or k-space data, such as 2D MR slice images, based on the signal from a respiratory surrogate assuming periodic respiration. However, RC-4DMRI provides higher soft-tissue contrast to visualize gross tumor volume (GTV) and nearby organs at risk (OARs) without ionizing radiation, higher image quality with less binning artifacts by using an internal respiratory surrogate to eliminate the uncertainties from an imperfect external-internal correlation, and higher precision in assessing the primary superior-inferior motion within image slices by imaging in the sagittal or coronal directions. Unlike 4DCT and RC-4DMRI, dynamic or TR-4DMRI does not assume periodic motion because it captures the images of a moving object on the fly. Therefore, TR-4DMRI is ideal to assess respiratory motion, which is often irregular. Furthermore, organ motion can be driven jointly by other involuntary motions, such as cardiac and digestive motions, or voluntary body motion. These motions are either random in nature or having a different rhythm; neither correlates well with respiration. So, the GTV/OAR motion may be complex and non-periodical. Therefore, TR-4DMRI is useful to assess respiratory motion in multi-breathing cycles with irregularities but no binning artifacts, providing higher 4DMRI image quality for GTV/ITV delineation and more realistic GTV/OAR motion for treatment planning and dose delivery assessment.

Historically, dynamic 3D cine MRI was first studied for imaging respiratory motion through direct acquisition (4). This would be the most desirable 4D imaging form to assess organ motion regardless it is regular or irregular, simple or complex, and voluntary or involuntary because it does not assume periodic motion and does not need any respiratory surrogate for reconstruction. However, only low spatial resolution 3D cine was achieved because it was limited by the slow physical MR

relaxation and large clinical field of view, even though parallel imaging and view-sharing approximation were applied. Despite the recent development of state-of-the-art MRI techniques (5–7), the basic limit of MR acquisition speed is still present. Facing the fundamental challenge, alternative approaches were developed to circumvent this limitation, including a super-resolution (SR) approach to achieving higher spatiotemporal resolution by combining two sets of MRI image series with complementary strength in either high temporal or high spatial resolution (8, 9). The dynamic 3D cine images acquired with low spatial resolution in free-breathing (FB) of a patient serve as the motion template to map the high spatial resolution from a breath-hold (BH) 3D MRI image of the same patient through deformable image registration (DIR).

In this short review article, we will start with a discussion on RC-4DMRI developments in section Respiratory-Correlated (RC) 4DMRI, with a different emphasis from two recent review articles on the RC-4DMRI (10, 11). The focus of this review will be on different approaches to reconstruct TR-4DMRI, with emphasis on the SR approach and its potentials, and on the discussion and comparison with RC-4DMRI in section Time-Resolved (TR) 4DMRI. The utility of the TR-4DMRI in clinical research and potentials in clinical applications will be discussed in section Clinical Evaluations and Applications of RC- and TR-4DMRI. Finally, we will summarize the recent advancements and provide an outlook for future development in the 4DMRI field.

RESPIRATORY-CORRELATED (RC) 4DMRI

Respiratory-correlated 4DMRI consists of several 3D images covering different respiratory states of one breathing cycle of a patient, similar to 4DCT. The similarity and difference among 4DCT, 2D/3D cine MRI, as well as RC- and TR-4DMRI are summarized in **Table 1**. It is worthwhile to mention that the use of an internal MR navigator as the surrogate reduces binning artifacts because it eliminates the uncertainty from the external-internal correlation if an external surrogate is used (12). In addition, the versatility of MR acquisition and reconstruction allows different ways for sorting the 2D MR slice images: (1) scanning k-space data via Cartesian, Radial, or Spiral acquisition, (2) orienting acquisition in axial, sagittal or coronal directions, (3) binning in image space or k-space, and (4) reconstructing the images prospectively or retrospectively.

In a prospective approach, the acquisition intends to fill the table with the row of respiratory state and the column of image slice. When all the table elements are acquired, the acquisition and reconstruction of RC-4DMRI are completed. Hu et al. developed a prospective T2-weighted 4D-MRI method with a respiratory amplitude-based triggering system to gate 2D MRI image acquisition (13). Either an internal or external surrogate can be applied, but the former produced superior image quality than the latter (12). In a retrospective approach, similar to 4DCT scan, more scans than a cycle may be acquired to reconstruct complete 3D images in all respiratory states. Cai et al. proposed a T2/T1-weighted 4DMRI retrospective phase sorting method that used body area extracting from images as the respiratory

TABLE 1 | Comparison of RC-4DMRI, 2D/3D cine MR, TR-4DMRI with 4DCT and 2D kV/MV.

Category	Sub-category	4DCT	RC-4DMRI	2DkV/MV	2D cine	3D cine	TR-4DMRI
Acquisition	Scanning	Projection	2D slice	Projection	2D slice	3D volume	3D
	Moving Couch	Yes	No	No	No	No	No
	Ionizing Radiation	Yes	No	Yes	No	No	No
	Preferred Scan Directions^	Axi	Axi/Sag/Cor	Obl/Sag/Cor	Obl/Sag/Cor	Sag/Cor	Sag/Cor
	3D Recon#	FBP	iFFT	NA	NA	iFFT	iFFT
	4D Recon#	Binning	Binning	NA	NA	NA	SR
Contrast	Lung&	High	High (T2)	Low	Mid	Mid	High (T2)
	Abdomen&	Low	High	Low	High	High	High
Motion	Respiratory Surrogate	External	Internal (/External)	NA	NA	NA	NA
	Cyclical Motion Assumption	Yes	Yes	No	No	No	No
	Binning Artifacts$	High	Low	No	No	No	No
	Multi-breathing Cycles	No	No	Yes	Yes	Yes	Yes

^Preferred scan directions include Axial (Axi), Sagittal (Sag), Coronal (Cor), and oblique (Obl). The coronal scan is often used in 3D cine due to shorter anterior-posterior separation of the human body.

#Reconstruction methods using the filtered back project (FBP), inverse fast Fourier transform (iFFT), and super-resolution (SR) methods.

&T1 or balanced steady-state free-precession (bSSFP) MR contrasts are used for real-time scan.

$High binning artifacts for irregular breathers in 4DCT. Low binning artifacts in RC-4DMRI when using an internal navigator with the Cartesian acquisition and No binning artifacts when using self-navigator in the Golden-angle radial acquisition.

surrogate (14). Van de Lindt et al. reported a self-sorting coronal 4DMRI technique for MR-Linac (15).

The reconstruction of RC-4DMRI in k-space can also be categorized as either prospective or retrospective. Akçakaya et al. developed a 4DMRI k-space respiratory gating method using an internal navigator as the surrogate to prospectively gate the acquisition of the central k-space data (16). Liu et al. developed a strategy that retrospectively reorders k-space data of MR images based on respiratory trajectories, allowing for finer segmentation of data in the time domain (17).

Recently, efforts have been devoted to making RC-4DMRI scan more efficient with higher image resolution and quality. Feng et al. developed a golden-angle radial acquisition technique with compressed sensing (CS) using local self-navigator(s) to resolve both respiratory and cardiac motions in RC-4DMRI (or 5D MRI) (18–20). Because of using radial acquisition, which is insensitive to motions, the RC-4DMRI images are essentially free from binning artifacts. Golden-angle acquisition and multi-navigators facilitate the CS scheme of reconstruction to resolve both respiratory and cardiac motions. Similarly, Wang et al. reported a spatiotemporal k-space scan and sorting technique to enhance RC-4DMRI image quality (21). Han et al. reported a rotating Cartesian k-space (ROCK) 4DMRI method that provides a $50 \times 40 \times 30$ cm^3 field of view, $1.2 \times 1.2 \times 1.6$ mm^3 voxel size, 8 respiratory states, and 5 min scan time (22, 23). This method was equivalent to spiral acquisition and was validated with a 4D phantom and tested in volunteers and patients, producing 1.0 ± 0.5 mm error at the diaphragm in comparison with 2D cine image. Mickevicius and Paulson compared 4 different 4DMRI image reconstruction algorithms and found that the two with CS outperformed the conventional techniques (24).

A similar conclusion was reported by Weiss et al. comparing conventional and CS-based liver scans using navigator-gated 4DMRI acquisition with a $1.2 \times 1.2 \times 3.0$ mm^3 resolution (25). Van Reeth et al. studied the proof of concept of an SR approach to achieve isotropic image resolution by combining anisotropic scans (26). Freedman et al. also developed an SR approach to gain T2w RC-4DMRI with an isotropic resolution ($1.0 \times 1.0 \times 1.0$ mm^3) by combining 2 acquisitions in axial and sagittal scans at $1.5 \times 1.5 \times 5.0$ mm^3 resolution (27).

So far, the image resolution is approaching or exceeding that of 4DCT, while the binning artifacts have been substantially reduced or virtually eliminated and scanning time has become clinically acceptable. With the advantages of soft-tissue contrast enhancement and elimination of ionizing radiation, RC-4DMRI technique has been increasingly tested in the clinic and have the potential to play a role in clinical applications (10, 11, 28, 29). However, RC-4DMRI shares the same limitation of one breathing cycle as 4DCT, while using a snapshot of patient respiration as the average motion has been questioned for its validity and reliability (30–32).

TIME-RESOLVED (TR) 4DMRI

Breathing irregularity has been recognized as a major clinical issue in radiotherapy because it causes a tumor to move differently from the motion assessment achieved at treatment simulation, so that the treatment delivery may not follow the treatment plan. Clinically, it has been observed that substantial breathing irregularities may occur, leading to a possible large variation in tumor motion (30–32). In addition, patient breathing

FIGURE 1 | A schematic graph of three reconstruction methods for time-resolved (TR) 4DMRI. **(A)** Super-resolution (SR) approach, **(B)** dynamic keyhole approach, and **(C)** motion modeling approach.
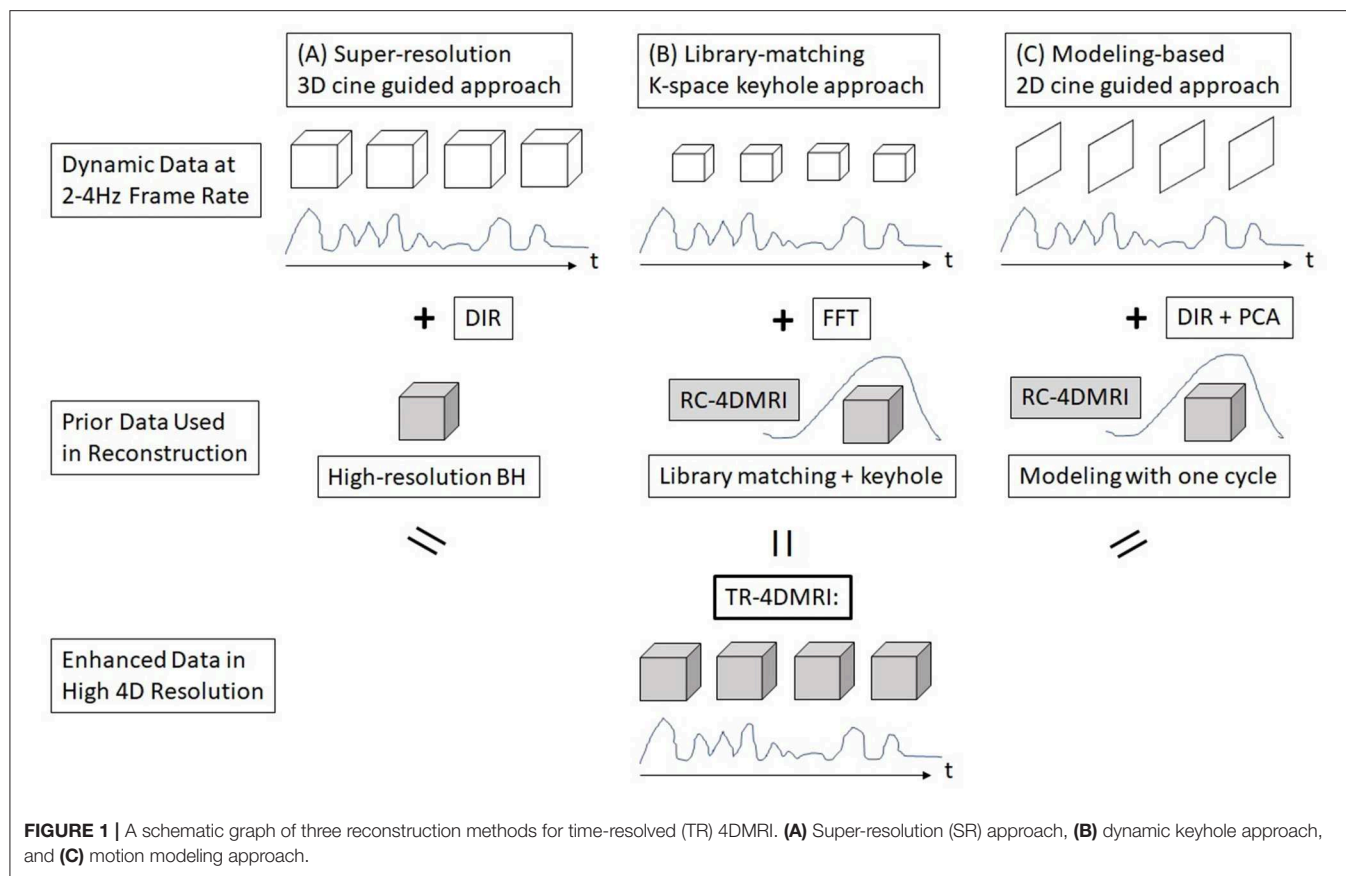
irregularities during motion simulation may affect RC-4DMRI image quality, which further affects tumor delineation, although it has been improved using internal navigator as the respiratory surrogate (12). On the contrary, TR-4DMRI provides multi-breath motion assessment and may immune from breathing irregularities if it is based on SR reconstruction. The major advantages of TR-4DMRI are summarized in **Table 1**. As direct acquisition of 3D cine suffers from low spatial resolution, three alternative methods have been developed to reconstruct TR-4DMRI with a clinically-adequate spatiotemporal resolution. Taking the advantage of 4D patient anatomy redundancy, incomplete but near real-time scans are sufficient to recover the missing information from *a priori* knowledge of the patient for TR-4DMRI reconstruction, as illustrated in **Figure 1**. It should be emphasized that only the SR approach (**Figure 1A**) is independent of RC-4DMRI without assuming motion periodicity, while the other two approaches depend on RC-4DMRI in various degrees for library building (**Figure 1B**) or motion modeling (**Figure 1C**).

Super-Resolution (SR) 3D-Cine-Guided TR-4DMRI Reconstruction

Super-resolution is a concept that has been proven effective to enhance the resolution of an imaging modality beyond its physical limitation (33). It achieves this objective by combining two image sets with complementary resolution strength using an independent method so that both strengths

will appear in the final synthesized image. The SR concept was applied to overcome the physical limitation of dynamic 3D cine MRI (8).

Li et al. reported an SR approach to reconstruct TR-4DMRI by combining two sets of MRI images with high temporal resolution (3D cine at 2 Hz and $5 \times 5 \times 5$ mm^3) in free-breathing (FB) and high spatial resolution in breath-hold (BH, at $2 \times 2 \times 2$ mm^3) through DIR (8), as shown in **Figure 1A**. Therefore, the resulting TR-4DMRI image will have a high spatiotemporal resolution (2 Hz and $2 \times 2 \times 2$ mm^3). In this approach, the dynamic FB 3D cine serves as targeting templates for DIR to map the high-resolution tissue texture from BH to FB images. Because the FB image records the actual organ motion without any restriction on the motion type, so this approach can image both regular or irregular organ motions, including breathing irregularities in multiple respiratory cycles. This method has been improved with enhanced deformation range for respiratory motion (9). Furthermore, the MR contrasts can be extended beyond T1w, including T2w TR-4DMRI (34). Therefore, the SR-based TR-4DMRI technique is promising to provide the accurate history and statistics of actual GTV/OAR motions during treatment simulation and/or treatment delivery.

Library-Matching Dynamic Keyhole TR-4DMRI Reconstruction

The dynamic keyhole method is derived from the conventional keyhole approach, a view-sharing technique, which divides the

k-space into the central (low frequency) and peripheral (high frequency) regions, where only the central data need to be newly acquired and updated while the peripheral data can be acquired separately and shared. The dynamic keyhole method requires anatomical matching between the central and peripheral k-space data so that the aliasing artifacts caused by anatomy mismatch can be minimized (35, 36).

Lee et al. reported a dynamic keyhole method by image matching at the moving diaphragm in a pre-acquired high-resolution RC-4DMRI library (37, 38). A 1D keyhole was applied to combine the central and peripheral k-space data to produce a dynamic TR-4DMRI image set. Liu et al. studied the dynamic volumetric keyhole method as a k-space SR approach for accelerated TR-4DMRI reconstruction without DIR (39). The RC-4DMRI image was applied to create the motion library with high-resolution, however the library can also be created by the TR-4DMRI images. Therefore, limited dependency of this approach to the RC-4DMRI exists. The major advantage of this approach is the fast reconstruction comparing with DIR-based reconstruction.

Comparing with CS, the keyhole method is inferior in both image quality and acceleration (36). Using statistical power in resolving sparse, incoherent signals is superior to mechanically-separated high- and low-frequency regions in the k-space. Yip et al. incorporated a dynamic view-sharing technique into the CS framework and developed a sliding-window prior-data-assisted CS (SW-PDACS) technique to track lung tumor motion (40, 41). In this approach, the k-space was divided into 3 regions, central, middle and peripheral regions for random sampling covering different parts of the k-space, allowing partial k-space update continuously. Therefore, dynamic cine can be acquired with reduced sampling and reconstructed by view sharing with acceptable image quality.

Model-Based 2D-Cine-Guided TR-4DMRI Reconstruction

A third method to reconstruct TR-4DMRI image is based on a motion model that is built on the RC-4DMRI and dynamic 2D cine to guide the model to deform to provide 4D volumetric images. Therefore, this method is fully dependent on RC-4DMRI. Dynamic 2D cine has been utilized for motion assessment and MR-guided radiotherapy on an MR-Linac (MRL) system (42, 43). The frame rate of 4 Hz is available in current commercial systems, although recent studies have shown that the temporal resolution can be further improved (6, 44).

Although 2D cine images are insufficient for volumetric motion assessment of GTV and OAR for treatment planning or treatment delivery, the missing volumetric data can be obtained from prior 4DMRI scans of the same patients. Harris et al. reported a method to retrieve the volumetric information from RC-4DMRI by building a patient-specific respiratory motion model, which can be deformed using the dynamic 2D cine as the guidance (45). This method requires RC-4DMRI, DIR, and principal component analysis (PCA) to build the motion model, similar to a method that was developed in 4DCT (46). Stemkens et al.

reported the same technique of TR-4DMRI but focusing on abdominal tumor (47) and 4DMRI can be also used for dynamic contrast-enhanced (DCE) imaging for better tumor delineation (48).

Other 2D cine-guided reconstruction methods of TR-4DMRI were also reported under certain clinical conditions and assumptions. Paganelli et al. acquired interleaved orthogonal 2D cine images, deformed them to an RC-4DMRI library, and extrapolated the 2D displacement vector fields (DVFs) for 3D image reconstruction (49). When deformation is small, 2D cine may be close to the same phase as a 3D image within volumetric RC-4DMRI (50). Park et al. used local rigid registration in a small region of interest that contains the tumor to retrieve the tumor motion to study internal-external motion correlation (51). The method served the purpose of extracting the tumor motion, but it may inherit a higher degree of uncertainty for full-image reconstruction, especially for motion that does not correlate well with respiration.

CLINICAL EVALUATIONS AND APPLICATIONS OF RC- AND TR-4DMRI

High soft-tissue contrast in RC-4DMRI facilitates tumor/organ delineation and registration for treatment planning. Zhang et al. reported organ segmentation based on T2w RC-4DMRI (28), including the heart, lungs, liver, and stomach in 10 volunteers and evaluated manual and DIR-propagated organ segmentation using STAPLE algorithm. A 95% confident-level ground truth was created to quantify the quality of individual contour with specificity, sensitivity, and Jaccard index. The DIR-propagated contours were found as good as human contours owing to the high soft-tissue contrasts in T2w 4DMRI. Zhang et al. also reported lung tumor delineation in 10 patients from 6 radiation oncologists and compared the results with those from 4DCT and T1w breath-hold images (29). All images were acquired on the simulation day within 2 h. It was found that T2w RC-4DMRI produced a similar GTV to 4DCT but with a much smaller variation among physicians, while T1w MRI based GTV is about 25% smaller. In addition, the tumor motion variation can be quite large, leading to a very different ITV. Gao et al. studied an accelerated 4DMRI for treatment planning (44). An abdominal tumor and two kidneys were delineated and compared between two acceleration imaging techniques with quantification of positioning, volume difference, Dice similarity index and mean distance to the agreement. Liu et al. explored the 4D diffusion-weighted imaging (DWI) MRI imaging for tumor delineation (52). A 4D digital phantom and volunteers were tested for the feasibility. The clinical workflow for RC-4DMRI has been investigated for MR-only treatment planning since only one-cycle motion is provided, like 4DCT (53), by converting MR voxel intensity to tissue electronic density for dose calculation and generating digitally-reconstructed radiography (DRR) with visualized fiducials for patient setup (43).

Assessment of radiotherapy treatment has also been performed using a TR-4DMRI technique since it records

exactly what happens during the beam-on time in MRL treatment. Thomas et al. reported the patient intra- and inter-fractional motion variations using dynamic 2D cine during 3-5 fraction SBRT treatments (30). Mostafaei et al. studied that the localization of the gallbladder is affected by both respiratory and peristaltic motions (32). The 2D cine images can be converted to volumetric TR-4DMRI using the DIR-PCA approach (**Figure 1C**) for retrospective treatment evaluation. Kontaxis et al. reported a strategy to perform online intrafraction replanning for free-breathing stereotactic body radiation therapy using MRL (54). The dosimetry consequence of motion variation from treatment planning viewpoint was studied using RC- and TR-4DMRI, in comparison with the ITV method based on 4DCT (55). Substantial dosimetry variations in ITV-based planning were found when the tumor motion range varied by 5 mm, and this is worthwhile for further evaluations.

In scanning proton therapy, motion interplay effect was found substantial (56) and respiratory-gated proton therapy may be a viable solution (57). Dolde et al. applied five repeated RC-4DMRI to simulate and evaluate motion variation and dose delivery in proton therapy (57). It was also found that the residual error of 2-3 mm in DIR has a large impact on dose assessment owing to the high dose conformality with a sharp dose falloff outside of the target for most proton plans (58).

For patient motion simulation, Stam et al. reported using 2D cine MRI to characterize kidney motions in FB (59). Park et al. reported that using an external surrogate to predict an internal tumor motion may suffer from its insensitivity to internal motion variation and a phase mismatch (51). Wilms et al. reported using multivariate regression approaches for diffeomorphic estimation of internal tumor motion based on surrogates (60). Milewski et al. reported a large phase shift between the external and internal motion based on the internal navigator echo (1D cine) and bellows data during FB 4DMRI acquisition and successfully enhanced the external-internal motion correlation by correcting the phase shift (61). Interestingly, the phase shift was found to be relatively stable over 7–13 min despite breathing irregularities. These studies could lead to the development of a robust patient-specific motion model for respiratory gating in the clinic.

TR-4DMRI provides actual patient breathing motion images over multiple breathing cycles, and therefore serves as an imaging tool for real-time motion monitoring in MRL and provides motion data for building a patient-specific breathing motion model for tumor motion prediction in non-MRL systems. Clinically, the scanning time of TR-4DMRI is determined by how much patient motion statistics is needed. This technique suggests that dynamic 3D/2D cine can be converted to volumetric TR-4DMRI with adequate spatiotemporal resolution for more clinical evaluations, including GTV and OAR motion and motion variations for radiotherapy treatment planning and delivery. An MR simulator or an MRL provides simulation or treatment motion images, which are useful for treatment planning, assessment of treatment delivery, and building a patient-specific multi-breath motion model.

The technical development and clinical application of RC- and TR-4DMRI are currently at their infancy and require further explorations to fully realize their potentials in radiotherapy with

new clinical workflows. Like any other techniques, the RC- and TR-4DMRI have their own limitations, which may lead to future research for improvements. For RC-4DMRI, the major limitations are a single breathing cycle, long acquisition time, and minor MR image distortion, while image resolution has been substantially improved to reach or exceed that of 4DCT. For TR-4DMRI, the major limitations include spatial image uncertainty and image reconstruction time owing to the DIR-based method, while the k-space or GPU-based reconstruction approaches may provide viable solutions. In clinical applications, physician training on target delineation based on 4DMRI images is essential to fully realize the value of the new techniques, because most radiation oncologists are trained for CT-based target delineation. Currently, dedicated MR simulators are only available in a few academic institutions but not in community clinics/hospitals, while MRL machines are even fewer in the world. Regardless of the dimension and form of MRI images, radiotherapy applications require reliable conversion from MR voxel intensity to tissue electron density for accurate radiation dose computation, especially for MR-only treatment planning. Despite these technical and financial limitations, RC- and TR-4DMRI are promising to play a role in radiotherapy because the unparallel ability to differentiate tumorous tissue from its surrounding normal tissues and to image patient continuously without motion periodicity assumption.

SUMMARY AND FUTURE PERSPECTIVE

Both RC- and TR-4DMRI have been developed in recent years and continued to be further studied in the future, especially for high spatial-resolution RC-4DMRI with reduced binning artifacts and high spatiotemporal-resolution TR-4DMRI with improved reconstruction. These 4DMRI techniques have been and will be evaluated under clinical conditions for applications, including tumor/organ delineation and motion assessment for treatment planning and treatment dose evaluation. These allow online or offline adaptive planning and treatment, using 4DMRI-based assessment of the delivered dose to a mobile tumor and surrounding OARs. High soft-tissue contrast and high spatial resolution of RC-4DMRI are useful to improve clinical target delineation, especially in the central thoracic and abdominal regions where 4DCT suffers from the inability to distinguish a tumor from surrounding healthy tissues. Because of the one-cycle motion image of RC-4DMRI similar to 4DCT and the fact of 5 years in advance of TR-4DMRI development, RC-4DMRI is more likely to be first applied to the clinic in the near future.

The main advantage of TR-4DMRI over RC-4DMRI is that it provides multi-breath motion images during the patient simulation and/or treatment. As common breathing irregularities may cause large uncertainties in radiotherapy, the TR-4DMRI technique can serve as a tool to assess and monitor motion irregularities in MRL, aiming to improve treatment accuracy. In light of MRL development, there is a strong clinical need to further develop TR-4DMRI and to make it clinically available to prospectively guide and retrospectively assess radiation

therapy treatment. Based on the multi-breath motion data from the simulation, a patient-specific motion model can be built that would be useful to provide intrafractional tumor motion guidance for non-MR Linac systems. With more representative motion assessment from TR-4DMRI, respiratory gating or tumor tracking could reliably be applied so that the motion margin can be reduced, resulting in more OAR sparing and thus allowing more potent dose prescription to treat a mobile tumor.

AUTHOR CONTRIBUTIONS

GL wrote the final version of the article. YL and XN participated drafting the initial version.

FUNDING

REFERENCES

1. Keall PJ, Mageras GS, Balter JM, Emery RS, Forster KM, Jiang SB, et al. The management of respiratory motion in radiation oncology report of AAPM Task Group 76. *Med Phys.* (2006) 33:3874–900. doi: 10.1118/1.2349696

2. Li G, Citrin D, Camphausen K, Mueller B, Burman C, Mychalczak B, et al. Advances in 4D medical imaging and 4D radiation therapy. *Technol Cancer Res Treat.* (2008) 7:67–81. doi: 10.1177/153303460800700109

3. Li G, Mageras G, Dong L, Mohan R. Image-guided radiation therapy. In: Khan FM, Gerbi BJ, editors. *Treatment Planning in Radiation Oncology.* 4th ed. Philadelphia, PA: Lippincott Williams & Wilkins (2016). p. 229–58.

4. Plathow C, Klopp M, Schoebinger M, Thieke C, Fink C, Puderbach M, et al. Monitoring of lung motion in patients with malignant pleural mesothelioma using two-dimensional and three-dimensional dynamic magnetic resonance imaging: comparison with spirometry. *Invest Radiol.* (2006) 41:443–8. doi: 10.1097/01.rli.0000208222.03256.ba

5. Tsao J, Kozerke S. MRI temporal acceleration techniques. *J Magn Reson Imaging.* (2012) 36:543–60. doi: 10.1002/jmri.23640

6. Hollingsworth KG. Reducing acquisition time in clinical MRI by data undersampling and compressed sensing reconstruction. *Phys Med Biol.* (2015) 60:R297–322. doi: 10.1088/0031-9155/60/21/R297

7. Wang H, Peng H, Chang Y, Liang D. A survey of GPU-based acceleration techniques in MRI reconstructions. *Quant Imaging Med Surg.* (2018) 8:196–208. doi: 10.21037/qims.2018.03.07

8. Li G, Wei J, Kadbi M, Moody J, Sun A, Zhang S, et al. Novel super-resolution approach to time-resolved volumetric 4-dimensional magnetic resonance imaging with high spatiotemporal resolution for multi-breathing cycle motion assessment. *Int J Radiat Oncol Biol Phys.* (2017) 98:454–62. doi: 10.1016/j.ijrobp.2017.02.016

9. Li G, Sun A, Nie X, Moody J, Huang K, Zhang S, et al. Introduction of a pseudo demons force to enhance deformation range for robust reconstruction of super-resolution time-resolved 4DMRI. *Med Phys.* (2018) 45:5197–207. doi: 10.1002/mp.13179

10. Stemkens B, Paulson ES, Tijssen RHN. Nuts and bolts of 4D-MRI for radiotherapy. *Phys Med Biol.* 63:21TR01. (2018). doi: 10.1088/1361-6560/aae56d

11. Paganelli C, Whelan B, Peroni M, Summers P, Fast M, van de Lindt T, et al. MRI-guidance for motion management in external beam radiotherapy: current status and future challenges. *Phys Med Biol.* (2018) 63:22TR03. doi: 10.1088/1361-6560/aaebcf

12. Li G, Wei J, Olek D, Kadbi M, Tyagi N, Zakian K, et al. Direct comparison of respiration-correlated four-dimensional magnetic resonance imaging reconstructed using concurrent internal navigator and external bellows. *Int J Radiat Oncol Biol Phys.* (2017) 97:596–605. doi: 10.1016/j.ijrobp.2016.11.004

13. Hu Y, Caruthers SD, Low DA, Parikh PJ, Mutic S. Respiratory amplitude guided 4-dimensional magnetic resonance imaging. *Int J Radiat Oncol Biol Phys.* (2013) 86:198–204. doi: 10.1016/j.ijrobp.2012.12.014

14. Cai J, Chang Z, O'Daniel J, Yoo S, Ge H, Kelsey C, et al. Investigation of sliced body volume (SBV) as respiratory surrogate. *J Appl Clin Med Phys.* (2013) 14:3987. doi: 10.1120/jacmp.v14i1.3987

15. van de Lindt T, Sonke JJ, Nowee M, Jansen E, van Pelt V, van der Heide U, et al. A self-sorting coronal 4D-MRI method for daily image guidance of liver lesions on an MR-LINAC. *Int J Radiat Oncol Biol Phys.* (2018) 102:875–84. doi: 10.1016/j.ijrobp.2018.05.029

16. Akcakaya M, Gulaka P, Basha TA, Ngo LH, Manning WJ, Nezafat R. Free-breathing phase contrast MRI with near 100% respiratory navigator efficiency using k-space-dependent respiratory gating. *Magn Reson Med.* (2014) 71:2172–9. doi: 10.1002/mrm.24874

17. Liu Y, Yin FF, Chen NK, Chu ML, Cai J. Four dimensional magnetic resonance imaging with retrospective k-space reordering: a feasibility study. *Med Phys.* (2015) 42:534–41. doi: 10.1118/1.4905044

18. Feng L, Axel L, Chandarana H, Block KT, Sodickson DK, Otazo R. XD-GRASP: Golden-angle radial MRI with reconstruction of extra motion-state dimensions using compressed sensing. *Mag Reson Med.* (2015) 75:775–88. doi: 10.1002/mrm.25665

19. Feng L, Benkert T, Block KT, Sodickson DK, Otazo R, Chandarana H. Compressed sensing for body MRI. *J Magn Reson Imaging.* (2017) 45:966–87. doi: 10.1002/jmri.25547

20. Feng L, Coppo S, Piccini D, Yerly J, Lim RP, Masci PG, et al. 5D whole-heart sparse MRI. *Magne Reson Med.* (2018) 79:826–38. doi: 10.1002/mrm.26745

21. Wang C, Subashi E, Yin FF, Chang Z, Cai J. A spatiotemporal-constrained sorting method for motion-robust 4D-MRI: a feasibility study. *Int J Radiat Oncol Biol Phys.* (2019) 103:758–66. doi: 10.1016/j.ijrobp.2018.10.004

22. Han F, Zhou Z, Cao M, Yang Y, Sheng K, Hu P. Respiratory motion-resolved, self-gated 4D-MRI using rotating cartesian k-space (ROCK). *Med Phys.* (2017) 44:1359–68. doi: 10.1002/mp.12139

23. Han F, Zhou Z, Du D, Gao Y, Rashid S, Cao M, et al. Respiratory motion-resolved, self-gated 4D-MRI using Rotating Cartesian K-space (ROCK): Initial clinical experience on an MRI-guided radiotherapy system. *Radiother Oncol.* (2018) 127:467–73. doi: 10.1016/j.radonc.2018.04.029

24. Mickevicius NJ, Paulson ES. Investigation of undersampling and reconstruction algorithm dependence on respiratory correlated 4D-MRI for online MR-guided radiation therapy. *Phys Med Biol.* (2017) 62:2910–21. doi: 10.1088/1361-6560/aa54f2

25. Weiss J, Notohamiprodjo M, Martirosian P, Taron J, Nickel MD, Kolb M, et al. Self-gated 4D-MRI of the liver: Initial clinical results of continuous multiphase imaging of hepatic enhancement. *J Magn Reson Imaging.* (2018) 47:459–67. doi: 10.1002/jmri.25784

26. Van Reeth E, Tan CH, Tham IW, Poh CL. Isotropic reconstruction of a 4-D MRI thoracic sequence using super-resolution. *Magn Reson Med.* (2015) 73:784–93. doi: 10.1002/mrm.25157

27. Freedman JN, Collins DJ, Gurney-Champion OJ, McClelland JR, Nill S, Oelfke U, et al. Super-resolution T2-weighted 4D MRI for image guided radiotherapy. *Radiother Oncol.* (2018) 129:486–93. doi: 10.1016/j.radonc.2018.05.015

28. Zhang J, Markova S, Garcia A, Huang K, Nie X, Choi W, et al. Evaluation of automatic contour propagation in T2-weighted 4DMRI for normal-tissue motion assessment using internal organ-at-risk volume (IRV). *J Appl Clin Med Phys.* (2018) 19:598–608. doi: 10.1002/acm2.12431

29. Zhang J, Srivastava S, Wang C, Beckham T, Johnson C, Dutta P, et al. Clinical evaluation of 4D MRI in the delineation of gross and internal tumor volumes in comparison with 4DCT. *J. Appl Clin Med Phys.* (2019) 20:51–60. doi: 10.1002/acm2.12699

30. Thomas DH, Santhanam A, Kishan AU, Cao M, Lamb J, Min Y, et al. Initial clinical observations of intra- and interfractional motion variation in MR-guided lung SBRT. *Br J Radiol.* (2017) 91:20170522. doi: 10.1259/bjr.20170522

31. Dhont J, Vandemeulebroucke J, Burghelea M, Poels K, Depuydt T, Van Den Begin R, et al. The long- and short-term variability of breathing induced

tumor motion in lung and liver over the course of a radiotherapy treatment. *Radiother Oncol.* (2018) 126:339–46. doi: 10.1016/j.radonc.2017.09.001

32. Mostafaei F, Tai A, Omari E, Song Y, Christian J, Paulson E, et al. Variations of MRI-assessed peristaltic motions during radiation therapy. *PLoS ONE.* (2018) 13:e0205917. doi: 10.1371/journal.pone.0205917

33. Robinson MD, Chiu SJ, Lo JY, Toth CA, Izatt JA, Farsiu S. Novel applications of super-resolution in medical imaging. In: Milanfar P, editor. *Super-Resolution Imaging.* Boca Raton, FL: CRC Press (2010). p. 383–412.

34. Li G, Nie X, Kadbi M, Saleh Z, Zakian K, Mechalakos J, et al. Feasibility of achieving T2w super-resolution time-resolved (TR) 4DMRI using T1w TR-4DMRI as the guidance. *Med Phys.* (2018) 45.

35. Chen Z, Zhang J, Pang KK. Adaptive keyhole methods for dynamic magnetic resonance image reconstruction. *Comput Med Imaging Graph.* (2007) 31:458–468. doi: 10.1016/j.compmedimag.2007.04.005

36. Geethanath S, Gulaka PK, Kodibagkar VD. Acceleration of conventional data acquisition in dynamic contrast enhancement: comparing keyhole approaches with compressive sensing. *Critical reviews in biomedical engineering.* 2014;42(6):437-450. doi: 10.1615/CritRevBiomedEng.2014010669

37. Lee D, Pollock S, Whelan B, Keall P, Kim T. Dynamic keyhole: a novel method to improve MR images in the presence of respiratory motion for real-time MRI. *Med Phys.* (2014) 41:072304. doi: 10.1118/1.4883882

38. Lee D, Greer PB, Pollock S, Kim T, Keall P. Quantifying the accuracy of the tumor motion and area as a function of acceleration factor for the simulation of the dynamic keyhole magnetic resonance imaging method. *Med Phys.* (2016) 43:2639. doi: 10.1118/1.4947508

39. Liu Y, Zakian K, Kadbi M, Deasy JO, Li G. Feasibility of fast k-space reconstruction of time-resolved 4DMRI with adequate spatiotemporal resolution using post-acquisition volumetric dynamic keyhole method. *Med Phys.* (2017) 44:3156.

40. Yip E, Yun J, Wachowicz K, Heikal AA, Gabos Z, Rathee S, et al. Prior data assisted compressed sensing: a novel MR imaging strategy for real time tracking of lung tumors. *Med Phys.* (2014) 41:082301. doi: 10.1118/1.4885960

41. Yip E, Yun J, Wachowicz K, Gabos Z, Rathee S, Fallone BG. Sliding window prior data assisted compressed sensing for MRI tracking of lung tumors. *Med Phys.* (2017) 44:84–98. doi: 10.1002/mp.12027

42. Bainbridge H, Salem A, Tijssen RHN, Dubec M, Wetscherek A, Van Es C, et al. Magnetic resonance imaging in precision radiation therapy for lung cancer. *Transl Lung Cancer Res.* (2017) 6:689–707. doi: 10.21037/tlcr.2017.09.02

43. Das IJ, McGee KP, Tyagi N, Wang H. Role and future of MRI in radiation oncology. *Br J Radiol.* (2019) 92:20180505. doi: 10.1259/bjr.20180505

44. Gao Y, Zhou Z, Han F, Cao M, Shaverdian N, Hegde JV, et al. Accelerated 3D bSSFP imaging for treatment planning on an MRI-guided radiotherapy system. *Med Phys.* (2018) 45:2595–602. doi: 10.1002/mp.12924

45. Harris W, Yin FF, Wang C, Zhang Y, Cai J, Ren L. Accelerating volumetric cine MRI (VC-MRI) using undersampling for real-time 3D target localization/tracking in radiation therapy: a feasibility study. *Phys Med Biol.* (2017) 63:01NT01. doi: 10.1088/1361-6560/aa9746

46. Li R, Lewis JH, Jia X, Zhao T, Liu W, Wuenschel S, et al. On a PCA-based lung motion model. *Phys Med Biol.* (2011) 56:6009–30. doi: 10.1088/0031-9155/56/18/015

47. Stemkens B, Tijssen RH, de Senneville BD, Lagendijk JJ, van den Berg CA. Image-driven, model-based 3D abdominal motion estimation for MR-guided radiotherapy. *Phys Med. Biol.*(2016) 61:5335–55. doi: 10.1088/0031-9155/61/14/5335

48. Stemkens B, Prins FM, Bruijnen T, Kerkmeijer LGW, Lagendijk JJW, van den Berg CAT, et al. A dual-purpose MRI acquisition to combine 4D-MRI and dynamic contrast-enhanced imaging for abdominal radiotherapy planning. *Phys Med Biol.* (2019) 64:06NT02. doi: 10.1088/1361-6560/ab0295

49. Paganelli C, Lee D, Kipritidis J, Whelan B, Greer PB, Baroni G, et al. Feasibility study on 3D image reconstruction from 2D orthogonal cine-MRI for MRI-guided radiotherapy. *J Med Imaging Radiation Oncol.* (2018) 62:389–400. doi: 10.1111/1754-9485.12713

50. Bjerre T, Crijns S, af Rosenschöld PM, Aznar M, Specht L, Larsen R, et al. Three-dimensional MRI-linac intra-fraction guidance using multiple orthogonal cine-MRI planes. *Phys Med Biol.* (2013) 58:4943–50. doi: 10.1088/0031-9155/58/14/4943

51. Park S, Farah R, Shea SM, Tryggestad E, Hales R, Lee J. Simultaneous tumor and surrogate motion tracking with dynamic MRI for radiation therapy planning. *Phys Med Biol.* (2018) 63:025015. doi: 10.1088/1361-6560/aaa20b

52. Liu Y, Zhong X, Czito BG, Palta M, Bashir MR, Dale BM, et al. Four-dimensional diffusion-weighted MR imaging (4D-DWI): a feasibility study. *Med Phys.* (2017) 44:397–406. doi: 10.1002/mp.12037

53. To DT, Kim JP, Price RG, Chetty IJ, Glide-Hurst CK. Impact of incorporating visual biofeedback in 4D MRI. *J Appl Clin Med Phys.* (2016) 17:128–37. doi: 10.1120/jacmp.v17i3.6017

54. Kontaxis C, Bol GH, Stemkens B, Glitzner M, Prins FM, Kerkmeijer LGW, et al. Towards fast online intrafraction replanning for free-breathing stereotactic body radiation therapy with the MR-linac. *Phys Med Biol.* (2017) 62:7233–248. doi: 10.1088/1361-6560/aa82ae

55. Xu L, Nie X, Zhang L, Wang C, Wang P, Srivastava S, et al. Tumor motion and dosimetry variability based on 4DMRI in comparison with 4DCT. In: *ASTRO Annual Meeting.* Chicago (2019). doi: 10.1016/j.ijrobp.2019.06.784

56. Dolde K, Zhang Y, Chaudhri N, Dávid C, Kachelrieß M, Lomax AJ, et al. 4DMRI-based investigation on the interplay effect for pencil beam scanning proton therapy of pancreatic cancer patients. *Radiat Oncol.* (2019) 14:30. doi: 10.1186/s13014-019-1231-2

57. Dolde K, Naumann P, Dávid C, Gnirs R, Kachelrieß M, Lomax AJ, et al. 4D dose calculation for PBS proton therapy of pancreatic cancer using repeated 4DMRI datasets. *Phys Med Biol.* (2018) 63:165005. doi: 10.1088/1361-6560/aad43f

58. Ribeiro CO, Knopf A, Langendijk JA, Weber DC, Lomax AJ, Zhang Y. Assessment of dosimetric errors induced by deformable image registration methods in 4D pencil beam scanned proton treatment planning for liver tumours. *Radiother Oncol.* (2018) 128:174–81. doi: 10.1016/j.radonc.2018.03.001

59. Stam MK, van Vulpen M, Barendrecht MM, Zonnenberg BA, Intven M, Crijns SP, et al. Kidney motion during free breathing and breath hold for MR-guided radiotherapy. *Phys Med Biol.* (2013) 58:2235–45. doi: 10.1088/0031-9155/58/7/2235

60. Wilms M, Werner R, Ehrhardt J, Schmidt-Richberg A, Schlemmer HP, Handels H. Multivariate regression approaches for surrogate-based diffeomorphic estimation of respiratory motion in radiation therapy. *Phys Med Biol.* (2014) 59:1147–64. doi: 10.1088/0031-9155/59/5/1147

61. Milewski A, Olek D, Deasy JO, Rimner A, Li G. Enhancement of long-term external-internal correlation by phase-shift detection and correction based on concurrent external bellows and internal navigator signals. *Adv Radiat Oncol.* (2019) 4:377–89. doi: 10.1016/j.adro.2019.02.001

Pretreatment Prediction of Adaptive Radiation Therapy Eligibility Using MRI-Based Radiomics for Advanced Nasopharyngeal Carcinoma Patients

Ting-ting Yu[1†], Sai-kit Lam[1†], Lok-hang To[1], Ka-yan Tse[1], Nong-yi Cheng[1], Yeuk-nam Fan[1], Cheuk-lai Lo[1], Ka-wa Or[1], Man-lok Chan[1], Ka-ching Hui[1], Fong-chi Chan[1], Wai-ming Hui[1], Lo-kin Ngai[1], Francis Kar-ho Lee[2], Kwok-hung Au[2], Celia Wai-yi Yip[2], Yong Zhang[3] and Jing Cai[1*]

[1] Department of Health Technology and Informatics, Hong Kong Polytechnic University, Hung Hom, Hong Kong, [2] Department of Clinical Oncology, Queen Elizabeth Hospital, Hong Kong, China, [3] Department of Physics, Xiamen University, Xiamen, China

Background and purpose: Adaptive radiotherapy (ART) can compensate for the dosimetric impacts induced by anatomic and geometric variations in patients with nasopharyngeal carcinoma (NPC); Yet, the need for ART can only be assessed during the radiation treatment and the implementation of ART is resource intensive. Therefore, we aimed to determine tumoral biomarkers using pre-treatment MR images for predicting ART eligibility in NPC patients prior to the start of treatment.

Methods: Seventy patients with biopsy-proven NPC (Stage II-IVB) in 2015 were enrolled into this retrospective study. Pre-treatment contrast-enhanced T1-w (CET1-w), T2-w MR images were processed and filtered using Laplacian of Gaussian (LoG) filter before radiomic features extraction. A total of 479 radiomics features, including the first-order ($n = 90$), shape ($n = 14$), and texture features ($n = 375$), were initially extracted from Gross-Tumor-Volume of primary tumor (GTVnp) using CET1-w, T2-w MR images. Patients were randomly divided into a training set ($n = 51$) and testing set ($n = 19$). The least absolute shrinkage and selection operator (LASSO) logistic regression model was applied for radiomic model construction in training set to select the most predictive features to predict patients who were replanned and assessed in the testing set. A double cross-validation approach of 100 resampled iterations with 3-fold nested cross-validation was employed in LASSO during model construction. The predictive performance of each model was evaluated using the area under the receiver operator characteristic (ROC) curve (AUC).

Results: In the present cohort, 13 of 70 patients (18.6%) underwent ART. Average AUCs in training and testing sets were 0.962 (95%CI: 0.961–0.963) and 0.852 (95%CI: 0.847–0.857) with 8 selected features for CET1-w model; 0.895 (95%CI: 0.893–0.896) and 0.750 (95%CI: 0.745–0.755) with 6 selected features for T2-w model; and 0.984 (95%CI: 0.983–0.984) and 0.930 (95%CI: 0.928–0.933) with 6 selected features for joint

T1-T2 model, respectively. In general, the joint T1-T2 model outperformed either CET1-w or T2-w model alone.

Conclusions: Our study successfully showed promising capability of MRI-based radiomics features for pre-treatment identification of ART eligibility in NPC patients.

Keywords: radiomics, nasopharyngeal carcinoma, adaptive radiation therapy, tumor shrinkage, magnetic resonance imaging

INTRODUCTION

Due to the high proximity of the primary NPC tumor to the surrounding critical organs (spinal cord, brainstem, parotid glands) and metastatic neck lymph nodes, NPC is rarely treated surgically; radiation therapy (RT) remains the mainstay of NPC treatment (1). Intensity-modulated radiation therapy (IMRT) with/without induction chemotherapy (IC) or adjuvant chemotherapy (AC) is currently the standard of care for NPC patients (1). In clinical practice, RT treatment plans are tailor-made based on anatomic information of individual patients from their pre-treatment planning computed tomography (CT) images to maximize the radiation dose to tumor while protecting nearby critical structures and maintaining sufficiently high dose coverage to surrounding nodal targets.

However, an abundance of research has shown that significant anatomic and geometric variations are not uncommon throughout the course of RT for NPC due to either body weight loss (BW loss) or tumor regression (2–8). Radiation-induced mucositis is a common and debilitating complication for RT to HNC patients, which can lead to severe pain and difficulty in eating, largely affecting one's nutritional intake and resulting in significant BW loss. A prospective study reported a 37% of BW loss > 5 kg by the end of treatment (9). Patients having significant BW loss tends to accompany with reduced skin separation at various levels of cervical spine and neck (10), causing positional variability due to possible head movement inside the thermoplastic cast. Consequently, such variations would leave the issue of whether the contour deviations induced significant dose deviations in target or organs at risk. For tumor regression, Hu et al. (6) conducted a retrospective study and reviewed the planning CT and re-CT images of 40 re-planned NPC patients and confirmed the significant clinical-target-volume shrinkage of 35.1%. Murat et al. (11) also reported median percentage change in GTV of HNC patients for primary (26.8%), nodal (43.0%), and total (31.2%) GTVs. Indeed, when significant tumor shrinkage occurs, those critical organs might move into the original high dose region, leading to deleterious dosimetric impact on the surrounding organs (3, 4, 12) and/or insufficient dose delivery to targets (4, 13). ART can compensate for these dosimetric impact and maintain desirable therapeutic index. The clinical and dosimetric benefits of ART for HNC and NPC cancer patients have been widely reported (14–17). Yet, the implementation of ART is limited by several reasons. First, the choice to ART can be resource intensive and time-consuming

for repeat imaging, re-contouring, re-planning, and analyzing dosimetric impacts between previous and new treatment plans, adding significant clinical burden and cost of patient care to an oncology center. Hence, performing ART on a patient basis is clinically impractical, especially for some busy units. Second, due to the nature of multifactorial ART eligibility, there is no universal selection protocol for ART that can be applied to all hospitals. In this regard, a huge amount of efforts has been constantly made to identify possible ART criteria for HNC and NPC cancer patients (5–7, 11, 18–21) to facilitate the clinical application of ART. Despite that, the current ART practice in most oncology centers, particularly for those busy units, is not efficient. The need for ART of each patient can now be only assessed during the RT treatment. Therefore, pre-treatment identification of high-risk NPC patients for ART is crucially favorable to achieve optimal personalized RT treatment, enabling radiation oncologists to more effectively and accurately prescribe ART for NPC patients and streamline resources management in clinical settings.

Recently, the field of radiomics together with rapid machine learning paradigms have increasingly gained popularity in the community of medical research, paving the way toward precision and personalized medicine (22). Radiomics, first introduced by Lambin et al. (22), is now shifting the role of medical imaging beyond the traditional diagnostic purposes. It allows for transformation of digitally encrypted medical images into mineable high-dimensional data, which can then be quantitatively analyzed to decode concealed genetic and molecular traits for decision making in oncology (23). While the predictive powers of radiomics in both cancer diagnosis and disease progression have been widely investigated (24–28), an extremely limited effort has yet been made to identify cancer patients for ART. Given the evidence of significant tumor shrinkage between two CT scans along RT treatment for re-planned NPC patients, we hypothesize that radiomic features extracted from 3-dimensional tumor volume contain predictive biomarkers for tumor shrinkage following cancer treatment—an implication for ART.

To our best knowledge, there is no research to include radiomics in predicting ART eligibility for NPC patients and its tumoral predictive biomarkers for ART has not been explored before. The objective of our study was to identify tumoral radiomic features using multi-parametric MR images, which are capable of predicting the ART eligibility for NPC patients. A study flow of current study is shown in **Supplementary Figure 1**.

METHODS AND MATERIALS

A Predefined Hypothesis

Radiomic features extracted from 3-dimensional tumor volume contain predictive biomarkers for tumor shrinkage following cancer treatment—an implication for ART.

Patients

Patient Source

The current research was approved by the Human Subjects Ethics Sub-committee of the Hong Kong Polytechnic University and Kowloon Central/Kowloon East Cluster Research Ethics Committee of the Hospital Authority. This is a retrospective study, based on analyses of anonymized radiographic data and clinical data, the requirement for individual informed consent was waived. A total of 100 newly diagnosed patients with biopsy-proven (II-IVB) NPC (According to 7th edition of American Joint Committee on Cancer/Union for International Cancer Control TNM staging system) who received primary radiation therapy with/without chemotherapy at the Department of Clinical Oncology of Queen Elizabeth Hospital (QEH) between April 2015 and February 2016 were retrospectively reviewed. Based on the inclusion and exclusion criteria (IEC), 70 eligible patients were enrolled in the current study and randomly stratified into training ($n = 51$) and testing ($n = 19$) sets, as illustrated in **Figure 1** (Details of the IEC is described in **Supplementary Material**).

Patient Data

Patient clinical data, including demographic information (age, gender) and tumor characteristics (T stage, N stage, histological subtype); imaging data (planning CT images, pre-treatment CET1-w and T2-w MR images); treatment-related data (contouring data, treatment machine, treatment strategies, dose fractionation scheme); outcome data (re-plan status and any replan-related medical records) were retrospectively collected.

Treatment

In general, patients with early-stage (I-II, $n = 3$) tumors were treated with curative RT alone, while those with advanced-stage (III-IVB, $n = 67$) were treated with radical concurrent chemoradiotherapy (CCRT), with/without IC or AC. Pre-treatment MRI and planning CT scans were performed within a week prior to the start of IC treatment for target delineation and during the last cycle of IC treatment, respectively. In our dataset, 7 out of 70 patients received IC, while only one underwent ART procedures, who subsequently refused further IC after completion of the first cycle due to repeated vomiting. See **Supplementary Material** for details of the chemotherapy and RT regimen.

Clinical Endpoint

The clinical endpoint of this study was defined as the re-plan status of patients: whether or not a patient received ART during RT treatment at the discretion of radiation oncologist.

FIGURE 1 | Inclusion and exclusion criteria used in the current study.
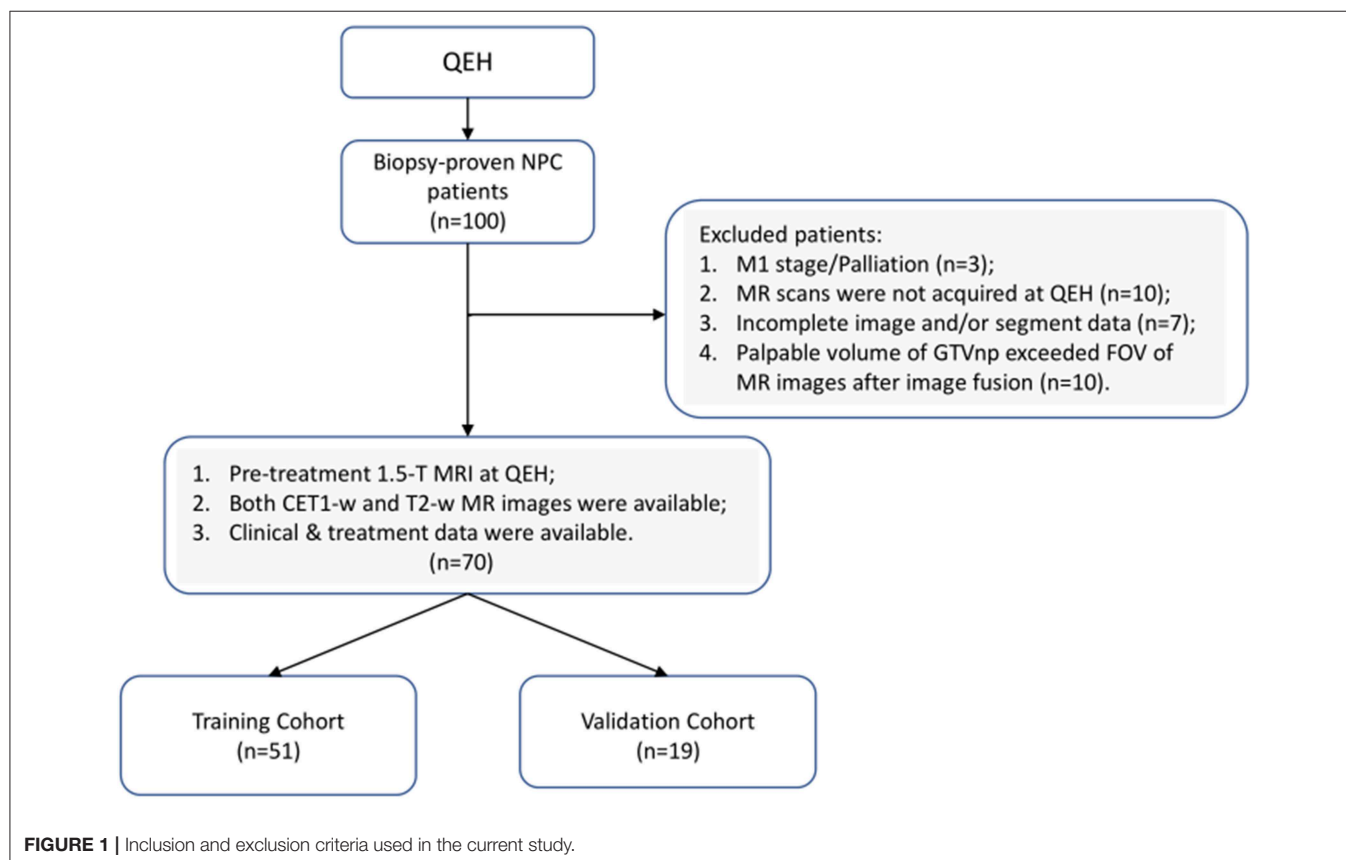
Multifactorial ART Eligibility

A daily megavoltage CT (MVCT) or cone beam CT (CBCT) or planar orthogonal X-rays was taken for all patients to correct for positional variations and to assess anatomic or geometric changes throughout the entire treatment chain. Additionally, weekly records of body weight were made to assess whether significant body weight loss (BWL > 10%) occurred.

The Radiation Oncology team reviewed daily scans on a regular basis, considering BWL of individual patients. When BWL > 10% occurred, possibly accompanied with noted change in body or neck contour, significant lymph nodes regression and/or loss of neck tissue, an adaptive review process was initiated, where the original plan was re-calculated on the MVCT scan for initial dosimetric evaluation to determine whether further actions (re-CT and/or re-plan) or continuous monitoring were appropriate. Patients who did not receive any actions from the first review session were then proceeded with original plan until the next review session for another dosimetric evaluation. On plan review, radiation oncologist assessed the geometric, volumetric and dosimetric variations of both target and organs at risk (OARs) structures through both visual inspection and dosimetric evaluation. The decision to generate a re-plan was at the discretion of the treating radiation oncologist. Considerations influencing ART implementation included risks of insufficient primary and nodal targets coverage, overdose to critical organs (such as spinal cord, optic chiasm, and brainstem), increase of high skin dose areas over neck, and unfit of thermoplastic cast for patient immobilization.

In our dataset, 39 (of 100) patients were initially enrolled into the adaptive review processes, while only 16 ultimately received re-planned procedures. Among the 16 patients, 13 were enrolled in our study, the replans were mostly done during week 4–5 and after the 20th fraction on average A diagram of leading causes for ART implementation are illustrated in **Figure 2**. A detailed qualitative summary of how those 39 patients were screened and selected can be found in **Supplementary Material**.

MRI Acquisition and Segmentation

All 70 patients were scanned with 1.5-T MRI (Avanto, Siemens, Germany) at QEH. We acquired T2-w and CET1-w Digital Imaging and Communications in Medicine (DICOM) images archived using Picture Archiving and Communication System (PACs). The MR images acquisition parameters can be found in **Supplementary Material**. Intravenous contrast enhanced computed tomography (CT) simulation was performed at 3 mm intervals from the vertex to 5 cm below the sternoclavicular notch with a 16-slice Brilliance Big Bore CT (Philips Medical Systems, Cleveland, OH). All segmentations (tumor, nodal volume and other organs-at-risk) were manually delineated on axial CT slices by an experienced radiation oncologist (with >20 years of experience), which was then fused with MR images for further processing.

MRI Image Preprocessing

Before extracting radiomic features, all MR images were processed using 3DSlicer (version 4.11.0). Isotropic resampling was performed by linear interpolation to obtain a voxel size of $1 \times 1 \times 1$ mm to account for variations in scanning parameters between studied MR series. MRI inhomogeneity correction was applied to account for the locally varying intensity using N4ITK algorithm. To ensure meaning comparison of the extracted features values across all patients, intensity normalization was conducted using brainstem as a reference ROI, which was chosen because its signal intensity is comparatively homogeneous. The existing contour of the brainstem structure for RT planning purpose was modified to exclude air. Image discretization with a fixed bin width of 5 to maintain constant intensity resolution across resampled images. Apart from the original images, image reconstructions were performed using Laplacian of Gaussian (LoG) filter with sigma values of 2, 3, 4, 5 mm to extract features at multiple scales of resolution, from fine, medium to coarse.

Feature Extraction and Preprocessing

A total of 479 radiomic features were extracted from GTVnp on CET1-w and T2-w MR images, respectively, using SlicerRadiomics in 3D Slicer (version 4.11.0). A representative example of axial pre-treatment MR images with GTVnp contour is shown in **Figure 3**. Extracted features included shape features ($n = 14$), first-order intensity features ($n = 90$), and texture features ($n = 375$) (See **Supplementary Material** for a detailed listing of extracted features). All extracted radiomics features were centered and scaled to a value with a mean of 0 and a standard deviation of 1 (z-score transformation) before further analysis using R software (version 3.5.2).

Feature Selection and Model Optimization Methodology

To avoid over sensitive model, we removed highly inter-correlated radiomics features. By using the R package "caret," we computed Pearson correlation coefficient (PCC) based on a correlation matrix to quantify the pair-wise correlations. If two radiomic features appeared a strong correlation with an absolute correlation coefficient (r) $\geq$ 0.9, we removed the feature with the largest mean absolute correlation. As a result, we obtained a primary feature set of 53 from 479.

Following this, we applied Least Absolute Shrinkage and Selection Operator (LASSO) algorithm in R package "glmet" to select the most predictive radiomic features based on the ART status of patients in the training set. The LASSO is typically applied to select high-dimensional biomarkers, and coefficients of the regression variables were penalized in the process of regularization to minimize the prediction error. The ratio of patients who did not receive ART ($n = 57$) to those who did ($n = 13$) was 4, approximately. Considering the imbalance data, we adopted our three-step feature screening strategy, as illustrated in **Figure 4**, to construct CET1-w, T2-w, and joint T1-T2 based radiomic models. The first two steps aimed to further eliminate less/least predictive features in terms of their frequency of occurrence among hundreds of generated models. With the reduced features, we performed PCC with $r \geq 0.8$ to avoid highly correlated features in our final models. Lastly, model trainings were performed with reduced number of input features using a double cross-validation approach, similar to the
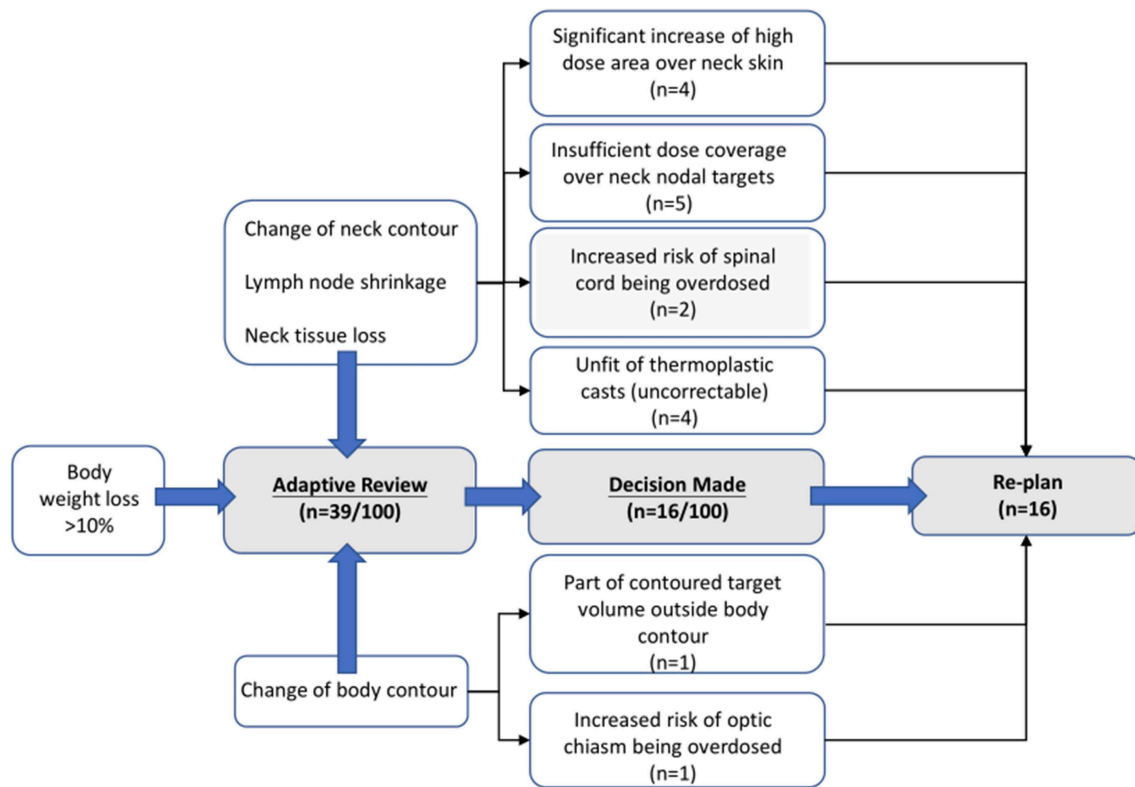
FIGURE 2 | An illustrative example of clinical decision on ART implementation.
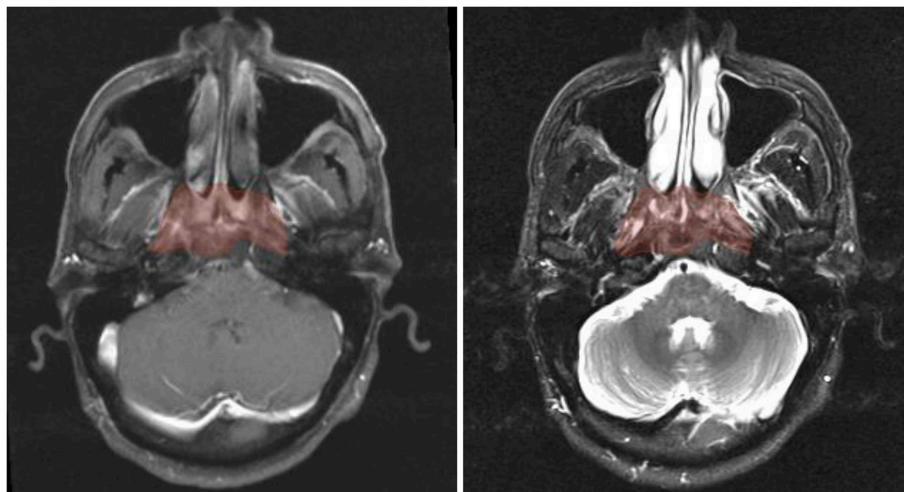
FIGURE 3 | Axial pre-treatment morphological MR images of a 44-year-old man with undifferentiated carcinoma of NPC (T3N2M0). Features of radiomics were extracted from the primary tumor area -GTVnp (red overlay). From left to right: CET1-w and T2-w MR image, respectively.

one adopted by Xu et al. (29) In short, 100 random sampling was conducted to balance the class distribution within the cross-validation partitions, which would result in a distribution of AUC values across the generated models and hence allow us to assess the model performance. A 3-fold nested cross-validation

was performed with 20 repetition to determine the optimal value for the model tuning parameter (λ). As a result, a total of 2,000 models were generated for each input set of features (See **Supplementary Material** for feature screening methodology). In total, 8 sets of radiomic features with number of variables ranging

from 3 to 10 were analyzed for the prediction capability in terms of AUCs using box and whisker plots and 95 percent confidence interval (CI).

Statistical Analysis

The statistical correlations between available clinical data and replan status were assessed using univariate logistic regression. All statistical analyses were performed using R software (version 3.5.2). The following R packages were used: The glmnet package was used for LASSO logistic regression. The caret package was used to perform Pearson correlation study. The ROCR package was employed to perform ROC analysis. All statistical tests were two-sided, and P-values of <0.05 were considered significant.

RESULTS

The demographic and tumor characteristics of 70 NPC patients are summarized in **Table 1**. Thirteen (18.6%) patients who underwent ART procedure were included. There is no statistical association between the available clinical data and re-plan incidence.

Figure 5 displays the AUC distributions for each feature set (from 3 to 10 features). **Figures 5A–C** shows the box and whisker plots of the three types of models (CET1-w, T2-w, and joint T1-T2) for training set; **Figures 5D–F** are for testing set; **Figures 5G–I** visualizes the range of 95% CI of AUCs in both training and testing sets for the three types of models. The optimal feature sets for each type of models were determined considering the overall distribution of AUC values and its stability. When adding one more feature to the current feature set made no/less difference to the AUC values, the current feature set was considered as the optimal feature set that would give optimal predictive performance of our models. Selected features for each model are listed in **Table 2**.

Average AUC values in training and testing sets were 0.962 (95%CI: 0.961–0.963) and 0.852 (95%CI: 0.847–0.857) with 8 selected features for CET1-w model; 0.895 (95%CI: 0.893–0.896) and 0.750 (95%CI: 0.745–0.755) with 6 selected features for T2-w model; and 0.984 (95%CI: 0.983–0.984) and 0.930 (95%CI: 0.928–0.933) with 6 selected features for joint T1-T2 model, respectively.

DISCUSSION

We successfully revealed the predictive capability of MRI-based radiomics in ART eligibility using our dataset. Eight features were identified for CET1-w model, including 2 shape features (sphericity, maximum 2D diameter slice) and 6 LoG-based features which include 3 first-order features (kurtosis, skewness) and 3 texture features (GLCM and GLDM). Six features were selected for T2-w model, including 2 shape features (sphericity, elongation) and 4 LoG-based features which include 1 first-order feature (kurtosis) and 3 texture features (GLDM, NGTDM). Six features were chosen for joint T1-T2 model, including 1 first-order feature (kurtosis) and 5 LoG-based features which consist of 2 first-order features (kurtosis, skewness) and 3 texture features (GLCM, GLDM), as shown in **Table 2**. With these

TABLE 1 | Patient characteristics in the present cohort.

Clinical factor	Category	Number (Percent)	P-values
Gender	Male	50 (71.4%)	0.2558
	Female	20 (28.6%)	
Age in years	<51	21 (30%)	0.386
	51–70	42 (60%)	
	>70	7 (10%)	
T stage	T1	2 (2.9%)	0.554
	T2	2 (2.9%)	
	T3	50 (71.4%)	
	T4	16 (22.8%)	
N stage	N1	5 (7.1%)	0.859
	N2	56 (80%)	
	N3	9 (12.9%)	
Overall stage	Stage II	3 (4.3%)	0.535
	Stage III	43 (61.4%)	
	Stage IV	24 (34.3%)	
Histology	Type I	3 (4.3%)	0.827
	Type II	1 (1.4%)	
	Type III	66 (94.3%)	
Treatment	EBRT-alone	14 (20%)	0.8411
	CCRT	37 (52.9%)	
	CCRT + AC	11 (15.7%)	
	IC + CCRT	7 (10%)	
	Others	1 (1.4%)	
Initial weight (kg) (average ± SD)	Replan Group	61.6 ± 15.5	0.929
	Non-replan Group	61.9 ± 12.2	

EBRT, External Beam Radiation Treatment; CCRT, Concurrent Chemotherapy Radiation Treatment; IC, Induction Chemotherapy; AC, Adjuvant Chemotherapy; Type I, Keratinizing squamous cell carcinoma; Type II, Non-keratinizing differentiated carcinoma; Type III, Non-keratinizing undifferentiated carcinoma.

selected features, we achieved average AUCs of 0.962 (0.852), 0.895 (0.750), 0.904 (0.930) in training (testing) set for CET1-w, Tw-2 and joint T1-T2 models, respectively, representing a promising result for pre-treatment prediction of ART eligibility in NPC patients.

Multiple groups have confirmed that significant tumor shrinkage occurs during RT, triggering the need for ART. Hu et al. (6) reviewed the planning CT and re-CT images of 40 re-planned NPC patients and confirmed the significant clinical-target-volume shrinkage of 35.1%. Murat et al. (11) reported median percentage change in GTV of HNC patients for primary (26.8%), nodal (43.0%), and total (31.2%) GTVs. Lee H et al. confirmed average volume reduction of GTVnp of 45.9 cm^3 (pre-RT) to 26.7 cm^3 (third week of RT) in 159 NPC patients. All these studies have suggested that tumor shrinkage serves as a favorable ART criterion. However, only a few studies have developed ART

FIGURE 4 | Feature selection and model optimization methodology. Superscript "a": T for training cohort; V for validation cohort. "b": The number inside the parentheses is either "1" or "0," representing "re-planned" and "not re-planned" patients; Numbers in front of the parentheses indicate number of patients. "c": 25 features remained in feature set 1c for CET1-w-based model; while 28 and 39 for T2-w-based and Joint T1-T2-based models, respectively. "d": 16 features remained in feature set 2 for CET1-w-beased model; while 13 and 22 for T2-w-based and Joint T1-T2-based models, respectively.

selection strategies based on the tumor volume reduction. Murat et al. (11) developed a decision tree for tumor shrinkage for HNC patients, incorporating initial target volumes and other clinical factors; although an accuracy of 88% was reported in predicting the tumor shrinkage in 48 patients, the validity was not tested and some of the clinical factors used may not be available in other clinics, such as tumor growth pattern (endophytic or exophytic), hindering the generalizability of the decision tree. Recently, Ramella et al. (30) explored the radiomic capability for ART in lung cancer patients and reported that radiomic features extracted from planning target volume (PTV) of lung cancer on CT images were capable of distinguishing patients between ART and non-ART group with AUC of 0.82, on the ground of tumor shrinkage during treatment. To our best knowledge, this study is the first to include radiomics in predicting ART eligibility for NPC patients and its tumoral predictive biomarkers for ART has not been explored before. Our promising results are also in line with the work done by Ramella et al. (30)

In our experience, we observed that the joint T1-T2 radiomic model outperformed either CET1-w or T2-w alone model in terms of AUCs in both training and testing sets. From **Figures 5G–I**, it can be observed that the joint T1-T2 model

gives a more consistent variation in 95% CI of AUCs against different feature sets in both training and testing sets, suggesting that joint T1-T2 model might be the preferable predictive system among the others. Another interesting observation was that the majority (5 of 6) of the selected features in the joint T1-T2 model were from CET1-w images, suggesting that features from CET1-w images might be more predictive than those from T2-w images. A possible reason could be attributed to the inherent limitation of LASSO; when pairwise correlations exist between predictors, the LASSO picks one correlated predictor and ignores the rest. To account for this, we performed another PCC with $r \geq 0.8$ prior to part III in our feature selection methodology (**Figure 4**) to avoid highly correlated features in our final models. Further investigations on the feature selection methodology will be part of our future studies.

On the other hand, NPC radiomics studies on MR images have been widely studied, focusing mainly on prediction of prognosis (disease progression) and treatment response to either induction chemotherapy (IC) or chemo-radiotherapy, while prediction of the need for replanning has not been previously reported. Besides, each study developed a unique radiomic signature for the same outcome prediction, which limits the feasibility

FIGURE 5 | Distribution of AUC values in training and testing cohorts against different number of features in the constructed models from 100 resampled iterations of 20-repeated 3-fold cross validation (CET1-w model: first column, T2-w model: second column, and Joint T1-T2 model: third column). The box and whisker plots in first **(A–C)** and second rows **(D–F)** display the AUC distributions with varying number of selected features in training cohort and testing cohort, respectively; the plots in third row **(G–I)** displays 95% confidence interval and average AUCs for both cohorts against number of selected features in the models.

to directly compare all the resultant features between studies. However, interestingly, categories of resultant features might be different depending on prediction outcomes, which might explain our results to some extent. For prognostic prediction, texture features were obviously dominant in their final radiomic signatures relative to first-order and shape features, and GLCM (Gray-Level Co-occurrence Matrix) was the only shared-feature category between studies. A possible rationale might be that the texture features were considered to reflect intra-tumor heterogeneity by depicting the spatial arrangement of voxels (regularity) and variability of local intensity within tumor, which was acknowledged as a characteristic of malignancy. For prediction of treatment response, while GLCM were still the only common resultant feature category between studies, however, first-order features were dominant in final radiomics signature. Wang et al. investigated the capability of MRI-based radiomic signatures to predict early response to IC for NPC patients using T1-w, CET1-w, and T2-w MR images. Among the 15 features selected in their joint-T1-CET1-T2-w model, 7 were first-order features, three were GLCM features, and the rest were Gabor and wavelet features. Another radiomic study by Hou et al. (31) exploring feasibility of CECT-based biomarkers to predict therapeutic response of esophageal carcinoma to chemo-radiotherapy reported that first-order features (skewness

and/or kurtosis) were identified as significant parameters for differentiating SDs (stable disease) from PRs (partial response) and SDs from CRs (complete response). In both studies, the tumor response was assessed according to the Response Evaluation Criteria in Solid Tumors (RECIST), which takes into account the reduction of tumor size following treatment. Similar to our study, we hypothesized that the image-based tumoral biomarkers are predictive to tumor shrinkage.

In our results, shape features (e.g., Sphericity, Elongation, Maximum 2D diameter slice) and/or first-order features (e.g., kurtosis and skewness) were generally dominant relative to texture features in our models, which is consistent with results from abovementioned radiomic studies for treatment response prediction. Interestingly, kurtosis and/or skewness and GLCM-based features are the common features shared in all three models. Kurtosis and skewness measure the peakiness and asymmetry of the histogram, respectively, while GLCM features quantify the spatial gray-level variation within local neighbors on a pixel basis. Nevertheless, the understanding of the meaningfulness of these features, especially in relation to the prediction outcome, is still largely unknown and deserves further investigations.

This study has several limitations. Firstly, the heterogeneity of image acquisition and reconstruction protocols and ART

TABLE 2 | Table of selected features in CET1-w, T2-w, and joint T1-T2 radiomics models.

MRI series	Selected radiomics features	CET1-w model	T2-w model	Joint T1-T2 model
CET1-w	Original shape Sphericity	✓		
CET1-w	Original shape Maximum 2D Diameter Slice	✓		
CET1-w	Log-sigma-2-0-mm-3D glcm MCC	✓		✓
CET1-w	Log-sigma-2-0-mm-3D first-order Kurtosis	✓		✓
CET1-w	Log-sigma-3-0-mm-3D first-order Skewness	✓		✓
CET1-w	Log-sigma-4-0-mm-3D first-order Kurtosis	✓		
CET1-w	Log-sigma-5-0-mm-3D gldm Dependence Entropy	✓		
CET1-w	Log-sigma-5-0-mm-3D gldm Small Dependence Low Gray Level Emphasis	✓		✓
CET1-w	Original first-order Kurtosis			✓
T2-w	Original shape Sphericity		✓	
T2-w	Original shape Elongation		✓	
T2-w	Log-sigma-2-0-mm-3D gldm Large Dependence High Gray Level Emphasis		✓	
T2-w	Log-sigma-2-0-mm-3D glcm Imc1		✓	
T2-w	Log-sigma-3-0-mm-3D ngtdm Strength		✓	
T2-w	Log-sigma-5-0-mm-3D first-order Kurtosis		✓	
T2-w	Log-sigma-3-0-mm-3D glcm Idn			✓

strategies in different medical centers limit the generalizability of the identified models and reproducibility of the selected features. In future study, we will perform testing on different datasets obtained from other oncology departments with patients undergoing MRIs on different scanners. Secondly, the rate of adaptive replannings in the small cohort is relatively low. A more convincible conclusion could be drawn by recruiting larger cohorts with more balanced dataset between patients who underwent replan and those did not, which will be part of our future efforts. Lastly, the retrospective nature of this study might account for the potential bias. However, the novelty of this study was to highlight the capability of using pre-treatment MRI radiomic features to predict which patients undergoing radiotherapy for NPC were selected for ART. In future study, radiomics features from other ROIs and other pertinent non-radiomic clinical data, such as volumetric and dosimetric data of tumor and nearby organs (e.g., lymph nodes and parotid glands), and geometric relations among these structures, will be incorporated into our radiomics models in future to yield a more comprehensive prediction.

CONCLUSION

The present study successfully demonstrated promising capability of MRI-based radiomics for pre-pretreatment identification of ART eligibility in NPC patients. In particular, the joint T1-T2 model with 6 selected radiomic features appears to be the preferable predictive system over other studied models. This would allow radiation oncologists to more effectively and accurately prescribe ART on individual patient basis to achieve true personalized radiotherapy for NPC patients, meanwhile streamlining resources management in clinical settings. In future work, multi-institution prospective studies with larger patient sample are warranted to improve the clinical efficacy of our models.

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 Human Subjects Ethics Sub-committee of the Hong Kong Polytechnic University and Kowloon Central/Kowloon East Cluster Research Ethics Committee of the Hospital Authority. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

YZ, KA, FL, JC, TY, and SL contributed to study design, methodology development, results interpretation, and manuscript review. CY offered administrative and material support for data collection. SL, TY, MC, KT, NC, YF, CL, KO, LT, KH, FC, WH, and LN collected patients clinical and imaging data. MC, KT, NC, YF, CL, KO, LT, KH, FC, and WH contributed to the image preprocessing and feature extraction. TY constructed the models. SL wrote the manuscript. JC supervised the study.

SUPPLEMENTARY MATERIAL

REFERENCES

1. Ng WT, Wong ECY, Lee VHF, Chan JYW, Lee AWM. Head and neck cancer in Hong Kong. *Jpn J Clin Oncol.* (2018) 48:13–21. doi: 10.1093/jjco/hyx151

2. Cheng HC, Wu VW, Ngan RK, Tang KW, Chan CC, Wong KH, et al. A prospective study on volumetric and dosimetric changes during intensity-modulated radiotherapy for nasopharyngeal carcinoma patients. *Radiother Oncol.* (2012) 104:317–23. doi: 10.1016/j.radonc.2012.03.013

3. Jin X, Han C, Zhou Y, Yi J, Yan H, Xie C. A modified VMAT adaptive radiotherapy for nasopharyngeal cancer patients based on CT-CT image fusion. *Radiat Oncol.* (2013) 8:277. doi: 10.1186/1748-717X-8-277

4. Zhao L, Wan Q, Zhou Y, Deng X, Xie C, Wu S. The role of replanning in fractionated intensity modulated radiotherapy for nasopharyngeal carcinoma. *Radiother Oncol.* (2011) 98:23–7. doi: 10.1016/j.radonc.2010.10.009

5. Lee H, Ahn YC, Oh D, Nam H, Noh JM, Park SY. Tumor volume reduction rate during adaptive radiation therapy as a prognosticator for nasopharyngeal cancer. *Cancer Res Treat.* (2016) 48:537–45. doi: 10.4143/crt.2015.081

6. Hu YC, Tsai KW, Lee CC, Peng NJ, Chien JC, Tseng HH, et al. Which nasopharyngeal cancer patients need adaptive radiotherapy? *BMC Cancer.* (2018) 18:1234. doi: 10.1186/s12885-018-5159-y

7. Tan W, Li Y, Han G, Xu J, Wang X, Li Y, et al. Target volume and position variations during intensity-modulated radiotherapy for patients with nasopharyngeal carcinoma. *Oncotargets Ther.* (2013) 6:1719–28. doi: 10.2147/OTT.S53639

8. Barker JL Jr, Garden AS, Ang KK, O'Daniel JC, Wang H, Court LE, et al. Quantification of volumetric and geometric changes occurring during fractionated radiotherapy for head-and-neck cancer using an integrated CT/linear accelerator system. *Int J Radiat Oncol Biol Phys.* (2004) 59:960–70. doi: 10.1016/j.ijrobp.2003.12.024

9. Munshi A, Pandey MB, Durga T, Pandey KC, Bahadur S, Mohanti BK. Weight loss during radiotherapy for head and neck malignancies: what factors impact it? *Nutr Cancer.* (2003) 47:136–40. doi: 10.1207/s15327914nc4702_5

10. Noble DJ, Yeap PL, Seah SYK, Harrison K, Shelley LEA, Romanchikova M, et al. Anatomical change during radiotherapy for head and neck cancer, and its effect on delivered dose to the spinal cord. *Radiother Oncol.* (2019) 130:32–8. doi: 10.1016/j.radonc.2018.07.009

11. Surucu M, Shah KK, Mescioglu I, Roeske JC, Small W Jr, Choi M, et al. Decision trees predicting tumor shrinkage for head and neck cancer: Implications for adaptive radiotherapy. *Technol Cancer Res Treat.* (2016) 15:139–45. doi: 10.1177/1533034615572638

12. Hansen EK, Bucci MK, Quivey JM, Weinberg V, Xia P. Repeat CT imaging and replanning during the course of IMRT for head-and-neck cancer. *Int J Radiat Oncol Biol Phys.* (2006) 64:355–62. doi: 10.1016/j.ijrobp.2005.07.957

13. Bhide SA, Davies M, Burke K, McNair HA, Hansen V, Barbachano Y, et al. Weekly volume and dosimetric changes during chemoradiotherapy with intensity-modulated radiation therapy for head and neck cancer: a prospective observational study. *Int J Radiat Oncol Biol Phys.* (2010) 76:1360–8. doi: 10.1016/j.ijrobp.2009.04.005

14. Schwartz DL, Garden AS, Shah SJ, Chronowski G, Sejpal S, Rosenthal DI, et al. Adaptive radiotherapy for head-and-neck cancer - dosimetric results from a prospective clinical trial. *Radiother Oncol.* (2013) 106:80–84. doi: 10.1016/j.radonc.2012.10.010

15. Schwartz DL, Garden AS, Thomas J, Chen Y, Zhang Y, Lewin J, et al. Adaptive radiotherapy for head-and-neck cancer: initial clinical outcomes from a prospective trial. *Int J Radiat Oncol Biol Phys.* (2012) 83:986–93. doi: 10.1016/j.ijrobp.2011.08.017

16. Chen AM, Daly ME, Cui J, Mathai M, Benedict S, Purdy JA. Clinical outcomes among patients with head and neck cancer treated by intensity-modulated radiotherapy with and without adaptive replanning. *Head Neck.* (2014) 36:1541–6. doi: 10.1002/hed.23477

17. Yang H, Hu W, Wang W, Chen P, Ding W, Luo W. Replanning during intensity modulated radiation therapy improved quality of life in patients with nasopharyngeal carcinoma. *Int J Radiat Oncol Biol Phys.* (2013) 85:e47–54. doi: 10.1016/j.ijrobp.2012.09.033

18. Wu Q, Chi Y, Chen PY, Krauss DJ, Yan D, Martinez A. Adaptive replanning strategies accounting for shrinkage in head and neck IMRT. *Int J Radiat Oncol Biol Phys.* (2009) 75:924–32. doi: 10.1016/j.ijrobp.2009.04.047

19. Fung WW, Wu VW, Teo PM. Developing an adaptive radiation therapy strategy for nasopharyngeal carcinoma. *J Radiat Res.* (2014) 55:293–304. doi: 10.1093/jrr/rrt103

20. Brown E, Owen R, Harden F, Mengersen K, Oestreich K, Houghton W, et al. Adaptive radiotherapy: Predicting the need for adaptive radiotherapy in head and neck cancer. *Radiother Oncol.* (2015) 116:57–63. doi: 10.1016/j.radonc.2015.06.025

21. Brouwer CL, Steenbakkers RJ, Langendijk JA, Sijtsema NM. Identifying patients who may benefit from adaptive radiotherapy: Does the literature on anatomic and dosimetric changes in head and neck organs at risk during radiotherapy provide information to help? *Radiother Oncol.* (2015) 115:285–94. doi: 10.1016/j.radonc.2015.05.018

22. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer.* (2012) 48:441–6. doi: 10.1016/j.ejca.2011.11.036

23. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol.* (2017) 14:749–62. doi: 10.1038/nrclinonc.2017.141

24. Zhang B, Tian J, Dong D, Gu D, Dong Y, Zhang L, et al. Radiomics features of multiparametric MRI as novel prognostic factors in advanced nasopharyngeal carcinoma. *Clin Cancer Res.* (2017) 23:4259–69. doi: 10.1158/1078-0432.CCR-16-2910

25. Zhang B, He X, Ouyang F, Gu D, Dong Y, Zhang L, et al. Radiomic machine-learning classifiers for prognostic biomarkers of advanced nasopharyngeal carcinoma. *Cancer Lett.* (2017) 403:21–7. doi: 10.1016/j.canlet.2017.06.004

26. Zhang B, Ouyang F, Gu D, Dong Y, Zhang L, Mo X, et al. Advanced nasopharyngeal carcinoma: pre-treatment prediction of progression based on multi-parametric MRI radiomics. *Oncotarget.* (2017) 8:72457–65. doi: 10.18632/oncotarget.19799

27. Ouyang FS, Guo BL, Zhang B, Dong YH, Zhang L, Mo XK, et al. Exploration and validation of radiomics signature as an independent prognostic biomarker in stage III-IVb nasopharyngeal carcinoma. *Oncotarget.* (2017) 8:74869–79. doi: 10.18632/oncotarget.20423

28. Dong D, Tang L, Li ZY, Fang MJ, Gao JB, Shan XH, et al. Development and validation of an individualized nomogram to identify occult peritoneal metastasis in patients with advanced gastric cancer. *Ann Oncol.* (2019) 30:431–8. doi: 10.1093/annonc/mdz001

29. Xu CJ, van der Schaaf A, Van't Veld AA, Langendijk JA, Schilstra C.Xu CJ, van dSA, van't VA, Langendijk JA, Schilstra C. Statistical validation of normal tissue complication probability models. *Int J Radiat Oncol Biol Phys.* (2012) 84:e123–e129. doi: 10.1016/j.ijrobp.2012.02.022

30. Ramella S, Fiore M, Greco C, Cordelli E, Sicilia R, Merone M, et al. A radiomic approach for adaptive radiotherapy in non-small cell lung cancer patients. *PLoS ONE.* 13:e0207455. doi: 10.1371/journal.pone.0207455

31. Hou Z, Ren W, Li S, Liu J, Sun Y, Yan J, et al. Radiomic analysis in contrast-enhanced CT: predict treatment response to chemoradiotherapy in esophageal carcinoma. *Oncotarget.* (2017) 8:104444–54. doi: 10.18632/oncotarget.22304

Prognostic Value of Texture Analysis Based on Pretreatment DWI-Weighted MRI for Esophageal Squamous Cell Carcinoma Patients Treated With Concurrent Chemo-Radiotherapy

Zhenjiang Li[1]*†, Chun Han[2]†, Lan Wang[2], Jian Zhu[1], Yong Yin[1] and Baosheng Li[1]*

[1] *Shandong Cancer Hospital and Institute, Shandong First Medical University and Shandong Academy of Medical Sciences, Shandong Cancer Hospital, Jinan, China,* [2] *Fourth Hospital of Hebei Medical University, Shijiazhuang, China*

Purpose: The purpose of the research was to assess the prognostic value of three-dimensional (3D) texture features based on diffusion-weighted magnetic resonance imaging (DWI) for esophageal squamous cell carcinoma (ESCC) patients undergoing concurrent chemo-radiotherapy (CRT).

Methods: We prospectively enrolled 82 patients with ESCC into a cohort study. Two DWI sequences ($b = 0$ and $b = 600$ s/mm^2) were acquired along with axial T2WI and T1WI before CRT. Two groups of features were examined: (1) clinical and demographic features (e.g., TNM stage, age and sex) and (2) changes in spatial texture characteristics of the apparent diffusion coefficient (ADC), which characterizes gray intensity changes in tumor areas, spatial pattern and distribution, and related changes caused by CRT. Reproducible feature sets without redundancy were statistically filtered and validated. The prognostic values associated with overall survival (OS) for each parameter were studied using Kaplan-Meier and Cox regression models for univariate and multivariate analyses, respectively.

Results: Both univariate and multivariate Cox model analyses showed that the energy of intensity histogram texture (IHIST_energy), radiation dose, mean of the contrast in distance 1 of 26 directions (m_contrast_1), extreme difference of the homogeneity in distance 2 of 26 directions (Diff_homogeneity_2), mean of the inverse variance in distance 2 of 26 directions (m_Inversevariance_2), high-intensity small zone emphasis (HISE), and low-intensity large zone emphasis (LILE) were significantly associated with survival. The results showed that 6 texture parameters extracted from the ADC images before treatment could distinguish among high-, medium-, and low-risk groups (log-rank $\chi^2 = 9.7$; $P = 0.00773$). The biased C-index value was 0.715 (95% CI: 0.708 to 0.732) based on bootstrapping validation.

Conclusions: The ADC 3D texture feature can be used as a useful biomarker to predict the survival of ESCC patients undergoing CRT. Combining ADC 3D texture features with conventional prognostic factors can generate reliable survival prediction models.

Keywords: esophageal squamous cell cancer, texture analysis, magnetic resonance imaging, diffusion-weighted magnetic resonance imaging, chemo-radiotherapy

INTRODUCTION

Esophageal squamous cell carcinoma (ESCC) is a disease with increasing incidence, and the diagnosis still carries a poor prognosis despite advances in therapy (1). Currently, chemo-radiotherapy (CRT) is the standard treatment for locally advanced unresectable ESCC. Due to tumor heterogeneity, these patients usually do not have the same response to a specific therapy. Thus, many patients may receive therapy that provides no benefit to them. Recently, a major research focus has been on how to provide individualized therapy. Individualized therapy requires the development of biomarkers to predict treatment prognosis and outcome. Imaging biomarkers, particularly those based on functional imaging techniques that can characterize biological effects at the cellular level, offer great potential to improve individualized therapy (2).

Diffusion-weighted magnetic resonance imaging (DWI) is a powerful MR functional imaging sequence sensitive to water diffusion (3) that can detect morphological changes in tumors at the molecular or cellular level. DWI has been studied for its potential to evaluate the treatment response to CRT for several types of cancers, including rectal cancer (4, 5). The quantitative apparent diffusion coefficient (ADC) map is obtained from two different b values to remove the T2 "shine-through" effects. This can allow quantitative assessment for a treatment response. The ADC can also be used to characterize hypercellularity, distinguish cystic lesion and solid regions, and monitor the change in cellularity within the tumor over time (6). A recent study found that the ADC map can be used to qualitatively assess the tumor area and detect metastatic lymph nodes (LNs) in esophageal cancer (EC) (7).

Recently, the application of texture analysis (TA) in tumor diagnostics has caught the attention of clinical researchers. Texture is an important feature of images that has been used in qualitative and quantitative classification and analysis of materials in industry and medicine. Medical applications of TA provide a quantitative means to analyze and characterize the properties of tumor tissues and their physiological and pathological stages (8). Previous studies have reported that texture analysis can predict the treatment response and predict patient survival (9–11). It was reported that 3D texture analysis can be more useful than 2D analysis in characterizing intra-tumor heterogeneity (12). In the study of ESCC, because the esophagus is a tubular organ, 3D TA is expected to provide richer spatial heterogeneity information than slice samples.

The objective of this study was to prospectively investigate the prognostic value of 3D DWI features in ESCC patients treated with CRT. By studying different types of global and regional 3D features, we evaluate their potential prognostic value in correlation with patient survival.

MATERIALS AND METHODS

Clinical Characteristics of the Patients

Eighty-two patients with newly diagnosed ESCC treated with CRT between 2010 and 2014 were initially enrolled in this prospective study. The inclusion criteria for the study included the following: (1) a confirmed diagnosis of ESCC with tissue pathology; (2) TNM staging according to AJCC 6th Edition, 2002; (3) a Karnofsky performance status (KPS) score >70; (4) no distant metastases under routine medical care; (5) informed consent to have DWI examinations before and during the course of CRT. Ten patients were excluded from the study because of a contraindication to MRI examination, such as those with pacemakers, metal objects, or a claustrophobic disorder. The clinical and treatment characteristics of the qualified 72 patients are summarized in **Table 1**. The mean age at the time of diagnosis was 62.8 ± 9.1 years (median, 62.5 years; range, 45–84 years). The male patients comprised 69.4%. Fifty-nine patients had T3 or T4 primary lesions, 53 patients were determined to have N1 (61%) with lymph node metastases, and 11 and 8 patients were at the N0 and N2 stages, respectively. No patient had distant metastases (**Table 1**).

MRI Acquisition

MR imaging in expiration breath-hold was performed before starting the treatment for tumor staging. The following imaging protocols were used:

- 2D T1-weighted fast low-angle shot (FLASH) sequence (TR/TE, 140/2.5 *ms*; flip angle, 70°; slice number, 24; gap: 5 mm; matrix, 512×384; field of view (FOV), 380×285 mm);
- Axial TSE T2-weighted MRI (T2WI; slice thickness, 4 mm; slice number, 24; gap: 5 mm; TR/TE, 1580/72 *ms*; flip angle, 140°; matrix: 512×512; FOV, 400×400 mm);
- DWI obtained in the axial planes using a single-shot spin-echo echo planar (SE-EPI) technique (TR/TE, 6800/70 ms; slice thickness, 4 mm; zero gap; matrix, 128×88; FOV, 430×295 mm); the b values (diffusion-sensitive factor) of 0 and 600 s/mm^2 were selected according to a previous pathological study at our institution (13), showing that the tumor lengths measured using a DWI scan of b = 600 s/mm^2 were close to the real tumor lengths based on a surgical specimen with a high concordance with pathology. As shown in **Figure 1**, the DWI scan with b = 600 s/mm^2 has a good image quality relative to the other DWI scans with b = 800 s/mm^2 or 1,000 s/mm^2.

TABLE 1 | Clinical and treatment characteristic.

Prognostic factors	N	%
No. of patients	72	
AGE		
>=62	38	52.8
<62	34	47.2
SEX		
Male	50	69.4
Female	22	30.6
LESION LENGTH		
<=5.5 cm		
>5.5 cm		
PRIMARY SITE		
Cervical	2	2.8
Upper esophagus	24	33.3
Middle esophagus	35	48.6
Lower esophagus	11	15.3
TNM STAGE		
T2	13	18.1
T3	20	27.8
T4	39	54.2
N0	11	15.3
N1	53	73.6
N2	8	11.1
M0	72	100
Treatment characteristics		
RADIATION DOSE		
50 Gy at 2 Gy/fx	1	1.4
54 Gy at 1.8 Gy/fx	5	6.9
54 Gy at 2 Gy/fx	2	2.8
56 Gy at 2 Gy/fx	1	1.4
57.6 Gy at 1.8 Gy/fx	1	1.4
59.4 Gy at 1.8 Gy/fx	4	5.6
60 Gy at 2 Gy/fx	47	65.3
61.2 Gy at 1.8 Gy/fx	5	6.9
63 Gy at 2 Gy/fx	6	8.3
RADIATION TYPE		
IMRT	58	80.6
3DCRT	14	19.4

3DRT, 3 dimensional conformal radiation therapy; IMRT, intensity modulated radiation therapy; fx, fraction.

Motion artifacts were minimized by acquiring all images with breath hold in the expiration phase. Because the tumor volumes could change slightly at the breath hold, two DWI scans with b = 0 and b = 600 were acquired in one cycle, improving the estimation of signal decay. This was expected to provide sufficient imaging information to describe tumor heterogeneity.

Image Preprocessing

An image preprocessing procedure was performed that included tumor segmentation and intensity normalization. The MRI data set in the DICOM format was imported into MATLAB (The Math Works Inc., Natick, MA). An in-house-developed radiomics image analysis program implemented in MATLAB was used for TA (available for share, https://pan.baidu.com/s/1Tl_PsXrQj-OBJt-1cNjaZQ). The method used in this study is described in the **Data Supplement**. To perform reliable measurements, as suggested by Collewet et al. [14], the MRI data were kept in the raw data form, and voxels within the tumor region with intensities outside the range $\mu \pm 3\delta$ were excluded in subsequent texture computations. Voxel intensity values were typically resampled in four discrete values (16, 32, 64, or 128):

$$p(x) = \left[Range \times \frac{I(x) - \min_{i \in \Theta} i}{\max_{i \in \Theta} i - \min_{i \in \Theta} i + 1} \right] \quad (1)$$

where "Range" is the discrete values chosen (16, 32, 64, or 128), $I(x)$ is the intensity of the original image, and Θ is the set of pixels in the delineated area. The use of different resampling schemes was tested. As discussed in the **Data Supplement**, 32 discrete values for renormalization produced the most reliable results.

Tumor Delineation Using MRI Data

The tumor was delineated based on abnormal regions from T2-weighted imaging (T2WI), DWI and ADC maps. Axial ADC maps were generated using an Extended Siemens MR Workspace workstation. The lesions showed relatively higher signals on DWI maps and lower signals on ADC maps. Pre-CRT MRI was first evaluated using the combination of corresponding T2WI and ADC images and matching between DWI and ADC to correctly position the regions of interest (ROIs) in the primary tumor. Axial T2 images in the same plane were referenced by the observers due to a lack of anatomic details because of the low signal-to-noise ratio (SNR) on DWI or ADC. Therefore, the registration accuracy between ADC/DWI and T1WI/T2WI was important for the process, which relies on careful registration by the alignment of local bone structures between ADC and T2WI images. The ROI was drawn along the border of the low signal of the tumor on the b = 600 mm/s^2 ADC images to cover the entire tumor area of each selected slice, avoiding regions with distortions or artifacts by verifying the lesion boundaries on T2WI. Delineation of the lesions was performed independently by two observers. Manual delineations were performed using MIM software (MIMvista Corp, Cleveland, OH). Independent samples t-test or the Kruskal-Wallis H test, where appropriate, was used to assess the differences between the features generated by reader 1 and those by reader 2, as well as between the twice-generated features by reader 1. Inter- and intra-class correlation coefficients (ICCs) were used to evaluate the intra- and inter-observer agreements of the contour agreement and feature extraction. ICC values >0.80 indicated good agreement. A free-hand ROI was drawn along the border of the low signal of the tumor on the b = 600 images to cover the entire tumor area of each selected slice, by referencing T2WI to verify the lesion boundaries and ensure inclusion of the entire tumor area (**Figure 2**).
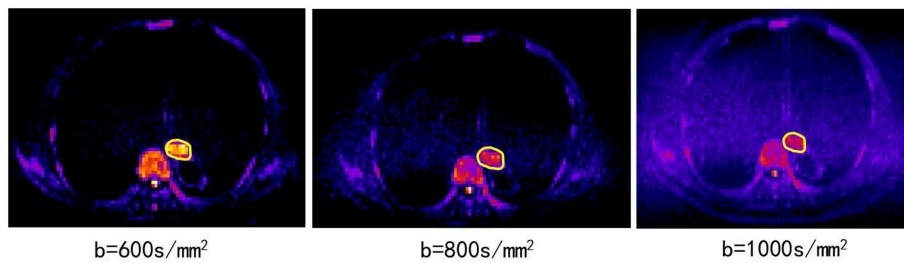
FIGURE 1 | The DWI scans of b = 600, 800, 1,000 s/mm^2 from the same patient (82 years old, male, T3N0M0).
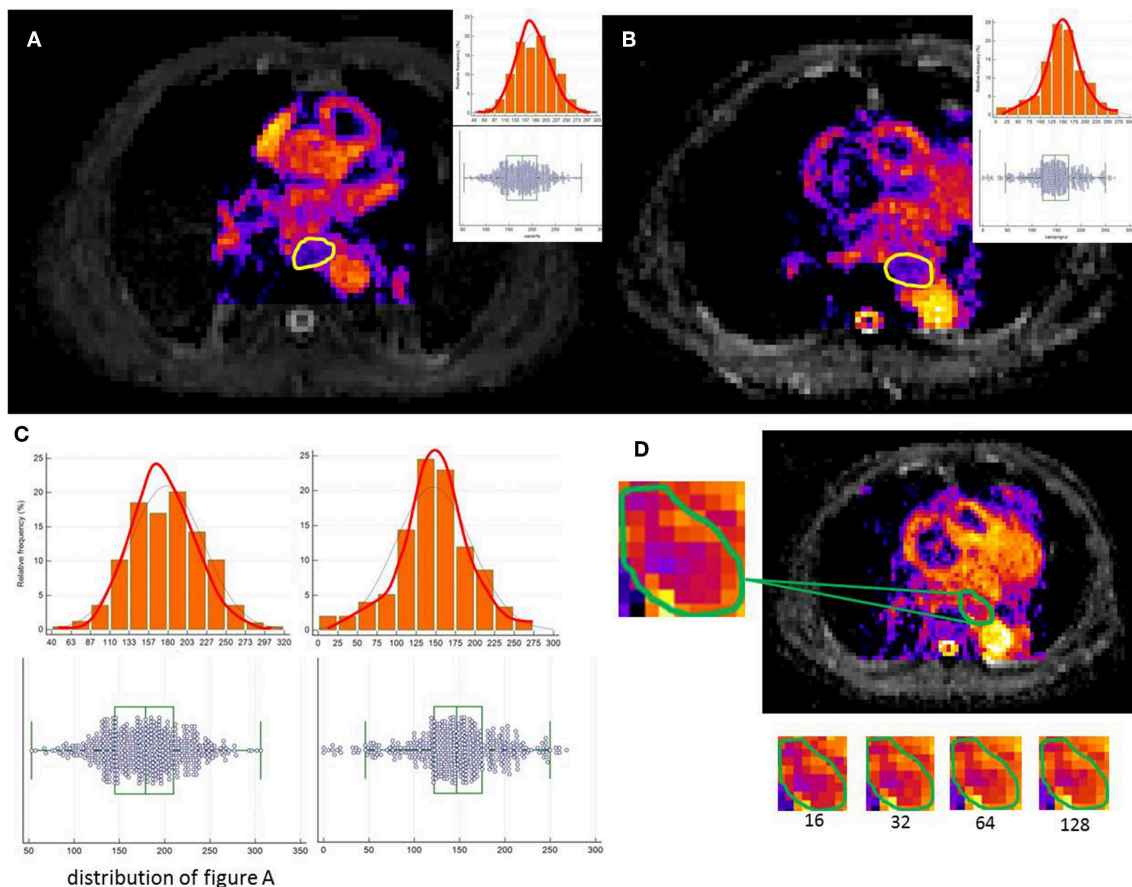
FIGURE 2 | A free-hand ROI was drawn along the border of the low signal of the tumor on the b = 600 images to cover the entire tumor area of each selected slice. **(A,B)** show image heterogeneity by the histogram **(C)**. **(D)** illustrates a tumor slice of the resulting resampled ROI for each of these discretization ranges.

Texture Analysis

From ADC images, four subset features were extracted (**Table 2**) to characterize tumor heterogeneity at global and regional levels using first-order and higher order statistics. These parameters were used to predict the patient response to CRT and survival. Global information was described by intensity histogram parameters (IHIST), including variance, mean, energy, entropy, skewness, and kurtosis, while regional information was characterized by intensity size-zone variability features (ISZFs). Regional heterogeneity information included intensity variability

in the size and tumor zones [see Tixier et al. (15) for the mathematical definition of the regional heterogeneity formula used in this study]. Local heterogeneity information was derived using the co-occurrence of the gray-level co-occurrence matrix (GLCM) and gray-level gradient co-occurrence (GLGCM). Twenty-six gray-level co-occurrence matrices were computed in the direction of the 26 uniform distributions on the sphere from each voxel data area. GLGCM was acquired in the original ADC image, and the corresponding gradient image of the ADC image and 15 Haralick features (16) were extracted from GLGCM. To

TABLE 2 | Extracted texture features.

IHIST	GLCM*	GLGCM*	ISZFs
Mean	Mean/diff energy	Small gradient emphasis	Small zone emphasis
Variance	Mean/diff entropy	Large gradient emphasis	Large zone emphasis
Median	Mean/diff correlation	No homogeneity of gray	Gray intensity change
Maximum	Mean/diff contrast	No homogeneity of gradient	Zone size change
Minimum	Mean/diff homogeneity	Energy	Zone percentage
Up quarter value	Mean/diff variance	Mean of gray	High intensity emphasis
Down quarter value	Mean/diff mean	Mean of gradient	Low intensity small zone emphasis
Energy	Mean/diff inertia	Variance of gray	High intensity small zone emphasis
Entropy	Mean/diff cluster shade	Variance of gradient	Low intensity large zone emphasis
Skewness	Mean/diff cluster tendency	Correlation of gradient	High intensity large zone emphasis
Kurtosis	Mean/diff max probability	Entropy of gray	
	Mean/diff inverse variance	Entropy of gradient	
	Mean/diff inverse difference moment	Mix entropy	
	Mean/diff sum mean	Inertia	
	Mean/diff sum entropy	Inverse difference moment	
	Mean/diff difference entropy		

IHIST, Intensity histogram texture; GLCM, gray level co-occurrence matrix; GLGCM, gray level gradient co-occurrence; ISZFs, Intensity size-zone variability features; diff, the extreme difference of feature.
*128 GLCM features are constructed by 64 mean values and 64 extreme difference values. Similarly, 60 gray gradient features were extracted from GLGCM.

obtain isotropy properties, the mean value and difference in the maximum and minimum value from the same Haralick features were computed in 26 directions and four distances (1, 2, 4, and 8 voxel distance). One hundred twenty-eight GLCM features were constructed by 64 mean values and 64 extreme difference values. Similarly, 60 gray gradient features were extracted from GLGCM. One hundred twenty-eight local heterogeneity features of co-occurrence matrices characterized variations in the intensity between consecutive voxels. For texture reporting with GLCM, the notation convention "method"_"feature"_"number" was used; for example, the mean of the contrast in distance 2 of 26 directions would be identified by m_contrast_2 and the extreme difference of cluster shade in distance 1 of 26 directions would be identified by Diff_clustershade_1. The GLGCM features were identified by GLGCM_"feature"_"distance"; the small gradient emphasis in distance 4 would be identified by GLGCM_Small gradient emphasis_4. The histogram-related features would be abbreviated by IHIST_"feature"; for instance, the histogram energy feature would be identified by IHIST_energy. The detailed feature information is shown in **Table 2**. The algorithms for texture feature extraction are described in the **Data Supplement**.

Feature Selection Methods

In this study, 229 features in four categories were selected. Notably, not all the features required evaluation because many features would be irrelevant or redundant. Therefore, the number of features tested must be reduced by feature extraction. The three major reasons to perform feature reduction are as follows: (1) to reduce the training time; (2) to improve the robustness; and (3) to enhance the reliability.

To assess texture feature reproducibility, Fried DV's method was used to perform test-retest scans from 10 independent patients (17). The results are shown in the **Data Supplement**. Reproducibility of the characteristic parameters is an important characteristic in repeated experiments. In this study, a concordance correlation coefficient (CCC) value >0.9 was considered to guarantee reproducibility. Another consideration was the use of a defined "dynamic range" (DR) metric to select highly differentiated features. Similar to CCC, DR ≥ 0.9 indicates that this feature has a large dynamic range (18). The R^2 of simple regression was equal to the square of the Pearson correlation coefficient. Values close to 1 indicate that the data points were close to the fitted line. These features were grouped by R^2 between them. We recursively repeated the process to cover all features. We also calculated R^2 between the remaining features to quantify the dependencies. Using the above methods, 38 features were chosen in the penalized model with highly reproducibility and a dynamic range.

To avoid an inadequate sample size to train and test, the "leave one out" cross validation method was used to test the model stability. Using many features, it was difficult to predict which parameters would be useful to indicate patient treatment responses and survival. Therefore, it was necessary to reduce the number of features to improve the predictability and reliability for analysis. The least absolute shrinkage and selection operator (LASSO) method was used to select the most useful predictive features from the primary data set.

The abovementioned features and clinically relevant features were entered into a penalized multivariate Cox proportional hazards model (Adaptive Elastic Net Cox model) that simultaneously performs covariate selection in addition to model development. The Adaptive Elastic Net method for the Cox model has a grouping effect (19, 20). By minimizing the opposite number of the Cox model first, and then adding the

appropriate penalty, the Elastic Net estimator for the Cox model was obtained:

$$\hat{\beta}(EN) = \arg\min \left\{ \frac{1}{n} \sum_{i=1}^{n} \left\{ -\beta^T X_i + \ln \left[\sum_{j \in R_i} \exp\left(\beta^T X_j\right) \right] \right\} + \lambda_1 \|\beta\|_1 + \lambda_2 \|\beta\|^2 \right\} \quad (2)$$

Statistical Analysis

Statistical analysis was performed using SPSS19.0 (IBM, Armonk, New York, United States) for Windows and R software (version 3.2.3; http://www.Rproject.org). The R packages (hdnom v 4.1, survival v 2.39-5, penalized v 0.9-47 and survcomp v 1.20.0) were used. OS was calculated from the date of the initial diagnosis to the date of death or time for the most recent follow-up, if the patients were still alive. The reported statistical significance levels were all two-sided, with statistical significance set at 0.05.

RESULTS

Overall Therapeutic Response and Survival

After the completion of CRT, an overall therapeutic response (TE) was estimated according to the RECIST 1.1 standard (21). Thirty-six patients (50%) were determined to have a complete response (CR), and 36 (50%) patients had a partial response (PR). The overall effective response rate was 100.0%.

All patients were followed up for over 1 year, and 27 patients (37.5%) were followed over 2 years. The median follow-up time was 16.5 months. The 1 and 2 year OS rates for all patients were estimated at 72.2 and 34.7%, respectively. According to the overall treatment response (CR, PR), the 1 and 2 year survival rates of CR patients were 86.1 and 38.9%, respectively, and those of PR patients were 58.3 and 30.1%, respectively. Significant differences were found between the two groups (log-rank test; $\chi^2 = 4.153$, $P = 0.042$).

Prognostic Value of ADC Radiomics Data

The possible association of ADC map features with survival was explored by Kaplan-Meier survival analysis. No significant correlation was found between any ADC value measurement (ADC_{mean}, ADC_{up}, ADC_{down}, ADC_{min}, ADC_{max}) in ESCC patients undergoing CRT ($P = 0.224, 0.534, 0.549, 0.328, 0.369$). The results of the log-rank analysis of conventional prognostic factors for OS in univariate analysis are given in **Table 3**.

Age, sex, tumor site, TNM stage, and treatment type were not significant prognostic factors according to the results of univariate analysis. In univariate analysis, the GTV (Gross Tumor Volume size), pathology lesion length, therapeutic effect and radiation dose were significant prognostic factors. Univariate analysis of image texture showed that the IHIST_energy, m_contrast_1, m_Cluster shade_2, Diff_Clusetr Tendency_2, Diff_homogeneity_2, m_lnversevariance_2, Small gradient emphasis_1, GLGCM_small gradient emphasis, high-intensity small zone emphasis (HISE) and low-intensity large

TABLE 3 | Conventional prognostic factors for patients.

Variables	OS,%		P-value	Hazard ratio	95% CI	
	1 year	2 year			Lower	Upper
GENDER						
Male	74	34		1		
Female	68.2	36.4	0.325	0.710	0.360	1.403
AGE						
>=62	66.7	25		1		
<62	77.8	44.4	0.168	1.002	0.991	1.055
TUMOR SITE						
Cervial	50	50		1		
Upper esophagus	79.2	33.3	0.884	0.851	0.098	7.356
Middle esophagus	71.4	34.3	0.626	1.284	0.469	3.513
Lower esophagus	63.6	36.4	0.499	1.394	0.531	3.658
PATHOLOGY LESION LENGTH						
<=5 (40)	77.5	35		1		
>5 (32)	65.6	34.4	**0.005**	1.149	1.042	1.268
T STAGE						
T2	0.85	0.38		1		
T3	0.75	0.3	0.9042	1.0763	0.4629	2.5024
T4	0.66	0.36	0.2153	1.6345	0.7702	3.4688
N STAGE						
N0	70	40		1		
N1	73.6	34.0	0.835	1.135	0.345	3.728
N2	77.8	44.4	0.741	1.172	0.458	2.996
TNM stage I–II	78.1	34.4		1		
III	67.5	35	0.120	1.274	0.939	1.727
GTV						
<=40.35	83.3	36.1		1		
>40.35	61.1	33.3	**0.015**	0.476	0.261	0.868
TE						
CR	86.1	38.9		1		
PR	58.3	30.6	**0.042**	1.851	1.024	3.346
RADIATION DOSE						
>=60	81.5	40.7		1		
<60	44.4	16.7	**0.050**	0.916	0.807	1.015
RADIATION TYPE						
IMRT	69.0	32.8		1		
3DCRT	85.7	42.9	0.906	1.043	0.516	2.108

3DRT, 3 dimensional conformal radiation therapy; IMRT, intensity modulated radiation therapy; GTV, Gross Tumor Volume; TE, therapeutic effect. The bold values show that the P ≤ 0.05.

zone emphasis (LILE) demonstrated a statistically significant difference in association with the OS rates.

Feature Selection

Thirty-eight texture features were reduced to 6 nonzero coefficients in the LASSO model with potential predictors based on 72 patients in the primary cohort. The detailed results used in this study were reported in the **Data Supplement**.

To further define the predictive values of ADC, multivariate Cox regression model analysis was performed using adjusted clinical factors. **Table 4** lists the multivariate analysis results.

TABLE 4 | Multivariate analysis of prognostic factor for patients with ESCC.

Variables	B	SE	Wald	df	P-value	Hazard ratio	95% CI	
							Lower	Upper
Radiation dose	−0.125	0.067	5.112	1	0.026	1.211	0.925	1.326
IHIST_energy	−0.056	0.021	7.482	1	0.007	0.952	0.911	0.995
m_contrast_1	0.146	0.028	19.47	1	0.001	1.152	1.128	1.195
Diff_homogeneity_2	−0.022	0.002	8.824	1	0.003	0.963	0.941	0.981
m_Inverencevariance_2	0.036	0.014	4.06	1	0.034	1.042	1.002	1.13
HISE	−0.053	0.018	8.016	1	0.004	0.942	0.913	0.952
LILE	0.067	0.033	9.735	1	0.003	1.085	1.033	1.139

IHIST_energy, the energy of intensity histogram texture; m_contrast_1, the mean of contrast in distance 1 of 26 directions; Diff_homogeneity_2, the extreme difference of homogeneity in distance 2 of 26 directions; m_Inverencevariance_2, the mean of inverse variance in distance 2 of 26 directions; HISE, high intensity small zone emphasis; LILE, low intensity large zone emphasis.

Validation of Model Performance

The study used the "hdnom" package to assess the model performance by time-dependent AUC using the "leave one out" cross-validation method. We validated the Adaptive Elastic Net multivariate Cox model performance every 6 months. **Figure 3** shows the mean, median, maximum, minimum, and 25 and 75% quartiles of time-dependent AUC at each time point across all fold predictions. The median and mean values could be considered the bias-corrected estimation of the model performance. The "leave one out" validation could ensure robust results. The figure shows that the median and mean values at each evaluation time point were relatively close. The results showed that the model had a relative high AUC value at each time point. The study used resampling methods of "leave one out" cross validation for internal model calibration. We split the samples into three risk groups according to the adaptive Elastic Net multivariate Cox model. The model calibration results (median of the predicted survival probability; median of the observed survival probability by the Kaplan-Meier method with 95% CI) are shown in **Figure 4**. The C-index for the prediction model was 0.720 (95% CI: 0.713 to 0.731) for the primary cohort, which was confirmed to be 0.715 (95% CI: 0.708 to 0.732) via bootstrapping validation. We used the Kaplan-Meier survival curve and values in the risk table to further analyze the survival differences among different risk groups. Here, we plotted the Kaplan-Meier survival curve and assessed the amount of risk in three risk groups from 1 to 3 years (**Figure 5**; $\chi^2 = 9.7$, Log-rank $P = 0.00773$).

DISCUSSION

DWI is a powerful MR sequence that provides unique information related to tumor cellularity and the integrity of the cellular membrane. The technique can be applied widely to detect and characterize tumors and to monitor the treatment response (6). The ADC map can be acquired by two DWIs (e.g., b values of 0 and 600 mm/s^2) using an MR workstation. The ADC map is independent of the magnetic field strength and can overcome the effects of T2 shine-through, thus allowing more meaningful comparison of the results. We also performed experiments using 800 mm/s^2 and 1,000 mm/s^2 (**Figure 1**). However, the results
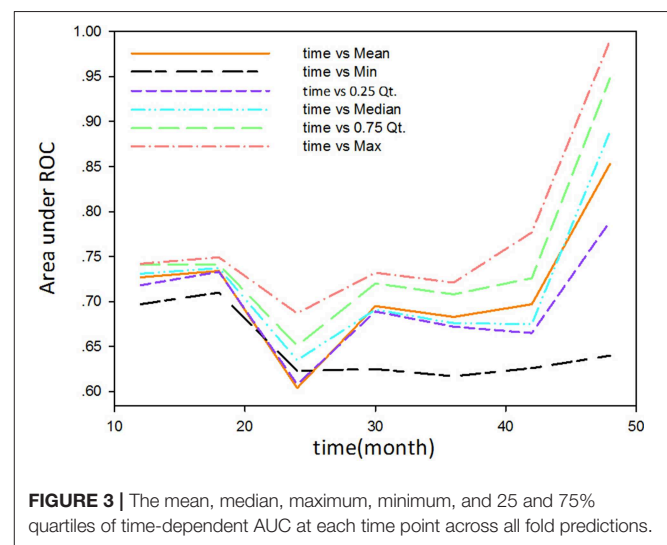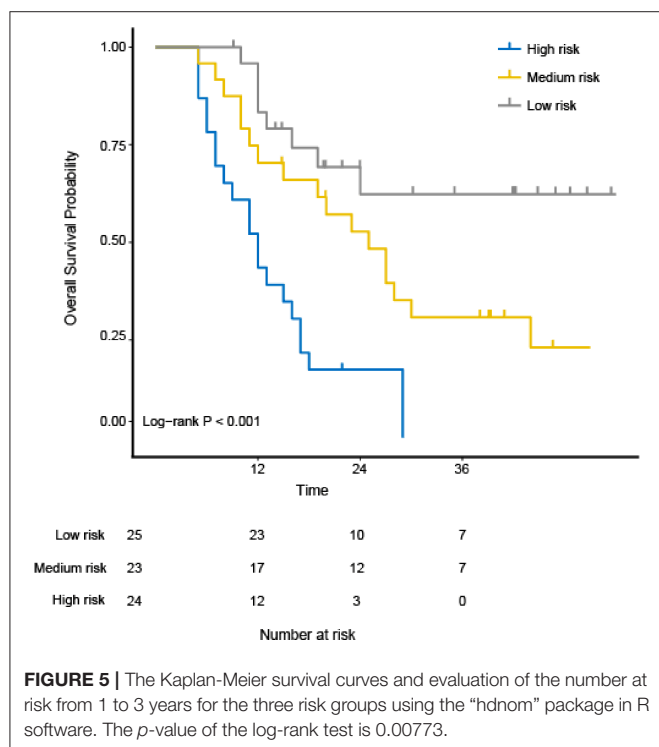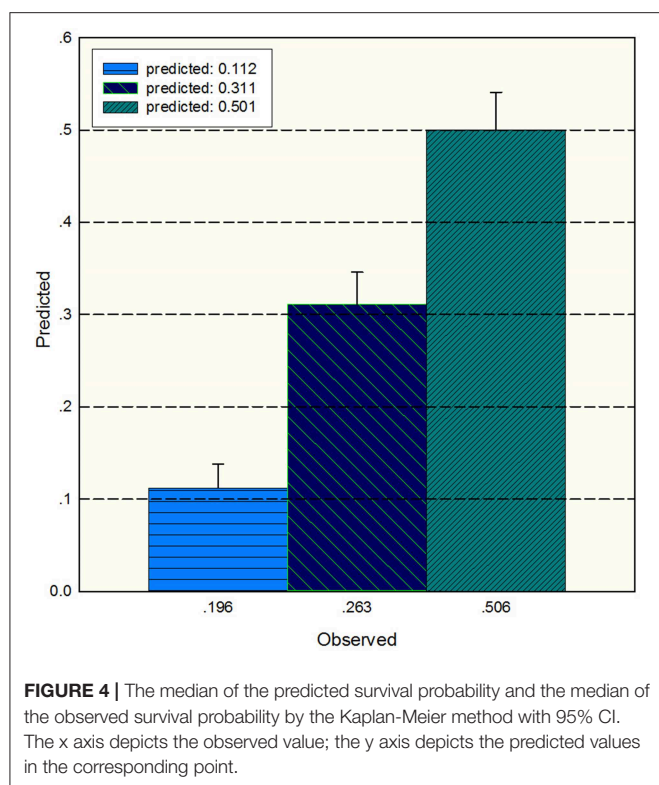

FIGURE 3 | The mean, median, maximum, minimum, and 25 and 75% quartiles of time-dependent AUC at each time point across all fold predictions.

were not reliable, the stability of the parameters was not high, and the repeatability was not good. The possible reasons for the above situation may be that a higher b value will introduce much more noise in chest tumors.

Recent investigations have demonstrated that the pretreatment ADC value may be applied as a biomarker to predict and detect early the treatment response in ESCC, but the results remain controversial. Koyama et al. (22) reported that tumors with a lower pretreatment ADC value and a higher signal intensity at DWI responded better to treatment. Koh et al. (6) discussed the mechanism of this phenomenon and showed that tumors with a high pretreatment ADC value were likely to be more necrotic than those with a low ADC value. Necrotic tumor tissues are frequently hypoxic, acidotic and poorly perfused, leading to diminished sensitivity to CRT. However, not all studies support this hypothesis. Aoyagi et al. studied 80 patients with advanced EC and found that tumors with a higher pretreatment ADC value responded better to treatment (23). They also performed a further study and found that the pretreatment ADC value was not significantly different between the responder and

FIGURE 4 | The median of the predicted survival probability and the median of the observed survival probability by the Kaplan-Meier method with 95% CI. The x axis depicts the observed value; the y axis depicts the predicted values in the corresponding point.

FIGURE 5 | The Kaplan-Meier survival curves and evaluation of the number at risk from 1 to 3 years for the three risk groups using the "hdnom" package in R software. The p-value of the log-rank test is 0.00773.

that simple ADC values only show limited information (one dimensional information) that only reflect variability (high or low), not including geometric distribution. Texture features can overcome the above defects, having the potential to show and quantify pixels or the voxel geometric distribution. With the development of imaging analysis, much evidence has suggested that TA can aid clinicians in cancer diagnosis (26), staging (27), prognoses (28), and response assessments (15). In our study of 82 patients with the diagnosis of ESCC, interestingly, we showed that the pretreatment DWI texture features can provide useful prognostic information for ESCC patients. Finally, previous studies were mostly focused on limited tumor areas, such as contouring ROIs in the largest section, rather than the global tumor volume. In our study, to compensate for the 2D texture feature defects (12, 29, 30), 3D texture parameters were chosen to evaluate the prediction potentials.

We first analyzed the intensity histogram features with highly reflected distribution of ADC values. The other texture features mainly focused on the local and regional scales, which were used to analyze the interrelationship between pairs of voxels and arrangement of voxels. From microscopy, the order of voxels reflected local non-uniformities. Our analysis showed that IHIST_energy, m_contrast_1, Diff_homogeneity_2, m_Inversevariance_2, HISE, and LILE have strong and independent associations with the OS rates. The IHIST_energy measures the homogeneity of gray distribution. The higher value depicts more homogeneity than the lower one. m_Contrast_1 is a measure of the contrast or amount of local variation present in the ADC. The tumor usually has a large amount of local variations present in the image compared with the normal part. The other parameters (Diff_homogeneity_2, m_Inversevariance_2) were a measure of homogeneity of the image. This represents the change in the tumor gray level and reflects the aggregation of tumor cells on the macro level. The HISE measures the joint distribution of small zones and high gray-level values. The LILE has opposite characteristic to HISE and measures the joint distribution of large zones and high gray-level values. These features represent spatial ADC variability in esophageal tumors, explaining why these ADC texture features are better prognostic factors than simple global ADC values.

The OS variation in ESCC patients treated with CRT is highly related to tumor heterogeneity due to its intra-tumor spatial variation in the cellularity, angiogenesis, extravascular, extracellular matrix, and areas of necrosis. The high tumor heterogeneity was shown to have a poorer prognosis and treatment resistance (31). Ganeshan et al. (32) found that tumor heterogeneity in EC could be reflected by TA. Our study showed that six 3D texture parameters extracted from ADC maps can distinguish among the high-, median-, and low-risk group (Log-rank $\chi^2 = 9.7$; $P = 0.00773$). The idea behind this performance is that texture parameters can reflect the movement of water molecules and tumor heterogeneity. This may become a major mechanism to explain why texture parameters can accurately associate with the OS of ESCC.

To ensure the model's stability, the test-retest method was used to test the selected feature stability in the feature selection step. In the model validation step, the "leave one out" cross

non-responder groups (24). Wang et al. also found no direct correlation between the pretreatment ADC value and treatment response in EC (25). The reasons for the controversy could be

validation for both model validation and model calibration was used. Compared with the general sampling test (splitting their data into test and validation sets), the LASSO regularization scheme was used to prevent over fitting (33).

This study has an important clinical significance. Uncertainties remain in the treatment of ESCC, including the scope of the radiotherapy target area, the dose of radiation therapy, the consolidation chemotherapy maintenance period, the assessment of the clinical effect and so on. The cause of the above uncertainty remains a lack of effective means to describe ESCC heterogeneity. The texture features combined with conventional prognostic factors may present a more accurate predictive tool. Our research showed that texture features can be used to evaluate the prognosis of ESCC after CRT at the early phase. However, our study is limited by several factors, the most important of which is the prospective nature of the assessment using a relatively small group of patients. It is necessary to expand the sample size for further study to clearly explore the relationship between the global ADC value and OS. Another limitation of this study is that the tumor regions of interest were drawn manually; inter- and intra-observer variation could be reduced if automated methods were used in the future, particularly for multicenter studies.

CONCLUSIONS

Based on the ADC images, the texture parameters extracted by computer semi-automatic extraction are related to ESCC patient survival. This study confirms that the combination of ADC textures (histogram feature, GLCM feature, and ISZF feature) and conventional prognostic factors (radiation dose) can be used to generate robust models to predict OS. Future work needs to further verify the practical value of related parameters in clinical application.

DATA AVAILABILITY STATEMENT

The datasets for this manuscript are not publicly available because involving patient confidential information. Requests to access the datasets should be directed to zhenjli@seu.edu.cn.

ETHICS STATEMENT

Registration number and name of registry: 20091124-2. Project: Prognostic Value of Pretreatment Diffusion Weighted Magnetic Resonance Imaging based Texture in Concurrent Chemoradiotherapy of Esophageal Squamous Cell Cancer. Institute: Shandong Cancer Hospital Affiliated to Shandong University, Department of Radiation Oncology, the Fourth Hospital of Hebei Medical University. Research Summary: In this retrospective cohort study, we aims to include a total 100 patients with esophageal squamous cell carcinoma in Chinese population. MR imaging examination was performed before starting the treatment for tumor staging. The imaging protocol included: (1) 2D Fast Low-Angle Shot imaging (FLASH)/T1-weighted and DWI was obtained in axial planes. Privacy Highlights: patients privacy. Comments of Institutional Review Board: After carefully examination of the relative information, including researcher qualifications, research programme, CRF, and informed consents, the study was approved by the Institutional Review Board.

AUTHOR CONTRIBUTIONS

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

FUNDING

ACKNOWLEDGMENTS

We also thank ESMO and ASTRO for publishing our initial work.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2019.01057/full#supplementary-material

REFERENCES

1. Falk GW. Risk factors for esophageal cancer development. *Surg Oncol Clin N Am*. (2009) 18:469–85. doi: 10.1016/j.soc.2009.03.005
2. Harry VN, Semple SI, Parkin DE, Gilbert FJ. Use of new imaging techniques to predict tumour response to therapy. *Lancet Oncol*. (2010) 11:11. doi: 10.1016/S1470-2045(09)70190-1
3. Chavhan GB, Alsabban Z, Babyn PS. Diffusion-weighted imaging in pediatric body MR imaging: principles, technique, and emerging applications. *Radiographics*. (2014) 34:E73–88. doi: 10.1148/rg.343135047

4. Hein PA, Kremser C, Judmaier W, Griebel J, Pfeiffer KP, Kreczy A, et al. Diffusion-weighted magnetic resonance imaging for monitoring diffusion changes in rectal carcinoma during combined, preoperative chemoradiation: preliminary results of a prospective study. *Eur J Radiol*. (2003) 45:214–22. doi: 10.1016/S0720-048X(02)00231-0
5. Kremser C, Judmaier W, Hein P, Griebel J, Lukas P, de Vries A. Preliminary results on the influence of chemoradiation on apparent diffusion coefficients of primary rectal carcinoma measured by magnetic resonance imaging. *Strahlenther Onkol*. (2003) 179:641–9. doi: 10.1007/s00066-003-1045-9

6. Koh DM, Collins DJ. Diffusion-weighted MRI in the body: applications and challenges in oncology. *Am J Roentgenol.* (2007) 188:1622–35. doi: 10.2214/AJR.06.1403

7. Sakurada A, Takahara T, Kwee TC, Yamashita T, Nasu S, Horie T, et al. Diagnostic performance of diffusion-weighted magnetic resonance imaging in esophageal cancer. *Eur Radiol.* (2009) 19:1461–9. doi: 10.1007/s00330-008-1291-4

8. Castellano G, Bonilha L, Li LM, Cendes F. Texture analysis of medical images. *Clin Radiol.* (2004) 59:1061–9. doi: 10.1016/j.crad.2004.07.008

9. Cook GJ, Yip C, Siddique M, Goh V, Chicklore S, Roy A, et al. Are pretreatment 18F-FDG PET tumor textural features in non-small cell lung cancer associated with response and survival after chemoradiotherapy? *J Nucl Med.* (2013) 54:19–26. doi: 10.2967/jnumed.112.107375

10. Harrison L, Dastidar P, Eskola H, Jarvenpaa R, Pertovaara H, Luukkaala T, et al. Texture analysis on MRI images of non-Hodgkin lymphoma. *Comput Biol Med.* (2008) 38:519–24. doi: 10.1016/j.compbiomed.2008.01.016

11. 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/ncomms5644

12. Ng F, Kozarski R, Ganeshan B, Goh V. Assessment of tumor heterogeneity by CT texture analysis: can the largest cross-sectional area be used as an alternative to whole tumor analysis? *Eur J Radiol.* (2013) 82:342–8. doi: 10.1016/j.ejrad.2012.10.023

13. Wang L, Han C, Zhu S, Shi G, Liu J, Liu L, et al. A pathological comparative study of diffusion-weighted imaging and computed tomography in determination of lesion length for esophageal carcinoma. *Chin J Radiat Oncol.* (2015) 24:373–6.

14. Collewet G, Strzelecki M, Mariette F. Influence of MRI acquisition protocols and image intensity normalization methods on texture classification. *Magn Reson Imaging.* (2004) 22:81–91. doi: 10.1016/j.mri.2003.09.001

15. Tixier F, Le Rest CC, Hatt M, Albarghach N, Pradier O, Metges JP, et al. Intratumor heterogeneity characterized by textural features on baseline 18F-FDG PET images predicts response to concomitant radiochemotherapy in esophageal cancer. *J Nucl Med.* (2011) 52:369–78. doi: 10.2967/jnumed.110.082404

16. Haralick RM, Shanmugam K, Dinstein I. Textural features for image classification. *IEEE Trans Syst Man Cybern.* (1973) 3:22. doi: 10.1109/TSMC.1973.4309314

17. Fried DV, Tucker SL, Zhou S, Liao Z, Mawlawi O, Ibbott G, et al. Prognostic value and reproducibility of pretreatment CT texture features in stage III non-small cell lung cancer. *Int J Radiat Oncol Biol Phys.* (2014) 90:834–42. doi: 10.1016/j.ijrobp.2014.07.020

18. Balagurunathan Y, Gu Y, Wang H, Kumar V, Grove O, Hawkins S, et al. Reproducibility and prognosis of quantitative features extracted from CT Images. *Transl Oncol.* (2014) 7:72–87. doi: 10.1593/tlo.13844

19. Chen X. Adaptive elastic-net sparse principal component analysis for pathway association testing. *Stat Appl Genet Mol Biol.* (2011) 10:48. doi: 10.2202/1544-6115.1697

20. Li C, Wei X, Dai H. Adaptive elastic net method for cox model. *Stat Theory. arXiv:1507.06371* (2015).

21. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer.* (2009) 45:228–47. doi: 10.1016/j.ejca.2008.10.026

22. Koyama T, Umeoka S, Saga T, Kataoka M, Tamai K, Watanabe G, et al. The ADC value in esophageal caners on diffusion-weighted MR images for predicting effect of chemoradiotherapy. *Proc Int Soc Magn Reson Med.* (2006) 14:2820.

23. Aoyagi T, Shuto K, Okazumi S, Shimada H, Kazama T, Matsubara H. Apparent diffusion coefficient values measured by diffusion-weighted imaging predict chemoradiotherapeutic effect for advanced esophageal cancer. *Dig Surg.* (2011) 28:252–7. doi: 10.1159/000328770

24. Imanishi S, Shuto K, Aoyagi T, Kono T, Saito H, Matsubara H. Diffusion-weighted magnetic resonance imaging for predicting and detecting the early response to chemoradiotherapy of advanced esophageal squamous cell carcinoma. *Dig Surg.* (2013) 30:240–8. doi: 10.1159/000351435

25. Wang L, Han C, Zhu S, Shi G, Wang Q, Tian R, et al. Investigation of using diffusion-weighted magnetic resonance imaging to evaluate the therapeutic effect of esophageal carcinoma treatment. *Oncol Res Treat.* (2014) 37:112–6. doi: 10.1159/000360210

26. Ba-Ssalamah A, Muin D, Schernthaner R, Kulinna-Cosentini C, Bastati N, Stift J, et al. Texture-based classification of different gastric tumors at contrast-enhanced CT. *Eur J Radiol.* (2013) 82:e537–43. doi: 10.1016/j.ejrad.2013.06.024

27. Dong X, Wu P, Sun X, Li W, Wan H, Yu J, et al. Intra-tumour 18F-FDG uptake heterogeneity decreases the reliability on target volume definition with positron emission tomography/computed tomography imaging. *J Med Imaging Radiat Oncol.* (2015) 59:338–45. doi: 10.1111/1754-9485.12289

28. Hatt M. Prognostic value of 18F FDG PET image-based parameters in o esophageal cancer and impact of tumour delineation methodology. *Eur J Nucl Med Mol Imaging.* (2011) 38:1191–202. doi: 10.1007/s00259-011-1755-7

29. Ganeshan B, Miles KA, Young RC, Chatwin CR, Gurling HM, Critchley HD. Three-dimensional textural analysis of brain images reveals distributed grey-matter abnormalities in schizophrenia. *Eur Radiol.* (2010) 20:941–8. doi: 10.1007/s00330-009-1605-1

30. Ganeshan B, Miles KA, Young RC, Chatwin CR. Three-dimensional selective-scale texture analysis of computed tomography pulmonary angiograms. *Invest Radiol.* (2008) 43:382–94. doi: 10.1097/RLI.0b013e3181690091

31. Marusyk A, Polyak K. Tumor heterogeneity: causes and consequences. *Biochim Biophys Acta.* (2010) 1805:105–17. doi: 10.1016/j.bbcan.2009.11.002

32. Ganeshan B, Skogen K, Pressney I, Coutroubis D, Miles K. Tumour heterogeneity in oesophageal cancer assessed by CT texture analysis: preliminary evidence of an association with tumour metabolism, stage, and survival. *Clin Radiol.* (2012) 67:157–64. doi: 10.1016/j.crad.2011.08.012

33. Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med.* (1996) 15:361–87.

STAT-ART: The Promise and Practice of a Rapid Palliative Single Session of MR-Guided Online Adaptive Radiotherapy (ART)

Kathryn E. Mittauer [1,2]*, Patrick M. Hill [1], Mark W. Geurts [1,3], Anna-Maria De Costa [1], Randall J. Kimple [1], Michael F. Bassetti [1] and John E. Bayouth [1]

[1] Department of Human Oncology, UW Carbone Cancer Center, University of Wisconsin-Madison, Madison, WI, United States, [2] Department of Radiation Oncology, Baptist Health South Florida, Miami Cancer Institute, Miami, FL, United States, [3] Department of Radiation Oncology, Aspirus Wausau Hospital, Aspirus Inc., Wausau, WI, United States

This work describes a novel application of MR-guided online adaptive radiotherapy (MRgoART) in the management of patients whom urgent palliative care is indicated using statum-adaptive radiotherapy (STAT-ART). The implementation of STAT-ART, as performed at our institution, is presented including a discussion of the advantages and limitations compared to the standard of care for palliative radiotherapy on conventional c-arm linacs. MR-based treatment planning techniques of STAT-ART for density overrides and deformable image registration (DIR) of diagnostic CT to the treatment MR are also addressed.

Keywords: online adaptive radiotherapy, MRgoART, MR-guidance, MRgRT, ART, MR-guided radiotherapy, palliative radiation, deformable image registration

INTRODUCTION

The American Cancer Society estimated that annually in 2019 there are 1.76 million new cases of cancer with 606,880 cancer deaths (1). Most cancer deaths are associated with a decreased quality of life and painful end-of-life, likely due to loco-regional or metastatic disease progression (2, 3). Palliative radiotherapy (RT) allows for the management of patients with advanced stage cancer. Palliative RT directly relieves obstructions, bleeding, and cancer-related pain for patients not a candidate for or responding to opioid medication (2–13).

As the sensitivity and accuracy of cancer detection and subsequently cancer treatments progressively improve, the life expectancy for cancer patients is steadily rising even with metastatic disease (3). Thus, the continued management of these patients is of great importance (3). Due to multiple steps in standard radiation planning processes, palliative treatment may take 3–7 days post-consultation before the first treatment fraction is performed (3). Improving the palliative RT workflow by utilizing advanced technology to offer rapid, same day treatment can reduce the pre-treatment time period, allowing for near-immediate pain-relief and an improved quality of life for these patients.

The MRIdian™ cobalt and more recently released MRIdian linac (ViewRay Inc., Cleveland, OH) is an MR-guided radiotherapy (MRgRT) platform that integrates magnetic resonance imaging (MRI), radiotherapy delivery, treatment planning, image registration, and treatment record and delivery into a single unit (14, 15). The integrated approach enables MR-guided online adaptive radiotherapy (MRgoART), where the care team designs and delivers a treatment plan based on patient anatomy and position at the time of treatment setup [Mittauer et al. (under review); (16, 17)]. MRgoART, or simply

online adaptive radiotherapy (ART), offers the opportunity for rapid and accurate palliative online adaptive radiation therapy, i.e., statum-ART (STAT-ART).

The purpose of this work is to describe a potential paradigm change in the management of palliative care in radiotherapy using STAT-ART. The implementation of STAT-ART, as performed at our institution, is presented including a discussion of the advantages and limitations compared to the standard of care for palliative radiotherapy performed on conventional c-arm linacs. MR-based treatment planning techniques for density overrides and deformable image registration (DIR) of diagnostic CT to the treatment MR are also addressed.
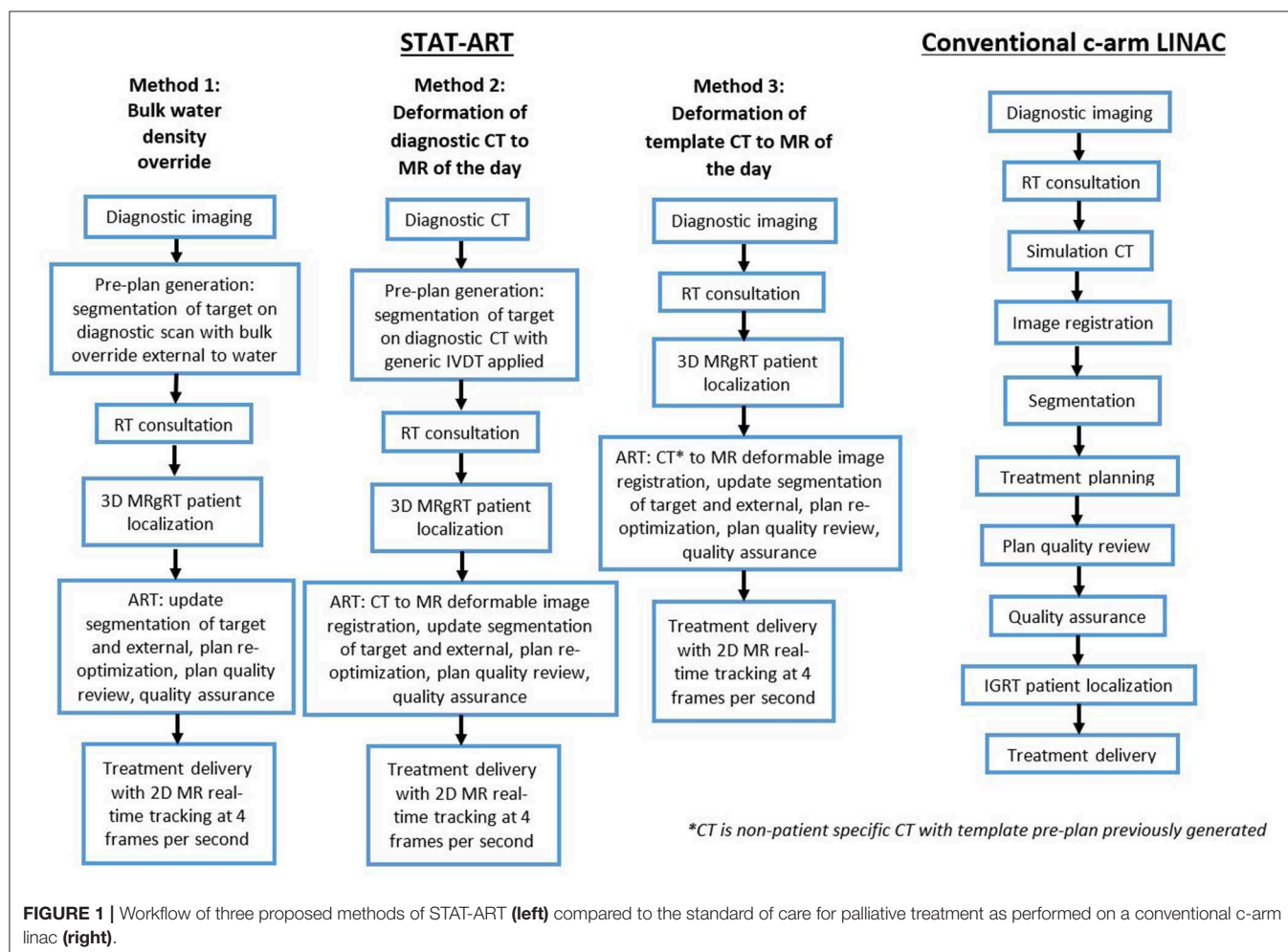
CONVENTIONAL LINAC WORKFLOW FOR URGENT PALLIATIVE TREATMENT

Conventional radiotherapy workflow utilizes a serial-based process map. Tasks are performed through multiple applications and platforms (i.e., PACS, simulator system, image registration software, segmentation software, treatment planning system, treatment delivery system, quality assurance software, record,

and verify system). Each step of the serial radiotherapy workflow is executed by a unique stakeholder (i.e., physician, physicist, dosimetrist, therapist).

The workflow process for conventional radiotherapy utilizing a c-arm linac will be briefly detailed here and is displayed in **Figure 1**. When a patient presents for initial consultation with their radiation oncologist, previously acquired diagnostic scans are reviewed at the time of consultation. Following consultation, the patient receives a CT simulation appointment in which a CT scan is acquired for purposes of treatment planning and treatment setup localization. Image registration of the diagnostic data to the planning CT scan is often performed to aid the physician in segmentation of the target and relevant surrounding anatomy. On the planning CT scan, the dosimetrist creates a treatment plan which subsequently undergoes plan quality review by the physician and physicist in addition to a quality assurance (QA) assessment. The patient is brought in for their treatment appointment, where x-ray based imaging is performed to localize the patient into the position as acquired at the time of the CT simulation, followed by radiotherapy treatment delivery.

To accelerate the radiotherapy process when a patient presents with urgent obstructions, bleeding, and/or pain, the above

FIGURE 1 | Workflow of three proposed methods of STAT-ART **(left)** compared to the standard of care for palliative treatment as performed on a conventional c-arm linac **(right)**.

workflow is consolidated at the potential cost of the overall treatment plan quality. The treatment plan complexity is often limited for palliative cases to a parallel-opposed beam geometry with rectangular fields, defined by jaws alone rather than the use of more sophisticated multileaf collimation (MLC). Such plans are advantageous for simplifying other radiotherapy workflow processes such as simulation, setup, and treatment. However, utilizing more conformal planning and delivery techniques with more beam angles and more sophisticated collimation schemes can reduce dose to uninvolved organs and tissues.

As previously described, the standard of care utilizes a workflow of acquiring a CT simulation of the patient in setup position followed by a treatment planning session. The process may take several hours to several days. Additional steps to accelerate the process for urgent palliative radiotherapy include forgoing the simulation scan and performing 2D treatment planning based on 2D MV imaging on the radiotherapy treatment unit and calculation of dose to an arbitrary point within the patient. Without knowledge of spatial orientation of the tumor in treatment position, large target margins, and non-conformal dose distributions are required.

STAT-ART WORKFLOW FOR URGENT PALLIATIVE TREATMENT

Overview of STAT-ART Workflow

The STAT-ART workflow takes advantage of the MRgoART features of the MRIdian to enable efficient adaptive planning capabilities and expedite the overall workflow [Mittauer et al. (under review); (16, 17)]. Prior to patient arrival and consultation, the generalized workflow (**Figure 1**) includes either generation of a patient-specific pre-plan or selection of a preexisting non-patient specific template plan. For the patient-specific pre-plan generation, the radiation oncologist performs segmentation of the target volume on the diagnostic dataset followed by the medical physicist or dosimetist performing the treatment planning (i.e., beam geometry, defining MLC aperture, beam weight optimization, and dose normalization). Details of the planning technique is described in the following sections.

Following consultation, the patient is setup in an arbitrary treatment position and localized using and a 3D volumetric balanced steady-state free precession sequence (TrueFISP) MR scan (**Figure 2**) (18). A template-plan or pre-plan from the diagnostic scan is then adapted online based on the actual treatment geometry including updates to segmentation, beam apertures, optimization, and/or dose normalization. A plan quality visual inspection and calculation-based QA are performed for plan fidelity (17), followed by treatment delivery. The STAT-ART workflow will be dependent on institutional resources and MRgoART staffing model, and is generally executed in real-time by multiple staff members including a combination of therapist(s), a medical physicist, and a radiation oncologist.

Planning Technique

Development of the STAT-ART protocol has emphasized efficiency. For the pre-plan generation, segmentation performed by the radiation oncologist is kept to a minimum with only two regions of interest requiring contouring: external or skin and
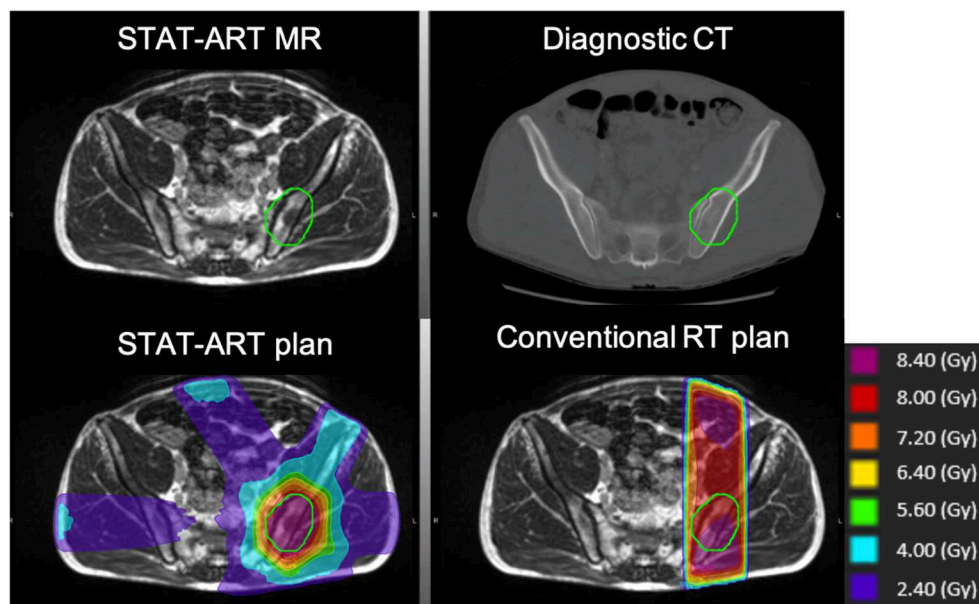
FIGURE 2 | Clinical STAT-ART case of a pelvic bony metastasis treated to 8 Gy in a single fraction to the planning target volume in green. Comparison of diagnostic CT used for pre-planning (top right) and improved gross disease visualization on the TrueFISP treatment planning MR obtained using the MRIdian (top left). The resulting dose distribution for STAT-ART plan of a six-beam cobalt-60 dose distribution (bottom left) in comparison to a conventional plan with 10 MV AP/PA beams (bottom right). The STAT-ART plan shows a marked reduction of the high dose volume, particularly in anterior regions of normal tissues.

a target volume. Initial planning of the pre-plan or template plan is performed by the medical physicist or dosimetrist with a single isocenter and 3D conformal beams defined by the MLC (19, 20). 3D conformal planning of the pre-plan or template plan involves defining an isocenter point of interest, inputting beams, setting gantry angles for ideal geometry, defining the MLC aperture, optimization of beam weights, and a Monte Carlo dose calculation with magnetic field corrections. As previously noted, online adaptive is then performed to make the corresponding plan adjustments to the patient's on table anatomy through segmentation updates to the target and corresponding plan MLC aperture shape and beam weight through plan re-optimization.

Typically, STAT-ART treatment plans use a six-beam arrangement, equivalent to two gantry positions for the three ^{60}Co sources on the MRIdian cobalt. For the MRIdian linac, the higher dose rate has enabled comparable delivery times to the MRIdain cobalt even with an increase in gantry rotation required for the six-beam arrangement. Due to the utilization of 3D conformal planning technique, STAT-ART plans have similar delivery times to conventional AP/PA beam arrangements while allowing for more comparatively superior dose distributions (**Figure 2**).

Electron Density for MR-Based Planning

MR-based planning requires electron density information for dose calculation purposes. We present three strategies (**Figure 1**) in the STAT-ART workflow for density propagation using bulk density overrides or deformable image registration for respective homogenous or heterogenous dose calculations.

Bulk Density Override

For non-thoracic based anatomical sites, a single bulk density override of the external region of interest to water can be used for electron density propagation. In this method of STAT-ART, a diagnostic CT scan or even a diagnostic MR scan can be utilized as the primary dataset of the STAT-ART pre-plan. The deformation of the CT scan to the MR scan in MRgoART is eliminated; advantageous when large anatomical mismatches are anticipated and therefore eliminates the need and time to perform manual density corrections. During the online adaptive process with a single bulk density override, the external contour is simply defined based on the treatment MR scan, enabling an efficient and robust method of density propagation.

Deformation of Diagnostic CT to MR of the Day

An alternative to density override(s) is to utilize electron density information obtained from a diagnostic CT scan. In the MRgoART workflow of the MRIdian, the pre-plan primary dataset of the diagnostic CT scan is deformably registered to the frame of reference of the treatment MR scan utilizing an inverse-consistent, free-form multi-modality DIR with a similarity metric of mutual information and regularization proportional to the Jacobian of the deformation vector field [Mittauer et al. (under review)].

Beyond anatomical setup differences that may require manual density corrections, additional deformation challenges may include a limited field of view on the diagnostic CT scan

with missing tissue information. When density corrections are necessary during the time of adaptation, an available override contour of air, bone, and/or soft tissue may be utilized to enable manual electron density edits. It is recommended to input such empty contours with pre-defined electron density of respective air, bone, and soft tissue to the pre-plan to enable the approach of manual electron density edits during online adaptation.

Deformation of Template CT to MR of the Day

When prior diagnostic imaging is not readily accessible, a template plan based on the anatomical site of interest (i.e., pelvis, abdomen, thorax, extremity, etc.) can be adapted to the patient's setup at the time of treatment. The alternative method of utilizing a template plan eliminates the time for pre-plan generation, and may be most applicable for urgent palliative cases providing machine availability. The deformation workflow of the CT scan to the treatment MR scan remains the same as the above method, with larger potential for density corrections to be required.

DISCUSSION

Adoption of Hypofractionated Palliative Care Toward Single Fraction

In a recent review of palliative radiotherapy, Rich et al. found the adoption of a single fraction course underutilized compared to conventional fractionated course (21), even with recent recommendations (22–24) emphasizing a single-fraction or short-course palliative care. The movement toward a single fraction course for palliative radiotherapy enables two key benefits: cost-reduction and patient time and convenience (21, 25–27). As patients with oligometastatic disease live longer, the need for a more sophisticated, short-course planning and delivery approach is evident over historical parallel-opposed techniques of 30 Gy in 10 fractions.

With MR linacs becoming more common place (28), the adoption of a single fraction course with MR-guidance is a viable approach. The STAT-ART workflow leverages existing volumetric imaging from radiology, eliminating an additional simulation CT. The STAT-ART technique utilizes routine clinical MRgRT workflows, i.e., adaptive radiotherapy. Moreover, MR-guidance enables greater confidence based on real-time image guidance and greater dose conformality, supporting a single fraction approach. A randomized trial comparing efficacy and toxicity between single fraction palliative care on CT-based IGRT vs. MR-based IGRT has yet to be conducted. While several randomized single fraction vs. multiple fraction trials have been carried out, the concerns of toxicities and efficacy of the 8 Gy in 1 fraction regimen still remain. Shuja et al. and Howell et al. described utilization of 3–5 cm margin and <2 cm margin, respectively, from the radiographic involvement for single fraction palliative radiotherapy on conventional c-arm linac modalities (29, 30). The superior bony metastasis visualization combined with improved soft tissue visualization allow for greater precision and enable the utility of a reduced margin in MRgRT of 0.3–0.5 cm as previously demonstrated by Mittauer et al. (16). Howell et al. specifically cited radiation oncologists' concerns of gastrointestinal (GI) toxicities associated with 8 Gy in

a single fraction (30). The online adaptive capabilities, to visualize adjacent organs at risk (OARs) and modify dose based on neighboring GI OARs, combined with a smaller margin required due to the reduction in setup uncertainty enable MR-guided single fraction radiotherapy a clear benefit.

Furthermore, to manage the increase in the number of oligometastatic patients, reducing the number of treatments to fewer fractions could potentially lessen the overall burden on hospitals, and ultimately reduce the number of machines required per patient population.

STAT-ART at the University of Wisconsin-Madison

Our STAT ART program at the University of Wisconsin-Madison was implemented in October 2015 on the MRIdian cobalt, and since then transitioned to the MRIdian linac. Our initial experiences of the STAT-ART program have been briefly reported by Hill et al. (31) and De Costa et al. (32), and includes a retrospective review of the first 18 patients treated with STAT-ART from October 2015 to November 2016.

The indication for STAT-ART included patients with metastatic cancer presenting with pain, obstruction, and bleeding. The majority of STAT-ART patients were treated with a prescription of 8 Gy in a single fraction. STAT-ART planning and treatment delivery was typically >30 min between the patient entering and exiting the treatment vault, compared with a mean time from CT simulation to delivery of first treatment of 29.5 h (95% CI, 23.7–35.2) for a similar sample of urgent palliative cases planned and treated with the conventional radiotherapy workflow. The median delivery time of STAT-ART was 122 s (N = 18 patients).

Excellent clinical outcomes were observed and were in line with historical and sampled controls: pain reduction in 11 of 14 patients, improvement of obstructive symptoms in 3 of 3 patients, and hemostasis in 1 of 1 patient. Overall, physician and patient response to the program has been positive, as plan quality has improved while time commitments have been comparable to or less than a conventional simulation-and-treatment workflow. Future efforts include characterizing the dose difference to organs at risk and conformality metrics between STAT-ART plans and conventional parallel-opposed beam geometries.

MR-Based Planning for Bony Metastases

MR-based treatment planning offers better soft tissue contrast for target delineation as compared to CT simulation (16). However, one particularly interesting finding of the STAT-ART program has been the ability of the MRIdian TrueFISP MR sequences used for treatment planning to identify contrast other than in soft tissues, as shown in **Figure 2** for bony metastases. Because pre-plan contours are routinely updated to encompass disease identified on the treatment planning MR, the ability to target disease in bone has been invaluable.

Additional Workflow Advantages

Through performing treatment planning, contouring, image registration, and treatment delivery on a single platform such as the MRIdian system not only can allow for improvement in

clinical efficiency, but also possibly allow for the decrease in the clinical errors from the use of multiple modalities and planning systems. The MRIdian system eliminates the time and need for treatment planning by dosimetry when utilizing template plans. Furthermore, simulation is done in the true treatment position on MRIdian, an advantage over conventional linac-based RT, enabling both enhanced contrast for target delineation and reduction of setup uncertainties.

Dose Calculation and Deformation Considerations

There are limitations in dose calculation accuracy of STAT-ART when performing bulk density overrides or using deformed diagnostic CT data. For example, with a bulk density override of water for patient anatomy of the pelvis and abdomen, the calculation error is likely on the order of 2%; as demonstrated by Lee et al. for pelvis with absolute dose differences ranging 0–5% inside the planning target volume for uniform density override of water compared to dose calculated on the respective CT scan (33). Larger magnitude of errors would present for other anatomical sites such as thorax/lung. Here, deformation of the diagnostic CT to the treatment MR would be more appropriate.

Electron density propagation of the diagnostic CT to the treatment MR presents has additional uncertainties. The image value to density conversion may not be characterized for the energy spectrum and/or applicable CT scanner of the diagnostic CT dataset at hand. While it may be feasible to characterize all CT scanners in an institution's radiology department, the body of work would be non-trivial and not inclusive of scanners from outside institutions for patients referred for treatment. A potentially more practical approach would be to incorporate CT energy-dependent image value to density table (IVDT) curves, as inter-scanner dependences are minimal to image value variation and on the order of acceptable dose calculation uncertainties for palliative care. A phantom with a range density inserts can be utilized to quantify the Hounsfield unit values as a function of CT energy. Repeat monoenergetic CT scans over an energy range would be acquired to benchmark the IVDT dependent curves. During initial planning the user would then select the respective IVDT curve based on the DICOM tag of the patient's diagnostic CT.

A limitation of the fidelity of the propagation of electron density is the overall voxel resolution. Partial-voxel effects can influence the deformable image registration quality as can be propagated from the initial diagnostic CT scan and/or in the resultant deformed CT, resampled in the resolution and frame of reference of the treatment planning MR. The user is recommended to note the influence partial-voxel effects on deformed electron density accuracy and the potential impact on dose calculation accuracy.

Another challenge of the deformable image registration of the diagnostic CT to the treatment MR is the large anatomical differences between the scans. **Figure 2** highlights the posterior anatomical deformations between the curved tabletop of the diagnostic CT and the flat tabletop of the radiotherapy system. Additional deformation differences may include patient arm

position or even missing tissue due to limited field of view on the diagnostic CT scan. All of these require review of the CT-MR deformation and may require additional effort and time during the online adaptive workflow to manually correct the electron density using segmentation and overrides.

Challenges to Implementation

The STAT-ART program relies on capabilities of MRgoART. This work has been presented based on the platform of the MRIdian system as has been implemented at our institution. Modified practice of STAT-ART for other systems with CT/MR on rails or other IGRT systems can be employed. The work to commission and to implement the deformable image registration and dose calculation of MRgoART has been previously described [Mittauer et al. (under review)].

There is some potential for errors to occur in the MRgoART workflow since the plan is adapted on the fly. However, for clinical MRgRT users, MRgoART has become a routine part of everyday workflow (34). The MRgoART utilizes a secondary calculation-based QA to verify plan fidelity. For 3D conformal plans this follows conventional radiotherapy workflow as more sophisticated measurement-based QA are not necessary if the beam model has been appropriately characterized and validated. Secondly, calculation-based methods have been previously shown to be in line with measured-based QA for the MRgoART process (17).

Another unique consideration when implementing MRgoART and STAT-ART is the overall time the patient is on the table in the treatment position. While the STAT-ART process rapidly decreases the time from consultation to treatment for these urgent palliative cases, the "table time" may be longer due to the adaption process. Some patients may not tolerate the 20–30 min on the treatment table as required for the STAT-ART process due to symptomatic pain.

Alternative Rapid Palliative RT on TomoTherapy

The University of Virginia have implemented a rapid palliative radiotherapy technique using CT-based IGRT of TomoTherapy (Accuracy Inc, Madison, WI) with their "STAT RAD" program (2, 3). The STAT RAD workflow utilizes the on-board MVCT capabilities of TomoTherapy to simulate the patient in treatment position followed by rapid treatment plan generation, quality assurance of the plan with exit dosimetry through the on-board CT detector, and treatment delivery. Since the treatment planning capabilities are not integrated into the simulation and

delivery console, plan generation is performed on a separate work station, eliminating an online adaptive approach. The University of Virginia has successfully piloted the program with 50 patient treats reported to date in 2012 (3).

CONCLUSION

The integration of a simulator, treatment planning system, and delivery system into a single platform enables the opportunity of STAT-ART, a rapid-access treatment for patients presenting with urgent palliative needs. Electron density information for MR-based planning of STAT-ART without formal CT simulation can be incorporated with either a bulk density override or deformable image registration of diagnostic CT to the treatment MR. The online adaptive features of STAT-ART enable adaptation of a preexisting pre-plan or template plan, reducing the time pressure for urgent palliative radiotherapy. Another key advantage of MRgoART is the superior plan and treatment quality as real-time plan adaptation is performed to the anatomy at treatment compared to the simulation day anatomy as performed in conventional radiotherapy. STAT-ART has great potential in the management of the palliative radiotherapy, making efficient use of both staffing time and resources and expediting palliative care with similarly successful clinical outcomes.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this manuscript will be made available by the authors, without undue reservation, to any qualified researcher.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by University of Wisconsin, Institutional Review Board. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

KM, PH, MG, MB, and JB designed the workflow. PH and A-MD performed the data summary. KM drafted the manuscript. KM, PH, MG, A-MD, RK, MB, and JB iteratively revised the manuscript.

REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *Cancer J Clin.* (2019) 69:7–34. doi: 10.3322/caac.21551
2. Jones RT, Handsfield L, Read PW, Wilson DD, Van Ausdal R, Schlesinger DJ, et al. Safety and feasibility of STAT RAD: Improvement of a novel rapid tomotherapy-based radiation therapy workflow by failure mode and effects analysis. *Pract Radiat Oncol.* (2015) 5:106–12. doi: 10.1016/j.prro.2014.03.016
3. Wilson D, Ke S, Wensha Y, Ryan J, Neal D, Paul R. STAT RAD: a potential real-time radiation therapy workflow. In: Natanasabapathi G, editor. *Modern*

Practices in Radiation Therapy. Rijeka: InTech (2012). p. 24–40. doi: 10.5772/34285
4. Bruera E, Kim HN. Cancer pain. *JAMA.* (2003) 290:2476–9. doi: 10.1001/jama.290.18.2476
5. Cleeland CS, Gonin R, Hatfield AK, Edmonson JH, Blum RH, Stewart JA, et al. Pain and its treatment in outpatients with metastatic cancer. *N Engl J Med.* (1994) 330:592–6 doi: 10.1056/NEJM199403033300902
6. McGuire DB. Occurrence of cancer pain. *J Natl Cancer Inst Monogr.* (2004) 32:51–6. doi: 10.1093/jncimonographs/lgh015

7. Gordon DB, Dahl JL, Miaskowski C, McCarberg B, Todd KH, Paice J, et al. American pain society recommendations for improving the quality of acute and cancer pain management: American pain society quality of care task force. *Arch Intern Med.* (2005) 165:1574–80. doi: 10.1001/archinte.165.14.1574

8. Mercadante S. Scoring the effect of radiotherapy for painful bone metastases. *Support Care Cancer.* (2006) 14:967–9. doi: 10.1007/s00520-006-0036-7

9. Regine WF, Scott C, Murray K, Curran W. Neurocognitive outcome in brain metastases patients treated with accelerated-fractionation vs. accelerated-hyperfractionated radiotherapy: an analysis from Radiation Therapy Oncology Group Study 91-04. *Int J Radiat Oncol Biol Phys.* (2001) 51:711–7. doi: 10.1016/S0360-3016(01)01676-5

10. Stephen L. Palliative whole-brain radiotherapy fractionation: Convenience versus cognition? *Cancer.* (2007) 110:2363–5. doi: 10.1002/cncr.23069

11. Aoyama H, Tago M, Kato N, Toyoda T, Kenjyo M, Hirota S, et al. Neurocognitive function of patients with brain metastasis who received either whole brain radiotherapy plus stereotactic radiosurgery or radiosurgery alone. *Int J Radiat Oncol Biol Phys.* (2007) 68:1388–95. doi: 10.1016/j.ijrobp.2007.03.048

12. Fairchild A, Harris K, Barnes E, Wong R, Lutz S, Bezjak A, et al. Palliative thoracic radiotherapy for lung cancer: a systematic review. *J Clin Oncol.* (2008) 26:4001–11. doi: 10.1200/JCO.2007.15.3312

13. Langendijk JA, Velde GP, Aaronson NK, de Jong JM, Muller MJ, Wouters EF. Quality of life after palliative radiotherapy in non-small cell lung cancer: a prospective study. *Int J Radiat Oncol Biol Phys.* (2000) 47:149–55. doi: 10.1016/S0360-3016(99)00540-4

14. Hill PM, Mittauer K. mr guidance in radiation therapy. In: Alaei P, Ding GX, editors. *Image Guidance in Radiation Therapy: Techniques, Accuracy, and Limitations. American Association of Physicists in Medicine Medical Physics Monograph No. 39.* Madison, WI: Medical Physics Publishing (2018). p. 1–30.

15. Mutic S, Dempsey JF. The ViewRay system: magnetic resonance-guided and controlled radiotherapy. *Semin Radiat Oncol.* (2014) 24:196-9. doi: 10.1016/j.semradonc.2014.02.008

16. Mittauer KE, Paliwal B, Hill P, Bayouth JE, Geurts MW, Baschnagel AM, et al. A new era of image guidance with magnetic resonance-guided radiation therapy for abdominal and thoracic malignancies. *Cureus.* (2018) 10:e2422. doi: 10.7759/cureus.2422

17. Mittauer K, Hill P, Bayouth J. Validation of plan verification QA and workflow process through benchmark plans for commissioning of an mr-guided online adapted radiotherapy (ART) Program. *Med Phys.* (2017) 44:2615–720.

18. Ginn JS, Agazaryan N, Cao M, Baharom U, Low DA, Yang Y, et al. Characterization of spatial distortion in a 0.35 T MRI-guided radiotherapy system. *Phys Med Biol.* (2017) 62:4525–40. doi: 10.1088/1361-6560/aa6e1a

19. Mittauer KE, Yadav P, Paliwal B, Bayouth JE. Characterization of positional accuracy of a double-focused and double-stack multileaf collimator on an MR-guided radiotherapy (MRgRT) Linac using an IC-Profiler Array. *Med Phys.* (in press).

20. Mittauer KE, Paliwal B, Hill P, Bayouth JE. Characterization and longitudinal assessment of daily quality assurance on an MR-guided Radiotherapy (MRgRT) Linac. *JACMP.* (2019) 99:S126.

21. Rich SE, Chow R, Raman S, Liang Zeng K, Lutz S, Lam H, et al. Update of the systematic review of palliative radiation therapy fractionation for bone metastases. *Radiother Oncol.* (2018) 126:547–57. doi: 10.1016/j.radonc.2018.01.003

22. Lutz S, Balboni T, Jones J, Lo S, Petit J, Rich SE, et al. Palliative radiotherapy for bone metastases: Update of an ASTRO evidence-based guideline. *Pract Radiat Oncol.* (2016) 7:4–12. doi: 10.1016/j.prro.2016.08.001

23. Ganesh V, Chan S, Raman S, Chow R, Hoskin P, Lam H, et al. A review of practice and clinical guidelines in the palliative radiation treatment of uncomplicated bone metastases. *Radiother Oncol.* (2017) 124:38–44. doi: 10.1016/j.radonc.2017.06.002

24. Ashworth A, Kong W, Chow E, Mackillop WJ. Fractionation of palliative radiation therapy for bone metastases in Ontario: do practice guidelines guide practice? *Int J Radiat Oncol Biol Phys.* (2016) 94:31–9. doi: 10.1016/j.ijrobp.2015.07.2291

25. Konski A. Radiotherapy is a cost-effective palliative treatment for patients with bone metastases from prostate cancer. *Int J Radiat Oncol Biol Phys.* (2004) 60:1373–8. doi: 10.1016/j.ijrobp.2004.05.053

26. Konski A, James J, Hartsell W, Leibenhaut MH, Janjan N, Curran W, et al. Economic analysis of radiation therapy oncology Group 97–14: multiple versus single fraction radiation treatment of patients with bone metastases. *Am J Clin Oncol.* (2009) 32:423–8. doi: 10.1097/COC.0b013e31818da9f7

27. Pollicino CA, Turner SL, Roos DE, O'Brien PC. Costing the components of pain management: analysis of Trans-Tasman Radiation Oncology Group trial (TROG 96.05): one versus five fractions for neuropathic bone pain. *Radiother Oncol.* (2005) 76:264–9. doi: 10.1016/j.radonc.2005.07.003

28. Bayouth JE, Low DA, Zaidi H. MRI-linac systems will replace conventional IGRT systems within 15 years. *Med Phys.* (2019) 46:3753–6. doi: 10.1002/mp.13657

29. Shuja M, Elghazaly AA, Iqbal A, Mohamed R1, Marie A1, Tunio MA, et al. Efficacy of 8 Gy Single Fraction Palliative Radiation Therapy in Painful Bone Metastases: A Single Institution Experience. *Cureus.* (2018) 10:e2036. doi: 10.7759/cureus.2036

30. Howell DD, James JL, Hartsell WF, Suntharalingam M, Machtay M, Suh JH, et al. Single-fraction radiotherapy versus multifraction radiotherapy for palliation of painful vertebral bone metastases—Equivalent efficacy, less toxicity, more convenient. *Cancer.* (2013) 119:888–96. doi: 10.1002/cncr.27616

31. Hill P, Geurts M, Mittauer K, Bayouth J. SU-F-J-110: MRI-guided single-session simulation, online adaptation, and treatment. *Med Phys.* (2016) 43:3432–2. doi: 10.1118/1.4956018

32. De Cost AMA, Mittauer KE, Ko HC, Hill PM, Rosenberg SA, Geurts MW, et al. Rapid access palliative radiation workflow using MRI-guided single-session simulation, online adaptation, and treatment. *Int J Radiat Oncol Biol Phys.* (2019) 99:S126 doi: 10.1016/j.ijrobp.2017.06.295

33. Lee YK, Bollet M, Charles-Edwards G, Flower MA, Leach MO, McNair H, et al. Radiotherapy treatment planning of prostate cancer using magnetic resonance imaging alone. *Radiother Oncol.* (2003) 66:203–16. doi: 10.1016/S0167-8140(02)00440-1

34. Henke L, Kashani R, Robinson C, Curcuru A, DeWees T, Bradley J, et al. Phase I trial of stereotactic MR-guided online adaptive radiation therapy (SMART) for the treatment of oligometastatic or unresectable primary malignancies of the abdomen. *Radiother Oncol.* (2018) 126:519–26. doi: 10.1016/j.radonc.2017.11.032

Predictive Values of MRI and PET Derived Quantitative Parameters for Patterns of Failure in Both p16+ and p16– High Risk Head and Neck Cancer

Yue Cao [1,2,3*], Madhava Aryal [1], Pin Li [4], Choonik Lee [1], Matthew Schipper [1,4], Peter G. Hawkins [1], Christina Chapman [1,5], Dawn Owen [1], Aleksandar F. Dragovic [1], Paul Swiecicki [6], Keith Casper [7], Francis Worden [6], Theodore S. Lawrence [1], Avraham Eisbruch [1] and Michelle Mierzwa [1]

[1] Department of Radiation Oncology, University of Michigan, Ann Arbor, MI, United States, [2] Department of Radiology, University of Michigan, Ann Arbor, MI, United States, [3] Department of Biomedical Engineering, University of Michigan, Ann Arbor, MI, United States, [4] Department of Biostatistics, University of Michigan, Ann Arbor, MI, United States, [5] Department of Radiation Oncology, VA Ann Arbor Healthcare System, Ann Arbor, MI, United States, [6] Department of Internal Medicine, University of Michigan, Ann Arbor, MI, United States, [7] Department of Otolaryngology, University of Michigan, Ann Arbor, MI, United States

Purpose: FDG-PET adds to clinical factors, such tumor stage and p16 status, in predicting local (LF), regional (RF), and distant failure (DF) in poor prognosis locally advanced head and neck cancer (HNC) treated with chemoradiation. We hypothesized that MRI-based quantitative imaging (QI) metrics could add to clinical predictors of treatment failure more significantly than FDG-PET metrics.

Materials and methods: Fifty four patients with poor prognosis HNCs who were enrolled in an IRB approved prospective adaptive chemoradiotherapy trial were analyzed. MRI-derived gross tumor volume (GTV), blood volume (BV), and apparent diffusion coefficient (ADC) pre-treatment and mid-treatment (fraction 10), as well as pre-treatment FDG PET metrics, were analyzed in primary and individual nodal tumors. Cox proportional hazards models for prediction of LRF and DF free survival were used to test the additional value of QI metrics over dominant clinical predictors.

Results: The mean ADC pre-RT and its change rate mid-treatment were significantly higher and lower in p16– than p16+ primary tumors, respectively. A Cox model identified that high mean ADC pre-RT had a high hazard for LF and RF in p16– but not p16+ tumors ($p = 0.015$). Most interesting, persisting subvolumes of low BV (TV_{bv}) in primary and nodal tumors mid-treatment had high-risk for DF ($p < 0.05$). Also, total nodal GTV mid-treatment, mean/max SUV of FDG in all nodal tumors, and total nodal TLG were predictive for DF ($p < 0.05$). When including clinical stage (T4/N3) and total nodal GTV in the model, all nodal PET parameters had a p-value of >0.3, and only TV_{bv} of primary tumors had a p-value of 0.06.

Conclusion: MRI-defined biomarkers, especially persisting subvolumes of low BV, add predictive value to clinical variables and compare favorably with FDG-PET imaging markers. MRI could be well-integrated into the radiation therapy workflow for treatment planning, response assessment, and adaptive therapy.

Keywords: MRI, head and neck cancer, radiation therapy, imaging biomarker, adaptive therapy

INTRODUCTION

Locoregional failure (LRF) remains a clinical challenge for poor prognosis locally advanced squamous cell carcinoma of the head and neck (HNSCC) treated with definitive chemoradiation therapy (CRT) (1). It is important to identify imaging markers of LRF that identify patients and tumor subvolumes that may benefit from intensified locoregional therapy in the form of radiation boost, targeted systemic therapy, or surgical intervention.

We and others have been developing prognostic and predictive imaging markers of PET and MRI for LRF, distant metastases, progression free survival (PFS), and overall survival (OS) (2-20). Retrospective studies of pre-treatment FDG-PET that quantify cellular glucose metabolism have identified metabolic tumor volume (MTV), total lesion glycolysis (TLG), and mean/max standard uptake value (SUV) in MTV as prognostic for LRF, PFS, and OS in HNSCC (2–6). Furthermore, FDG-PET has been incorporated into standard of care work-up and follow-up for HNSCC (7, 8). Functional MRI incorporating diffusion and perfusion parameters is an emerging advanced imaging modality in HNSCC. In particular, apparent diffusion coefficient (ADC) correlates with locoregional and distant progression (9–11). Poorly perfused and low oxygenation tumors have been shown to be associated with LRF and worse survival outcomes (12–18).

Despite this progress, it has been difficult to determine which imaging biomarkers should be used to individualize treatment for the patients with locally advanced HNSCC. Most head and neck cancer imaging studies to date include heterogeneous populations of various disease sites, stages, and prognosis. Few imaging studies investigate how p16 status affects imaging parameters pre- and mid-treatment. With respect to ADC in particular, no study to date has evaluated ADC changes during RT for p16+ vs. p16– tumors. A single study investigated ADC differences between HPV+ and HPV– HNSCC, including only 6 HPV+ patients (8%), and found that pre-treatment ADC in HPV+ HNSCC patients was significantly lower than in HPV– patients (19). Furthermore, at the tumor and subtumor level, there is no report on imaging biomarker differences between tumors with local, regional, or distant failure as site of first failure compared to disease free patients. This is an important issue, as it would help stratify the patients for local or systemic intensified or de-intensified therapy. Finally, poorly perfused tumor subvolumes are largely spatially distinct from areas of high FDG uptake and high restricted water diffusion in the same patients, and the spatial correlation between high glucose metabolism and high restricted water diffusion varies greatly

from patient to patient (20, 21). These studies question whether both FDG PET and MRI biomarkers are necessary to guide adaptive RT in HNSCC.

This study aimed to (1) investigate p16+ effects on imaging parameters and their early response rates; (2) assess differences between imaging biomarkers of tumors with local, regional or distant progression and those with no evidence disease (NED), and (3) compare the predictive values of MRI and PET biomarkers. We hypothesized that p16+ status could affect imaging biomarkers and their early response rates, and MRI-based QI metrics could add to clinical predictors of treatment failure more significantly than FDG-PET metrics for local, regional and distant failure.

METHODS

Patients

Imaging analysis was performed on 54 patients [median age of 61 years; 7 females; 31 p16+ (57%)] with advanced HNSCC who were enrolled in a randomized phase II clinical trial between March 2014 and January 2018 (**Table 1**). The trial was approved by the Institutional Review Board of the University of Michigan, including a parallel imaging study to investigate the predictive values of QI metrics for tumor progression. Written consent was obtained from all enrolled patients. Eligibility included patients with p16+ T4/N3 squamous cell carcinoma of oropharynx or locally advanced p16– HNSCC if planned to undergo definitive CRT. All patients were evaluated for p16 status by immunohistochemistry. After completion of CRT, patients were followed up every 2–3 months per standard care for oncologic outcomes as well as toxicity. Tumor recurrences were scored as LF, RF, or DF, or a combination thereof.

MRI and PET Acquisition

Patients underwent FDG-PET/CT scans pre-RT within 4 weeks of RT as a part of standard care. Clinical FDG-PET/CT scans were performed on various PET scanners by following the standard clinical protocol (22).

MRI scans were acquired pre-RT (within 2 weeks) and at fraction 10 (20 Gy) as a part of the protocol. All MRI scans were acquired on a 3T scanner (Skyra, Siemens Healthineers), including anatomic, diffusion weighted (DW), and DCE T1-weighted imaging series. All patients were scanned in the treatment position using an individual-patient immobilization 5-point mask and bite block or aquaplast mold as required for treatment. DW images were acquired with spatial resolution of $\sim$1.2 $\times$ 1.2 $\times$ 4.8 mm and b-values of 50 and 800 s/mm^2 by either a 2D spin-echo single shot echo-planar pulse sequence or

TABLE 1 | Patient characteristics.

Variable	Mean	Range	Median
Age	61	31–85	61
Smoking pack years	37	0–150	30
Gross tumor volume			
Primary tumor (CC)	77	10–595	61
Nodal tumor (CC)	22	0.4–242	6

	Category	Count	Percentage %
Primary tumor site	Nasopharynx	3	5
	Oral cavity	6	11
	Oropharynx	35	65
	Larynx	2	4
	Hypopharynx	6	11
	Nasal sinonasal	2	4
p16	Negative	23	43
	Positive	31	57
Smoking status	Never	6	11
	Former	34	63
	Current	14	26
T stage	1	1	2
	2	2	4
	3	6	11
	4	45	83
N stage	0	6	11
	1	3	6
	2	38	70
	3	7	13
Dose	70 in 35 Fx	34	63
	80 in 35 Fx	20	37
Chemo	Carboplatin	26	48
	Cisplatin	28	52

a readout segmentation of long variable echo-trains (RESOLVE) pulse sequence that reduced geometric distortion (23). Sixty T1-weighted DCE image volumes were acquired using a 3D gradient echo pulse sequence in a sagittal orientation with voxel size ~1.5 × 1.5 × 2.5 mm during an injection of one standard dose of Gd-DTPA. Post-Gd T1-weighted images were acquired in the axial plane with spatial resolution of 0.875 × 0.875 × 3.3 mm by a 2D fast spin echo sequence with fat saturation.

Image Analysis and Registration

Blood volume (BV) maps were quantified from DCE-MRI using the modified Tofts model implemented in an in-house imFIAT Analysis Tool, which was validated using a digital reference object (24). ADC maps were calculated from DW images with b-values of 50 and 800 to mitigate the perfusion effect by using in-house software that was technically validated in a QIN collaborative project (25). Since using the individual-patient immobilization devices reduced gross movement of head and neck during scanning dramatically, BV and ADC maps were reformatted to match voxel-by-voxel of post-Gd T1-weighted images acquired in the same session using coordinates in DICOM

headers. SUV of FDG-PET was calculated. Pre-RT FDG-PET/CT and mid-treatment MR images were co-registered to pre-RT post-Gd T1-weighted images using rigid body transformation and mutual information. Target displacement errors, including image mis-registration and geometric distortion in ADC maps, between image series were assessed and reported previously (20). Reproducibility of BV maps was 16%, which was reported previously (26).

Tumor Volumes and Subvolumes

Gross tumor volume (GTV) of primary and nodal disease was contoured individually on post-Gd T1-weighted images by treating attending head and neck radiation oncologists and reviewed by the trial PI (MM). For this cohort of patients with locally advanced HNSCC, gross cystic or necrotic regions and tumor invasion into blood vessels occurred in many tumors, and therefore were excluded from the GTVs for following analyses of quantitative image (QI) metrics by applying simple thresholds. For the ADC analysis, a threshold of $>2.7 \times 10^{-3}$ mm^2/s (10% below free water diffusion) was used to exclude gross necrosis and blood vessels, and a threshold of $<0.0001 \times 10^{-3}$ mm^2/s was used to exclude air. Then, a low BV subvolume of the GTV (TV$_{BV}$) was created using a threshold of BV <7.64 ml/100 g reported previously based upon a histogram analysis (16). The low ADC subvolume of the GTV (TV$_{ADC}$) was defined as ADC $< 1.2 \times 10^{-3}$ mm^2/s based on an ADC-histogram analysis (20), which is also consistent with the mean ADC reported by others (21). A MTV was defined as FDG SUV $>50\%$ of a value averaged over 4 voxels with maximum SUVs (MTV$_{50}$).

Quantitative Imaging Metrics

QI metrics in tumor volumes and their mid-treatment changes were analyzed for prediction of LF, RF, and DF. Tumor volume metrics included GTV, TV$_{BV}$, TV$_{ADC}$, MTV$_{50}$. Mean values of ADC and BV in GTV excluding blood vessels and necrosis, mean and max SUVs in MTV$_{50}$, and TLG of MTV$_{50}$ were calculated for each primary or nodal tumor as well as for all tumors in each patient.

Treatment

The patients were randomized to a standard arm of RT (70 Gy in 35 fractions) or an experimental arm. In the experimental arm, a union of the persisting TV$_{BV}$ pre-RT to 2 weeks and persisting TV$_{ADC}$ pre-RT to 2 weeks received 2.5Gy per fraction for the last 15 of 35 fractions. If the union of persisting subvolumes pre-RT to 2 weeks was <1 cc, the patient was entered into an observation arm and treated by the standard RT (70 Gy in 35 fractions). Patients were planned to receive weekly cisplatin 40 mg/m^2, and patients considered to be cisplatin ineligible were treated with weekly carboplatin AUC2.

Statistical Analysis

First, we assessed the p16 effect on imaging parameters and parameter change rates at 2 weeks compared to pre-RT using the Mann-Whitney U-test. Secondly, we assessed whether MRI and PET biomarkers had similar predictive values for LRF and DF free survival. For the analysis of LRF, most previous

analyses considered either LF, RF, or LRF as an event, of which the model was useful for stratification of the patients but not for stratification of the tumors for intensified adaptive RT. Tumor progression could occur in one or a few treated tumors (primary or nodal tumor) or in none. Therefore, we applied Cox proportional hazards models to individual (primary or nodal) tumors for prediction of failure. The individual tumor failure free rate (ITFFR) was defined from the start of RT to the date of progression of the tested (primary or nodal) tumor. ITFFR times were censored for all tumors from a patient at the earlier of DF, death or last follow-up. Whether primary and nodal tumors can be analyzed together was tested for each imaging parameter. To compare the predictive values of MRI and FDG PET biomarkers, imaging metrics were assessed one at a time in models also including p16 as a co-variable, which is the most important clinical variable for LRF (27–29). Distant failure free survival (DFRS) was defined as the time interval from the start of RT to the date of DF. The Cox models were fitted including a single QI metric and clinical stage T4/N3 vs. other (non-T4/N3) as the sole clinical variable (30–32), and entering one imaging parameter at a time. Each QI metric was summed up or averaged over all nodal tumors for volume-related or intensity-related metrics, respectively. In the DFFS model, patients were censored at the first occurrence of any local or regional failure, death or last follow-up. If there were any significant differences of imaging parameters between p16– and p16+ tumors, we considered an interaction term in the Cox model or an analysis in different Cox models as appropriate. Since multiple comparisons were made, p-values were corrected using false discovery rate (FDR) control, and corrected $p < 0.10$ were considered significant. Finally, we assessed if there were any significant differences of imaging biomarkers between the tumors that never progressed, those that demonstrated local or regional progression, and those that were locoregionally controlled but metastasized distantly. This landmark analysis used outcomes at 18 months as a cutoff. The tumors were excluded from the analysis if the tumor had local or regional progression after 18 months or the tumor had no progression but the follow-up was shorter than 18 months. As the data were not Gaussian distributed, non-parametric tests were used: Kruskal-Wallis test for the three-group comparison and Wilcoxon rank test for the comparison between local or regional failure and NED. The p-values were corrected with FDR control, and <0.1 were considered as significant. Since 37% of the patients received higher doses, we tested the dose effect before performing the proposed analyses.

RESULTS

Treatment Failure

This cohort of 54 patients with locally advanced HNSCC had large primary GTVs with a median value of 60.5 cc (range: 10.2–595.2 cc; SD: 86.8 cc; **Table 1**), which was several times greater than most reported studies (2–6, 9–11, 33). Eleven patients (20%) (3 p16+) have had local recurrence. Nine patients (17%) (2 p16+) have had regional recurrence, including one patient (p16–) who failed regionally at two separate treated lymph node locations, and 2 (1 p16– and 1 p16+) who had RF at

the locations of non-enlarged/non-FDG avid nodes before RT. Fourteen patients (7 p16+) had distant failure with or without local and regional failure. All cases with LF or RF alone were confirmed pathologically, and distant metastases were diagnosed pathologically or by overt radiographic presentation. Twelve patients have died of HNC (3 p16+), and one patient died cancer-free of other causes. For the patients who did not have progression at the time of analysis, median follow-up was 24 months (range: 10–58 months).

Effects of p16 on Imaging Parameters and Change Rates

We found that both baseline ADC and ADC change after radiation were significantly different between p16+ and p16– primary tumors. The p16– primary tumors had significantly greater mean ADCs pre-RT [1.48 ± 0.05(SEM) µm²/ms], and significantly smaller rates of increase after 10 fractions of RT (10.0% ± 1.2%) than p16+ primary tumors (1.34 ± 0.04 µm²/ms and 21.2% ± 3.1%, $p = 0.04$, and $p = 0.009$, respectively). However, there was no significant difference in mean ADC between p16– and p16+ nodal tumors pre-RT or at 2 weeks as well as ADC increased rates ($p > 0.7$), see **Figure 1**. Pre-RT GTVs of p16– primary tumors (75 ± 12.1 cc) as well as change rates at 2 weeks (−16.2% ± 3.9%) were similar to p16+ ones (79.2 ± 18.9 cc, and −16.7% ± 3.3%, respectively). Mean GTVs as well as change rates at 2 weeks for p16– and p16+ nodal tumors were not significantly different ($p > 0.5$), 24.1 ± 8.1 cc and 21.1 ± 5.9 cc of GTVs and −22.4% + 6.7% and −16.5% + 4.5% of change rates for respective p16– and p16+ nodal tumors Also, there was no significant difference in other imaging parameters between p16+ and p16– primary or nodal tumors ($p > 0.1$). Examples of images are shown in **Figure 2**.

Predictive Values of MRI and PET Imaging Parameters for Local and Regional Progression

First, we did not detect significant difference in local and regional control rates between two-dose arms yet so that the patients who received different doses were analyzed together. For prediction of local progression, mean ADC pre-RT of primary tumors was the only parameter found significant in a univariate Cox model. Since there was no significant difference in mean ADC between primary and nodal tumors, we combined primary and nodal tumors in a single model (53 primary tumors and 82 nodal tumors). For prediction of ITFFR, considering the p16 effect on mean ADC of primary tumors, the Cox model included p16 status, pre-RT mean ADC, and the interaction of pre-RT mean ADC and p16 status. We found that p16 had a significant effect on tumor control (HR p16+ vs. p16– of 0.21, $p = 0.005$), pre-RT mean ADC had a significant effect in p16– tumors (HR per 1 SD increase in ADC = 1.9, $p = 0.015$) but no effect in p16+ tumors (HR = 1.0, $p = 1.0$). The interaction between p16 status and ADC was not statistically significant ($p = 0.24$, **Table 2**).

Since QI metrics other than mean ADC were significantly different between primary and nodal tumors ($p < 0.05$), the QI

FIGURE 1 | (A) Mean ADC in p16– and p16+ primary tumors pre-RT and after 10 fractions (2 weeks) of radiation therapy. (B) Mean ADC in p16– and p16+ nodal tumors pre-RT and after 10 fractions (2 weeks) of radiation therapy. (C) Mean ADC change rates in p16– and p16+ primary and nodal tumors after 10 fractions of radiation therapy compared to pre-treatment. $*p < 0.05$, $**p < 0.01$.
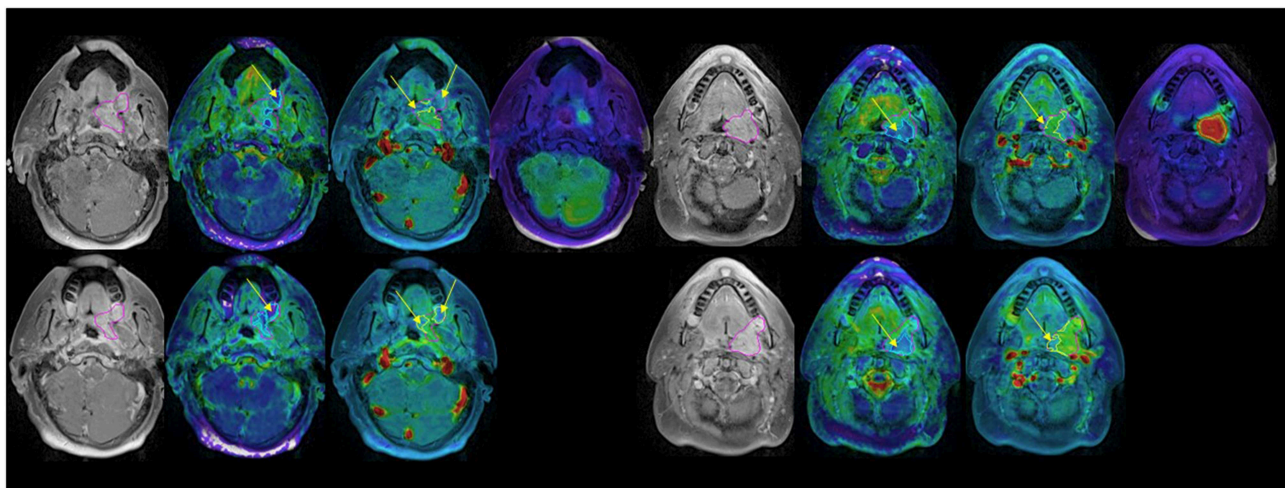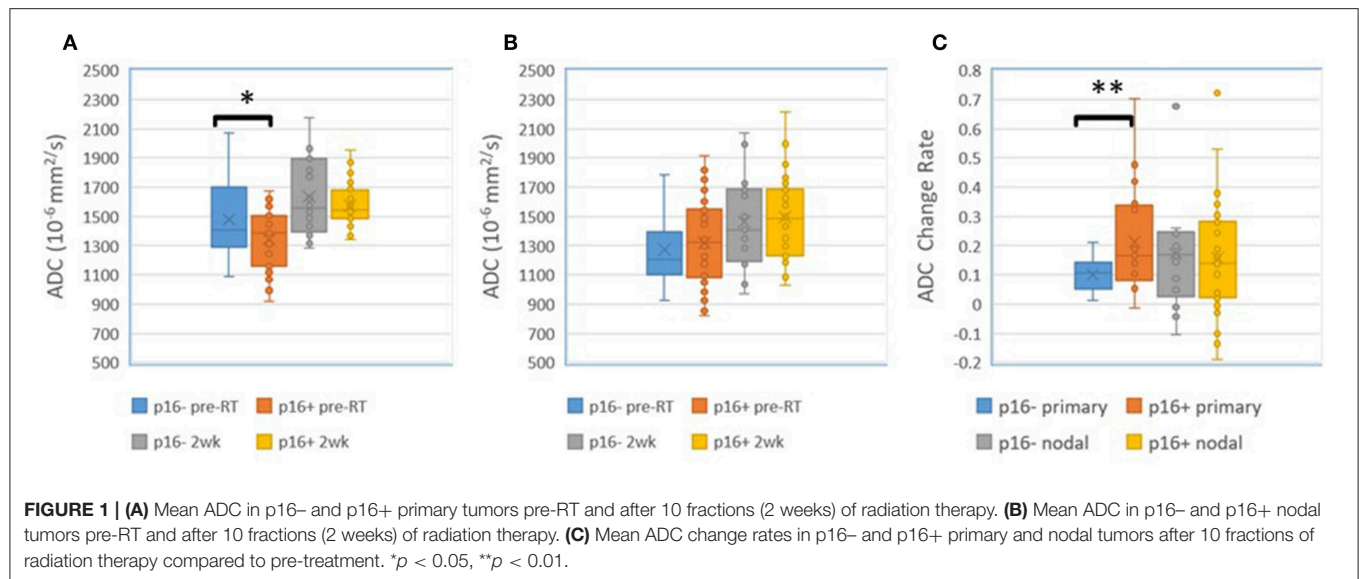

FIGURE 2 | Post Gd T1 weighted images (left), ADC (second left), blood volume (second right), and SUV (right) FDG PET pre-RT (top) and after 10 fractions of radiation therapy (bottom). GTV: magenta; low BV subvolume: yellow; low ADC subvolume: cyan; MTV: red. Note persistent low BV and low ADC subvolumes after 10 fractions of radiation therapy (yellow arrows).

metrics of nodal tumors were tested separately for prediction of regional failure free rates. In Cox models of 82 nodal tumors with p16 status as a co-variate, GTV pre-RT and at 2 weeks, TV_{BV} at 2 weeks, mean and max SUV in MTV_{50} pre-RT, MTV_{50} pre-RT, TLG pre-RT, and change in GTV at 2 weeks vs. pre-RT were significant with $p < 0.07$ with FDR control, see **Table 3**. It is interesting to note that GTV pre-RT and at 2 weeks as well as mean SUV and TLG pre-RT have the highest c-index (> 0.9). However, MTV_{50} and TLG as well as TV_{BV} were strongly correlated with GTV pre-RT (range of r between 0.88 and 0.90), suggesting that these metrics are not independent of GTV. The mean and max SUV in MTV_{50} were strongly correlated each other ($r = 0.98$) but modestly correlated with GTV pre-RT (range of r between 0.65 and 0.67).

Predictive Values of Imaging Biomarkers for Distant Progression

For prediction of distant progression, Cox models identified that TV_{BV} of primary tumors at 2 weeks, total TV_{BV} of all nodal tumors pre-RT and at 2 weeks, total GTV of all nodal tumors at 2 weeks, mean and max SUV of all nodal MTV50 pre-RT, and TLG pre-RT of all nodal tumors had a nominal $p < 0.05$ without FDR. With FDR control, total GTV of all nodal tumors at 2 weeks, mean and max SUV of all nodal MTV50 pre-RT, and TLG pre-RT had a $p < 0.1$, see **Table 4**. We tested whether the significant predictors could provide any complimentary information to clinical stage of T4/N3 and the sum of all nodal GTVs at 2 weeks for prediction of DF, and found that neither total TV_{BV}, nor mean and max SUV, nor total TLG of all nodal tumors had a $p < 0.3$,

TABLE 2 | Cox model of mean ADC effects.

	Coef	HR	Lower 95%	Upper 95%	p-value
p16+ effect*	−1.578	0.206	0.069	0.618	0.005
ADC effect for p16−	0.645	1.905	1.135	3.198	0.015
ADC effect for p16+	0.000	1.000	0.383	2.611	1.0
ADC*p16 effect interaction (ratio of ADC effects between p16+ and p16−)	−0.644	0.525	0.177	1.553	0.2

*At mean ADC value.

TABLE 3 | Cox models for RFFS.

	Hazard ratio	C index	Pr(>|z|)nominal P-value	P-value with FDR adjustment
GTV preRT	1.120	0.90	0.02	0.05*
GTV 2 weeks	1.142	0.91	0.001	0.01*
TV_{BV} 2 weeks	1.514	0.88	0.005	0.02*
MTV_{50} preRT	1.524	0.88	0.03	0.07*
Mean SUV preRT	1.406	0.90	0.005	0.02*
Max SUV preRT	1.238	0.88	0.008	0.02*
TLG preRT	1.073	0.90	0.001	0.01*
Change in GTV	1.288	0.72	0.004	0.02*

82 nodal tumors were included in the analysis. Change in GTV was after 10 fractions of RT compared to preRT. *Indicates significant.

TABLE 4 | Cox models for DFFS.

	Hazard ratio	Pr(>|z|)	P with FDR
PRIMARY TUMOR			
Mean BV 2 weeks	0.797	0.008	0.18
NODAL TUMOR			
Sum of TV_{BV} preRT	1.290	0.05	0.15
Sum of GTVs 2 weeks	1.091	0.01	0.07*
Sum of TV_{BV} 2 weeks	1.290	0.02	0.12
Mean SUV of all MTV_{50} pre	1.363	0.01	0.09*
Max SUV of all MTV_{50} pre	1.233	0.01	0.09*
TLG of all MTV_{50} pre	1.054	0.02	0.09*

T4/N3 was included as a co-variable. *Indicates significant.

and only TV_{BV} of primary tumors at 2 weeks showed marginally significant ($p = 0.06$).

Imaging Biomarkers for Differentiation of Tumors With LF (or RF), DF, and NED

For primary tumors, the subvolumes of low BV pre-RT showed a descending trend from LF, DF, and NED with a marginally significant p-value of <0.06 without FDR control, see **Table 5**. **Figure 3** shows the subvolumes of low BV of primary tumors with LF, DF, and NED pre-RT and at 2 weeks as well as its change rates after 10 factions of RT. Post *ad hoc* analysis showed that the change rates of low BV subvolume were significant smaller in

primary tumors with DF (−0.05% ± 0.16%) than tumors with LF (−0.49 ± 0.08%) and tumors with NED (−0.45 ± 0.09%) with p values of <0.03 and <0.015, respectively.

For nodal tumors, GTV pre-RT and at 2 weeks, the subvolume of low BV pre-RT, mean ADC at 2 weeks, mean BV at 2 weeks, and mean/max SUV of MTV_{50} pre-RT were different among DF, RF, and NED groups with $p < 0.05$ without FDR control and $p \leq 0.1$ with FDR control, see **Table 5**. Again, GTV of nodal tumors was a strongest parameter to differentiate the three groups with different outcomes. Regarding the difference between DF and RF groups, only mean BV values at 2 weeks had a $p < 0.05$ without FDR control but $p > 0.1$ with FDR control. **Figure 4** shows GTVs, the subvolumes of low BV, mean ADC, and mean BV of nodal tumors with RF, DF, and NED pre-RT and at 2 weeks. **Figure 5** shows mean SUV, max SUV, and TLG of nodal tumors with RF, DF, and NED pre-RT.
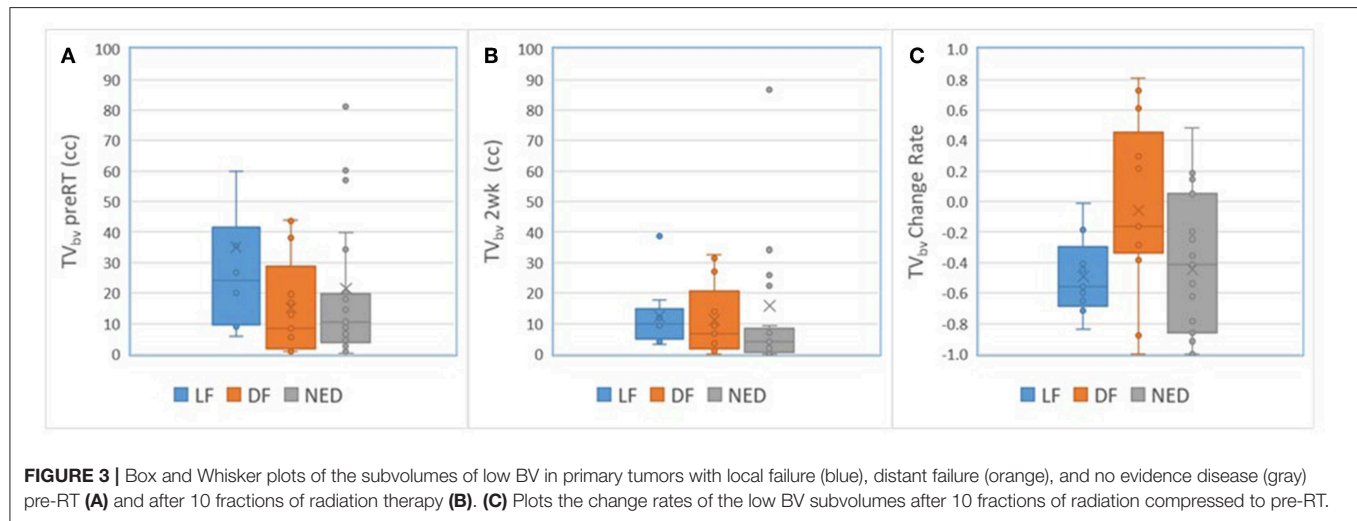
DISCUSSION

In this study, we investigated p16 effects on MRI and PET QI metrics, imaging biomarker differences as a function of tumor control (local, regional, or distant), and the predictive values between MRI and PET biomarkers for tumor progression in locally advanced poor prognosis HN cancers. Our cohort of patients had large tumor volumes compared to previously reported literature (2–6, 9–11, 33). We found the p16− primary tumors had elevated ADC values pre-RT and low early response rates compared to p16+ tumors; the latter of which has not been previously reported. Also, high mean ADC value pre-RT is a hazard for local and regional failure of p16− tumors. Multiple MRI and PET imaging parameters (including GTV, ADC, BV, SUV, and TLG) predicted RF and DF, but the nodal GTV defined on anatomic MRI was the strongest predictor. Most interesting, we report for the first time that the persistent low BV in primary and nodal tumors during the early course of CRT is associated with high-risk for distant failure. In order to identify patients who may benefit from intensified local therapy in the form of a radiation boost or surgical intervention, or from intensified systemic therapy (30, 34), we analyzed the significant imaging predictors found in Cox modeling for differentiation of the tumors that were controlled compared to those with LF, RF, or DF. The performance of MRI related parameters is stronger than PET parameters. Although PET is a part of standard care, MRI could play an important role from treatment planning, to early response assessment, and boost target definition.

We found a p16 effect on ADC and ADC change rates during the early course of RT. The p16− primary tumors had significantly greater mean ADC values pre-RT and smaller increases in ADC after 2 weeks of CRT than p16+ primary tumors. Furthermore, the p16− tumors from patients with local or regional failure had significantly greater mean ADC values pre-RT and mid-treatment than those from disease free patients. These results are consistent with previous reports that the pre-RT high ADC is negatively prognostic for HN cancers (9–11). A recent study shows that ADC is significantly and inversely correlated with cell density but also significantly and positively correlated with

TABLE 5 | Differences of imaging biomarkers among tumors with LF (or RF), DF, and NED.

	Median DF ($n = 12$)	Median LF ($n = 10$)	Median NED ($n = 30$)	KW test P value	KW test P value with FDR	WR test P value	WR test P value with FDR
Primary tumor							
TV$_{BV}$ pre	11.37	24.24	7.10	0.06	0.3	0.08	0.3
	DF ($n = 17$)	LF ($n = 8$)	NED ($n = 57$)	KW test P value	KW test P value with FDR	WR test P value	WR test P value with FDR
Nodal tumor							
GTV pre	8.68	32.14	5.78	0.01*	0.1*	0.07	0.5
GTV 2 weeks	6.57	24.94	3.97	0.02*	0.1*	0.08	0.5
Change in GTV	−1.07	−1.27	−0.44	0.83	0.9	0.8	0.8
TV$_{BV}$ pre	2.77	4.77	1.77	0.02*	0.1*	0.2	0.5
TV$_{BV}$ 2 weeks	1.94	7.40	1.18	0.06	0.1	0.2	0.5
Mean ADC pre	1.38	1.40	1.19	0.06	0.1	0.9	0.9
Mean ADC 2 weeks	1.55	1.64	1.35	0.02*	0.1*	0.4	0.6
Mean BV pre	8.26	10.11	10.21	0.46	0.6	0.5	0.6
Mean BV 2 weeks	8.83	13.33	10.93	0.04*	0.1	0.04*	0.5
Mean SUV of MTV$_{50}$ pre	3.91	5.38	2.35	0.02*	0.1*	0.3	0.5
max SUV of MTV$_{50}$ pre	5.96	7.92	3.50	0.03*	0.1	0.3	0.6
TLG of MTV$_{50}$ pre	0.718	3.245	0.420	0.11	0.2	0.3	0.5

*Tumor volume is a unit of cc. ADC is in unit of 10^{-3} mm^2/s. BV is in unit of (ml/100 g). SUV of FDG is in unit of g/ml. TLG is in unit of 100 g. *Indicates significant.*

FIGURE 3 | Box and Whisker plots of the subvolumes of low BV in primary tumors with local failure (blue), distant failure (orange), and no evidence disease (gray) pre-RT **(A)** and after 10 fractions of radiation therapy **(B)**. **(C)** Plots the change rates of the low BV subvolumes after 10 fractions of radiation compressed to pre-RT.

the percentage area of stroma in laryngeal and hypopharyngeal carcinoma (35). The former finding has been reported previously in animal studies, prostate cancer and lymphomas (36–39), and is related to restricted water diffusion due to high cellularity. The latter finding suggests that a large percentage area of stroma in HN cancers is associated with a high ADC. Stroma has been shown to be negatively prognostic in several cancers, to promote tumor growth and invasion, and to potentially protect tumors from delivery of chemotherapy (40–45). ADC behaviors in the p16– tumors could be explained by their increased stroma. HPV-related oropharynx cancers are histologically basaloid in

histology with significant tumor lymphocytic infiltration, which is associated with improved prognosis (46, 47) and decreased ADC. ADC, although a promising QI metric for differentiation of local and regional failure, and even distant failure, is affected by multiple biologic and physiologic factors, including cell density and stroma as well as cyst and necrosis (In this study, we excluded grossly cystic and necrotic regions for QI metric analysis).

The low BV in primary tumors and persisting during the early course of RT have reported previously to be associated with LF (12–17). However, there is no report that the low BV and its low response rate in HNSCC during the early course
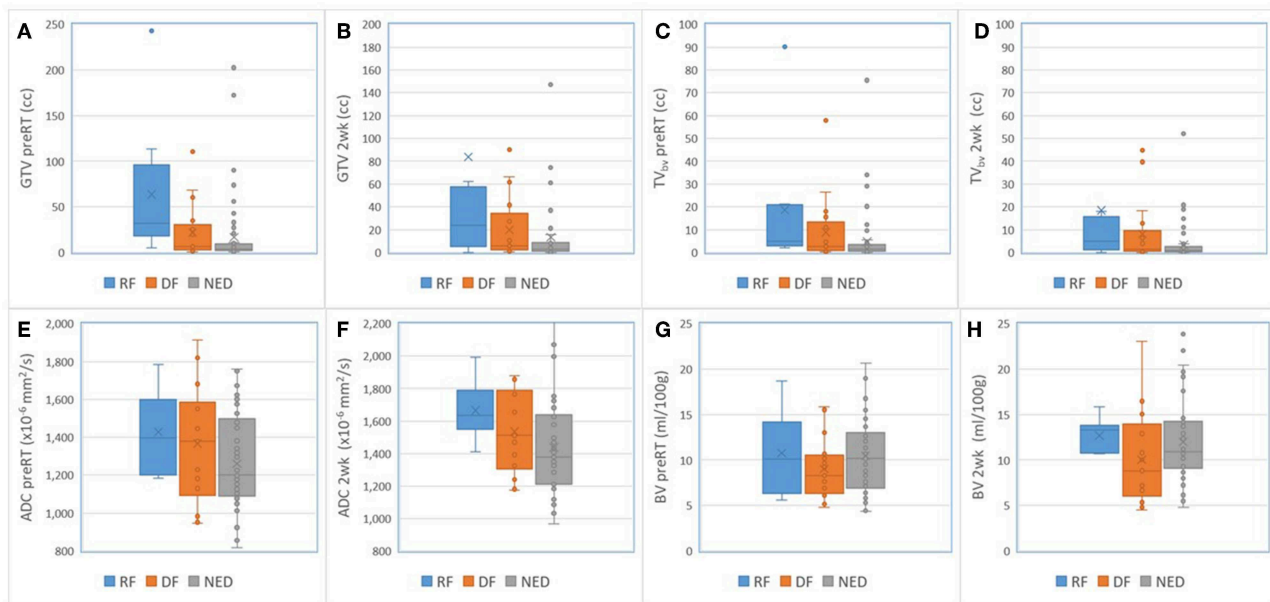
FIGURE 4 | Box and Whisker plots of the GTVs (A,B), the subvolumes of low BV (C,D), mean ADC (E,F), and mean BV (G,H) of nodal tumors with RF (blue), DF (orange), and NED (gray) pre-RT and at 2 weeks.
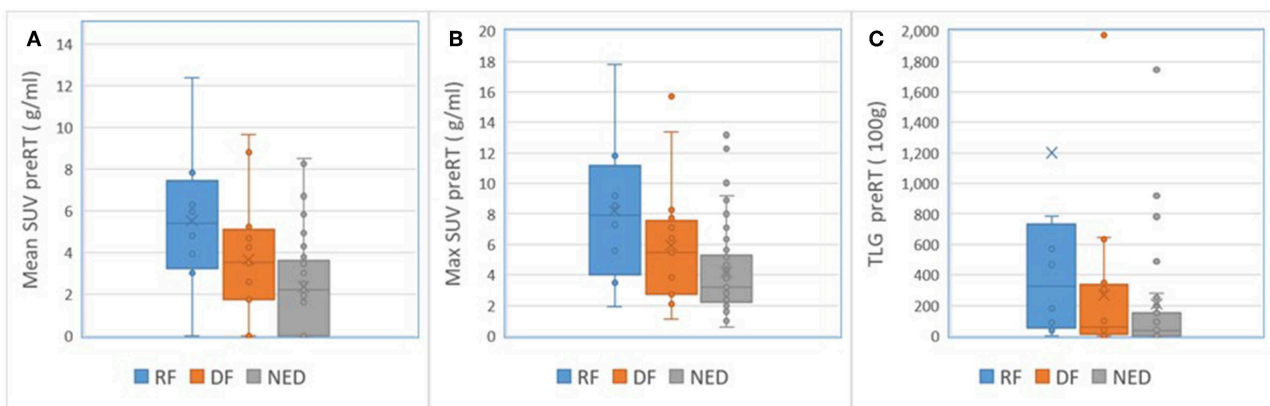

FIGURE 5 | Box and Whisker plots of the mean SUV (A), max SUV (B), and TLG (C) of nodal tumors with RF (blue), DF (orange), and NED (gray) pre-RT.

of RT is associated with DF. The subvolumes of low BV in primary tumors show a descending trend from LF, to DF and NED. The response rate of low BV could be used to differentiate the tumor at high-risk for LF or DF from NED, and thereby adapting intensified local or systematic therapy for the patients with different progression risks.

Pretreatment FDG QI metrics, including MTV, TLG and mean/max, have been reported to be correlated with PFS and OS in the patients with HN cancers treated with CRT (3–6). We found that the high mean/max SUV and large TLG in nodal tumors were risk factors for nodal failure, and that the sum of TLG over all nodal MTVs was a negative prognostic factor for DFFS, which is consistent with several previous reports (2–4). Although TLG accounts for both the size and SUV of MTV, we found that nodal TLG was strongly correlated

with MRI-defined GTV, and nodal GTV was the strongest predictor for RF in our study. For prediction of RF and DF, several other MRI parameters (including GTV, ADC, and BV) perform as well as FDG PET related parameters. When including T4/N3 and total nodal GTV in the Cox model, no other imaging parameters including PET were found to be significant. Finally, there were no FDG PET related parameters that could predict LF.

Radiomics analysis of CT and PET features is another area of imaging analysis that could provide complimentary information to the present study. Radiomics analysis that extracts the large amounts of quantitative textural features from CT, PET, and MRI has been investigated for the prediction of local control, PFS, and OS in head and neck cancers (48–52). Through the feature selection and reduction processes, a small number of

features have been found to have prognostic or predictive value. These features include general categories of statistical energy, shape compactness, gray level non-homogeneity, and gray level non-uniformity. These features may represent different tumor phenotypes. However, it is hard to link the feature to tumor physiology, pathology and biology. Furthermore, radiomics approaches require a large amount of high quality image data, and high-throughput.

A limitation of the present analysis includes RT boost of tumor subvolumes with persistent low BV and low ADC on our clinical trial. This could affect QI metrics that are identified for prediction of treatment failure. We will perform this analysis on patients who are on the standard treatment arm when the trial is completed and the data have matured. Nevertheless, we found that persistent low BV in primary and nodal tumors carries a high-risk for nodal and distant failure, the low response rate of low BV has a high-risk for distant failure, and the low response rate of ADC is for p16– primary tumors. MRI derived biomarkers perform at least as well as FDG PET defined ones. As MRI based planning is already well-integrated into radiation therapy, our findings suggest that MRI based response assessment will be a valuable guide in adaptive radiation therapy.

DATA AVAILABILITY STATEMENT

The image data that were collected in the patients with at least 10 months follow-up in the study are included in the manuscript/supplementary files.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the institute review board of University Michigan. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

All authors contributed significantly for study design, patient enrollment, image acquisition and analysis, statistic analysis, data interpretation, or writing.

FUNDING

REFERENCES

1. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. *CA Cancer J Clin*. (2011) 61:69–90. doi: 10.3322/caac. 20107

2. Chotchutipan T, Rosen BS, Hawkins PG, Lee JY, Saripalli AL, Thakkar D, et al. Volumetric (18) F-FDG-PET parameters as predictors of locoregional failure in low-risk HPV-related oropharyngeal cancer after definitive chemoradiation therapy. *Head Neck*. (2019) 41:366–73. doi: 10.1002/hed. 25505

3. Lin YC, Chen SW, Hsieh TC, Yen KY, Yang SN, Wang YC, et al. Risk stratification of metastatic neck nodes by CT and PET in patients with head and neck cancer receiving definitive radiotherapy. *J Nucl Med*. (2015) 56:183–9. doi: 10.2967/jnumed.114.148023

4. Pollom EL, Song J, Durkee BY, Aggarwal S, Bui T, von Eyben R, et al. Prognostic value of midtreatment FDG-PET in oropharyngeal cancer. *Head Neck*. (2016) 38:1472–8. doi: 10.1002/hed.24454

5. Schwartz DL, Harris J, Yao M, Rosenthal DI, Opanowski A, Levering A, et al. Metabolic tumor volume as a prognostic imaging-based biomarker for head-and-neck cancer: pilot results from Radiation Therapy Oncology Group protocol 0522. *Int J Radiat Oncol Biol Phys*. (2015) 91:721–9. doi: 10.1016/j.ijrobp.2014.12.023

6. Tang C, Murphy JD, Khong B, La TH, Kong C, Fischbein NJ, et al. Validation that metabolic tumor volume predicts outcome in head-and-neck cancer. *Int J Radiat Oncol Biol Phys*. (2012) 83:1514–20. doi: 10.1016/j.ijrobp.2011. 10.023

7. Gupta T, Master Z, Kannan S, Agarwal JP, Ghsoh-Laskar S, Rangarajan V, et al. Diagnostic performance of post-treatment FDG PET or FDG PET/CT imaging in head and neck cancer: a systematic review and meta-analysis. *Eur J Nucl Med Mol Imaging*. (2011) 38:2083–95. doi: 10.1007/s00259-011-1893-y

8. Isles MG, McConkey C, Mehanna HM. A systematic review and meta-analysis of the role of positron emission tomography in the follow up of head and neck squamous cell carcinoma following radiotherapy or chemoradiotherapy. *Clin Otolaryngol*. (2008) 33:210–22. doi: 10.1111/j.1749-4486.2008. 01688.x

9. Hatakenaka M, Nakamura K, Yabuuchi H, Shioyama Y, Matsuo Y, Ohnishi K, et al. Pretreatment apparent diffusion coefficient of the primary

lesion correlates with local failure in head-and-neck cancer treated with chemoradiotherapy or radiotherapy. *Int J Radiat Oncol Biol Phys*. (2011) 81:339–45. doi: 10.1016/j.ijrobp.2010.05.051

10. Kim S, Loevner L, Quon H, Sherman E, Weinstein G, Kilger A, et al. Diffusion-weighted magnetic resonance imaging for predicting and detecting early response to chemoradiation therapy of squamous cell carcinomas of the head and neck. *Clin Cancer Res*. (2009) 15:986–94. doi: 10.1158/1078-0432.CCR-08-1287

11. Lambrecht M, Van Calster B, Vandecaveye V, De Keyzer F, Roebben I, Hermans R, et al. Integrating pretreatment diffusion weighted MRI into a multivariable prognostic model for head and neck squamous cell carcinoma. *Radiother Oncol*. (2014) 110:429–34. doi: 10.1016/j.radonc.2014. 01.004

12. Bernstein JM, Kershaw LE, Withey SB, Lowe NM, Homer JJ, Slevin NJ, et al. Tumor plasma flow determined by dynamic contrast-enhanced MRI predicts response to induction chemotherapy in head and neck cancer. *Oral Oncol*. (2015) 51:508–13. doi: 10.1016/j.oraloncology.2015.01.013

13. Cao Y, Popovtzer A, Li D, Chepeha DB, Moyer JS, Prince ME, et al. Early prediction of outcome in advanced head-and-neck cancer based on tumor blood volume alterations during therapy: a prospective study. *Int J Radiat Oncol Biol Phys*. (2008) 72:1287–90. doi: 10.1016/j.ijrobp.2008.08.024

14. Chawla S, Kim S, Loevner LA, Hwang WT, Weinstein G, Chalian A, et al. Prediction of disease-free survival in patients with squamous cell carcinomas of the head and neck using dynamic contrast-enhanced MR imaging. *AJNR Am J Neuroradiol*. (2011) 32:778–84. doi: 10.3174/ajnr.A2376

15. Hermans R, Meijerink M, Van den Bogaert W, Rijnders A, Weltens C, Lambin P. Tumor perfusion rate determined noninvasively by dynamic computed tomography predicts outcome in head-and-neck cancer after radiotherapy. *Int J Radiat Oncol Biol Phys*. (2003) 57:1351–6. doi: 10.1016/S0360-3016(03)00764-8

16. Wang P, Popovtzer A, Eisbruch A, Cao Y. An approach to identify, from DCE MRI, significant subvolumes of tumors related to outcomes in advanced head-and-neck cancer. *Med Phys*. (2012) 39:5277–85. doi: 10.1118/1. 4737022

17. Brizel DM, Dodge RK, Clough RW, Dewhirst MW. Oxygenation of head and neck cancer: changes during radiotherapy and impact on treatment outcome. *Radiother Oncol*. (1999) 53:113–7. doi: 10.1016/S0167-8140(99) 00102-4

18. Stadler P, Becker A, Feldmann HJ, Hansgen G, Dunst J, Wurschmidt F, et al. Influence of the hypoxic subvolume on the survival of patients with head and neck cancer. *Int J Radiat Oncol Biol Phys.* (1999) 44:749–54. doi: 10.1016/S0360-3016(99)00115-7

19. Driessen JP, van Bemmel AJ, van Kempen PM, Janssen LM, Terhaard CH, Pameijer FA, et al. Correlation of human papillomavirus status with apparent diffusion coefficient of diffusion-weighted MRI in head and neck squamous cell carcinomas. *Head Neck.* (2016) 38 (Suppl 1):E613–8. doi: 10.1002/hed.24051

20. Teng F, Aryal M, Lee J, Lee C, Shen X, Hawkins PG, et al. Adaptive boost target definition in high-risk head and neck cancer based on multi-imaging risk biomarkers. *Int J Radiat Oncol Biol Phys.* (2018) 102:969–77. doi: 10.1016/j.ijrobp.2017.12.269

21. Houweling AC, Wolf AL, Vogel WV, Hamming-Vrieze O, van Vliet-Vroegindeweij C, van de Kamer JB, et al. FDG-PET and diffusion-weighted MRI in head-and-neck cancer patients: implications for dose painting. *Radiother Oncol.* (2013) 106:250–4. doi: 10.1016/j.radonc.2013.01.003

22. Daisne JF, Sibomana M, Bol A, Doumont T, Lonneux M, Gregoire V. Tri-dimensional automatic segmentation of PET volumes based on measured source-to-background ratios: influence of reconstruction algorithms. *Radiother Oncol.* (2003) 69:247–50. doi: 10.1016/S0167-8140(03)00270-6

23. Zhao M, Liu Z, Sha Y, Wang S, Ye X, Pan Y, et al. Readout-segmented echo-planar imaging in the evaluation of sinonasal lesions: a comprehensive comparison of image quality in single-shot echo-planar imaging. *Magn Reson Imaging.* (2016) 34:166–72. doi: 10.1016/j.mri.2015.10.010

24. Cao Y, Li D, Shen Z, Normolle D. Sensitivity of quantitative metrics derived from DCE MRI and a pharmacokinetic model to image quality and acquisition parameters. *Acad Radiol.* (2010) 17:468–78. doi: 10.1016/j.acra.2009.10.021

25. Newitt DC, Malyarenko D, Chenevert TL, Quarles CC, Bell L, Fedorov A, et al. Multisite concordance of apparent diffusion coefficient measurements across the NCI Quantitative Imaging Network. *J Med Imaging.* (2018) 5:011003. doi: 10.1117/1.JMI.5.1.011003

26. Aryal M, Lee C, Hawkins PG, Chapman C, Eisbruch A, Mierzwa M, et al. Real-time quantitative assessment of accuracy and precision of blood volume derived from DCE MRI in individual patients during a clinical trial. *Tomography.* (2019) 5:61–7. doi: 10.18383/j.tom.2018.00029

27. Chung CH, Zhang Q, Kong CS, Harris J, Fertig EJ, Harari PM, et al. p16 protein expression and human papillomavirus status as prognostic biomarkers of nonoropharyngeal head and neck squamous cell carcinoma. *J Clin Oncol.* (2014) 32:3930–8. doi: 10.1200/JCO.2013.54.5228

28. Lassen P, Lacas B, Pignon JP, Trotti A, Zackrisson B, Zhang Q, et al. Prognostic impact of HPV-associated p16–expression and smoking status on outcomes following radiotherapy for oropharyngeal cancer: the MARCH-HPV project. *Radiother Oncol.* (2018) 126:107–15. doi: 10.1016/j.radonc.2017.10.018

29. Rosenthal DI, Harari PM, Giralt J, Bell D, Raben D, Liu J, et al. Association of human papillomavirus and p16 status with outcomes in the IMCL-9815 phase III registration trial for patients with locoregionally advanced oropharyngeal squamous cell carcinoma of the head and neck treated with radiotherapy with or without cetuximab. *J Clin Oncol.* (2016) 34:1300–8. doi: 10.1200/JCO.2015.62.5970

30. Cohen EE, Karrison TG, Kocherginsky M, Mueller J, Egan R, Huang CH, et al. Phase III randomized trial of induction chemotherapy in patients with N2 or N3 locally advanced head and neck cancer. *J Clin Oncol.* (2014) 32:2735–43. doi: 10.1200/JCO.2013.54.6309

31. Hawkins PG, Mierzwa ML, Bellile E, Jackson WC, Malloy KM, Chinn SB, et al. Impact of American Joint Committee on Cancer Eighth Edition clinical stage and smoking history on oncologic outcomes in human papillomavirus-associated oropharyngeal squamous cell carcinoma. *Head Neck.* (2019) 41:857–64. doi: 10.1002/hed.25336

32. Spector ME, Chinn SB, Bellile E, Gallagher KK, Ibrahim M, Vainshtein J, et al. Matted nodes as a predictor of distant metastasis in advanced-stage III/IV oropharyngeal squamous cell carcinoma. *Head Neck.* (2016) 38:184–90. doi: 10.1002/hed.23882

33. Martens RM, Noij DP, Koopman T, Zwezerijnen B, Heymans M, de Jong MC, et al. Predictive value of quantitative diffusion-weighted imaging and 18-F-FDG-PET in head and neck squamous cell carcinoma treated by (chemo)radiotherapy. *Eur J Radiol.* (2019) 113:39–50. doi: 10.1016/j.ejrad.2019.01.031

34. Lin TA, Garden AS, Elhalawani H, Elgohari B, Jethanandani A, Ng SP, et al. Radiographic retropharyngeal lymph node involvement in human papillomavirus-associated oropharyngeal carcinoma: patterns of involvement and impact on patient outcomes. *Cancer.* (2019) 125:1536–46. doi: 10.1002/cncr.31944

35. Driessen JP, Caldas-Magalhaes J, Janssen LM, Pameijer FA, Kooij N, Terhaard CH, et al. Diffusion-weighted MR imaging in laryngeal and hypopharyngeal carcinoma: association between apparent diffusion coefficient and histologic findings. *Radiology.* (2014) 272:456–63. doi: 10.1148/radiol.14131173

36. Barajas RF Jr, Rubenstein JL, Chang JS, Hwang J, Cha S. Diffusion-weighted MR imaging derived apparent diffusion coefficient is predictive of clinical outcome in primary central nervous system lymphoma. *AJNR Am J Neuroradiol.* (2010) 31:60–6. doi: 10.3174/ajnr.A1750

37. Gibbs P, Liney GP, Pickles MD, Zelhof B, Rodrigues G, Turnbull LW. Correlation of ADC and T2 measurements with cell density in prostate cancer at 3.0 Tesla. *Invest Radiol.* (2009) 44:572–6. doi: 10.1097/RLI.0b013e3181b4c10e

38. Wang XZ, Wang B, Gao ZQ, Liu JG, Liu ZQ, Niu QL, et al. Diffusion-weighted imaging of prostate cancer: correlation between apparent diffusion coefficient values and tumor proliferation. *J Magn Reson Imaging.* (2009) 29:1360–6. doi: 10.1002/jmri.21797

39. Zelhof B, Pickles M, Liney G, Gibbs P, Rodrigues G, Kraus S, et al. Correlation of diffusion-weighted magnetic resonance data with cellularity in prostate cancer. *BJU Int.* (2009) 103:883–8. doi: 10.1111/j.1464-410X.2008.08130.x

40. Neesse A, Michl P, Frese KK, Feig C, Cook N, Jacobetz MA, et al. Stromal biology and therapy in pancreatic cancer. *Gut.* (2011) 60:861–8. doi: 10.1136/gut.2010.226092

41. Mueller MM, Fusenig NE. Friends or foes - bipolar effects of the tumour stroma in cancer. *Nat Rev Cancer.* (2004) 4:839–49. doi: 10.1038/nrc1477

42. Wang K, Ma W, Wang J, Yu L, Zhang X, Wang Z, et al. Tumor-stroma ratio is an independent predictor for survival in esophageal squamous cell carcinoma. *J Thorac Oncol.* (2012) 7:1457–61. doi: 10.1097/JTO.0b013e318260dfe8

43. de Kruijf EM, van Nes JG, van de Velde CJ, Putter H, Smit VT, Liefers GJ, et al. Tumor-stroma ratio in the primary tumor is a prognostic factor in early breast cancer patients, especially in triple-negative carcinoma patients. *Breast Cancer Res Treat.* (2011) 125:687–96. doi: 10.1007/s10549-010-0855-6

44. Huijbers A, Tollenaar RA, v Pelt GW, Zeestraten EC, Dutton S, McConkey CC, et al. The proportion of tumor-stroma as a strong prognosticator for stage II and III colon cancer patients: validation in the VICTOR trial. *Ann Oncol.* (2013) 24:179–85. doi: 10.1093/annonc/mds246

45. Wiseman BS, Werb Z. Stromal effects on mammary gland development and breast cancer. *Science.* (2002) 296:1046–9. doi: 10.1126/science.1067431

46. Ward MJ, Thirdborough SM, Mellows T, Riley C, Harris S, Suchak K, et al. Tumour-infiltrating lymphocytes predict for outcome in HPV-positive oropharyngeal cancer. *Br J Cancer.* (2014) 110:489–500. doi: 10.1038/bjc.2013.639

47. Chakravarthy A, Henderson S, Thirdborough SM, Ottensmeier CH, Su X, Lechner M, et al. Human papillomavirus drives tumor development throughout the head and neck: improved prognosis is associated with an immune response largely restricted to the oropharynx. *J Clin Oncol.* (2016) 34:4132–41. doi: 10.1200/JCO.2016.68.2955

48. Bogowicz M, Leijenaar RTH, Tanadini-Lang S, Riesterer O, Pruschy M, Studer G, et al. Post-radiochemotherapy PET radiomics in head and neck cancer - The influence of radiomics implementation on the reproducibility of local control tumor models. *Radiother Oncol.* (2017) 125:385–91. doi: 10.1016/j.radonc.2017.10.023

49. Ou D, Blanchard P, Rosellini S, Levy A, Nguyen F, Leijenaar RTH, et al. Predictive and prognostic value of CT based radiomics signature in locally advanced head and neck cancers patients treated with concurrent chemoradiotherapy or bioradiotherapy and its added

value to Human Papillomavirus status. *Oral Oncol.* (2017) 71:150–5. doi: 10.1016/j.oraloncology.2017.06.015

50. 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

51. Jansen JF, Lu Y, Gupta G, Lee NY, Stambuk HE, Mazaheri Y, et al. Texture analysis on parametric maps derived from dynamic contrast-enhanced magnetic resonance imaging in head and neck cancer. *World J Radiol.* (2016) 8:90–7. doi: 10.4329/wjr.v8.i1.90

52. Zhang H, Graham CM, Elci O, Griswold ME, Zhang X, Khan MA, et al. Locally advanced squamous cell carcinoma of the head and neck: CT texture and histogram analysis allow independent prediction of overall survival in patients treated with induction chemotherapy. *Radiology.* (2013) 269:801–9. doi: 10.1148/radiol.13130110

Detection of Dominant Intra-prostatic Lesions in Patients With Prostate Cancer Using an Artificial Neural Network and MR Multi-modal Radiomics Analysis

Hassan Bagher-Ebadian [1], Branislava Janic [1], Chang Liu [1], Milan Pantelic [2], David Hearshen [2], Mohamed Elshaikh [1], Benjamin Movsas [1], Indrin J. Chetty [1] and Ning Wen [1]*

[1] Department of Radiation Oncology, Henry Ford Health System, Detroit, MI, United States, [2] Department of Radiology, Henry Ford Health System, Detroit, MI, United States

Purpose: The aim of this study was to identify and rank discriminant radiomics features extracted from MR multi-modal images to construct an adaptive model for characterization of Dominant Intra-prostatic Lesions (DILs) from normal prostatic gland tissues (NT).

Methods and Materials: Two cohorts were retrospectively studied: Group A consisted of 98 patients and Group B 19 patients. Two image modalities were acquired using a 3.0T MR scanner: Axial T2 Weighted (T2W) and axial diffusion weighted (DW) imaging. A linear regression method was used to construct apparent diffusion coefficient (ADC) maps from DW images. DILs and the NT in the mirrored location were drawn on each modality. One hundred and sixty-eight radiomics features were extracted from DILs and NT. A Partial-Least-Squares-Correlation (PLSC) with one-way ANOVA along with bootstrapping ratio techniques were recruited to identify and rank the most discriminant latent variables. An artificial neural network (ANN) was constructed based on the optimal latent variable feature to classify the DILs and NTs. Nineteen patients were randomly chosen to test the contour variability effect on the radiomics analysis and the performance of the ANN. Finally, the trained ANN and a two dimension (2D) convolutional sampling method were combined and used to estimate DIL-NT probability map for two test cases.

Results: Among 168 radiomics-based latent variables, only the first four variables of each modality in the PLSC space were found to be significantly different between the DILs and NTs. Area Under Receiver Operating Characteristic (AUROC), Positive Predictive and Negative Predictive values (PPV and NPV) for the conventional method were 94%, 0.95, and 0.92, respectively. When the feature vector was randomly permuted 10,000 times, a very strong permutation-invariant efficiency ($p < 0.0001$) was achieved. The radiomic-based latent variables of the NTs and DILs showed no statistically significant differences ($F_{statistic} < Fc = 4.11$ with Confidence Level of 95% for

all 8 variables) against contour variability. Dice coefficients between DIL-NT probability map and physician contours for the two test cases were 0.82 and 0.71, respectively.

Conclusion: This study demonstrates the high performance of combining radiomics information extracted from multimodal MR information such as T2WI and ADC maps, and adaptive models to detect DILs in patients with PCa.

INTRODUCTION

Radiation Therapy (RT) has been proven to be an effective form of treatment for prostate cancer (PCa) and still is considered as one of the standard treatment options available. The current practice is to treat the entire prostate with a homogeneous dose distribution (1, 2). Escalated dose conformal radiotherapy has shown an advantage in biochemical progression-free survival but it is associated with the increase in acute and late toxicities (3). Simultaneous dose escalation to the dominant intra-prostatic lesions (DILs), while maintaining acceptable doses to the whole prostate gland has potential to improve therapeutic ratio for prostate cancer patients. A median dose to the entire gland could prevent the disease recurrence in the prostate from satellite tumors and significantly reduce the side effects associated with escalated radiation dose to the entire gland. A boosting dose to the DIL can maintain the effectiveness of focal therapy to treat the DIL that is the main determinant for tumor progression and prognosis. For this strategy to be successful, key requirements are the ability to accurately and reliably identify clinically significant tumors in the prostate gland.

Among different imaging techniques, Magnetic Resonance Imaging (MRI) is used increasingly and provides clinicians and researchers with useful information for delineation of the prostate gland and clinically significant tumors in PCa patients (1, 2, 4). While multi-parametric (MP) MRI is well-established (5, 6) for detection of lesions and for staging of the disease, the sensitivity for small and lower grade lesions as well as spare tumors has been low (7) and MP-MRI has failed to improve the detection accuracy of lesions in the central gland (8). Furthermore, accurate and automatic delineation of DILs from prostate glandular tissue which is not a common practice, still remains a challenge. Radiomics analysis, which is defined as the post-processing for high throughput extraction of textural and intensity-based information from medical images, can play a central role toward detecting biomarkers for diagnosis and/or therapy of patients with cancer (9, 10).

This study aims to identify discriminant radiomics features in the real radiomics-feature space and the latent-variable space (constructed from radiomics features in the space of Partial Least Square Correlation, PLSC) for construction of an adaptive model to classify DILs and NTs. The discriminant feature set in the PLSC latent-variable space can also be used for intra-tumoral segmentation and treatment response evaluation.

METHODS AND MATERIALS
Patient Population, and Pre-processing

A total of hundred-seventeen patients consisted of the following two groups were studied:

Group A: This group consisted of 98 PCa patients collected in Radboud University Nijmegen Medical Centre (11) and evaluated with Computer-Aided Diagnosis (CAD) (12, 13). Each MR study was read and reported by or under the supervision of an expert radiologist (Barentsz), with more than 20 years of experience in prostate MR. The radiologist indicated areas of suspicion with a score per modality using a point marker. If an area was considered likely for cancer a biopsy was performed. All biopsies were performed under MR-guidance and confirmation scans of the biopsy needle *in situ* were made to confirm accurate localization. Biopsy specimen were subsequently graded by a pathologist and the results were used as ground truth. Gleason grade groups for these patients are listed in **Table 1**, GroupA.

All MR studies included T2-weighted (T2W) and diffusion-weighted (DW) imaging. The images were acquired on two

TABLE 1 | Gleason Grade Group and PSA level of PCa patients for the two groups are shown in the table.

Gleason grade group	No. of patients
GROUP A	
1	30
2	39
3	19
4	5
5	5
GROUP B	
1	5
2	7
3	3
4-5	4
PSA level	**No. of patients**
<4	5
4-10	8
10–20	2
>20	4

The PSA levels are not available for group A.

different types of Siemens 3T MR scanners, the MAGNETOM Trio and Skyra. T2W images were acquired using a turbo spin echo sequence and had a resolution of around 0.5 mm in plane and a slice thickness of 3.6 mm. the DWI series were acquired with a single-shot echo planar imaging sequence with a resolution of 2 mm in-plane and 3.6 mm slice thickness and with diffusion-encoding gradients in three directions. Three b-values were acquired [50, 400, and 800 (sec-mm^{-2})], and subsequently, the ADC map was calculated by the scanner software. All images were acquired without an endorectal coil, as per the PI-RADS guidelines for acquisition of prostate MRI (14).

Group B: Consisted of 19 patients (age range: 56–84, mean: 67) collected in our hospital, presented with increased PSA levels, suspicion in MR images, and biopsy-proven localized prostate carcinoma with no prior treatment. PSA and Gleason score of these patients are listed in **Table 1**, GroupB. All patients underwent an MP MRI study. An ultrasound guided needle biopsy was performed to confirm the diagnosis. Among 19 patients, 15 had histopathologically identified cancer in peripheral zone and 4 in the central gland. Two image modalities were acquired from the pelvis of all patients using a 3.0 T MR scanner (Ingenia, Philips Medical System, Best, the Netherlands) using small field of view as follows: Axial T2W Images (T2WI) acquired with Fast-Spin-Echo (TE/TR: 4389/110 ms, Flip Angle: 90° with image resolution of $0.42 \times 0.42 \times 2.4$ mm^3) and axial Diffusion Weighted Images (DWI) with two *b*-values [TE/TR:4000/85 ms, FA:90°, $1.79 \times 1.79 \times 0.56$ mm^3, *b*-values:0 and 1000 (sec-mm^{-2})]. The voxel-wise Apparent Diffusion Coefficient (ADC) map was constructed using two DWIs with two *b*-values. A large field of view transverse T2W sequences was also acquired to access the pelvic bones and lymph nodes. Image registration and lesion contouring was performed on in-house developed software.

Data Contouring and Harmonization

For each patient of group B, a radiologist with over 20 years of experience evaluated the axial T2WI and ADC maps and used the following criteria for delineation of DIL: Areas with a well-circumscribed, hypo-intense with the highest Gleason score in the prostate on T2WI and ADC map. DIL and the equivalent region in contralateral (normal prostatic glandular tissues, NT) were contoured on axial T2WI and ADC maps, respectively. To harmonize the data and make them independent from MR scanner gains (can affect weighted images), for each patient of both groups, the signal intensity of their DIL was normalized to the mean value of their corresponding normal volume prior to the radiomics analysis.

Radiomics Analysis

All data processing was performed off-line using a commercial software package (MATLAB 2016a, the MathWorks Inc., Natick, MA, 2000). For each patient, 168 radiomics features (15), from eight different categories, were extracted from DIL and NT volumes contoured on ADC maps and T2W images. The 8 feature categories (15), as detailed below and in **Table 2**, were classified as follows: Intensity Based Histogram Features (IBHF—9 features), Gray Level Run Length (GLRL—7 features),

Law's Textural information (LAWS—18 features), Discrete Orthonormal Stockwell Transform (DOST—18 features), Local Binary Pattern (LBP—6 features), Two-Dimensional Wavelet Transform (2DWT—48 features), Two Dimensional Gabor Filter (2DGF—40 features), and Gray Level Co-Occurrence Matrix (GLCM—22 features) (15).

Feature Selection and Statistical Analysis

A Partial Least Square Correlation (PLSC) (16) technique combined with one-way analysis of variance (ANOVA) were recruited to identify the most discriminant PLSC latent variables constructed from radiomics features extracted from NTs and DILs of multimodal MR information (T2WI and ADC map). PLSC method which is also called as projection to latent structures, can relate the information present in two MR modalities in which collect measurements on the same set of observations (16, 17). The goal of the PLSC is to find pairs of latent vectors with maximal covariance and with the additional constraints that the pairs of latent vectors made from two different indices are uncorrelated and the coefficients used to compute the latent variables are normalized. As shown in **Figure 1**, two observation matrices were constructed using 168 radiomics features extracted from the two image modalities (T2WI and ADC) from total patients. A singular value decomposition (SVD) technique was used to analyze the common and discriminant information between the two observation matrices. For each MR modality, a latent vector was computed by the SVD technique and then it was tested by the ANOVA (with homoscedasticity assumption and confidence level of 0.95) to identify the most discriminant features in latent variable space between the features extracted from DIL and NT volumes in both groups. The Holm–Bonferroni method (18) was also used for circumventing the problem of multiple comparisons for the *p*-values. This method of *p*-value adjustment controls the familywise error rate and offers a uniform test, which is more powerful than the classic Bonferroni correction (18). Using the discriminant latent variable set identified by ANOVA, an optimal feature set for both modalities was identified and constructed.

Feature Ranking Using Bootstrapping Ratio Technique

A bootstrapping ratio (16, 19, 20) and permutation test (10,000 times randomly repeated) were performed on the latent vectors of the features sets (extracted from T2WI and ADC) and the SVD was computed for each configuration and distribution of eigen values was used to estimate the ranking and efficiency of the radiomics features against random permutation. For radiomics feature ranking, bootstrap ratios were computed by dividing the mean of the bootstrapped distribution of a significant latent variable by its standard deviation. The bootstrap ratio is akin to a Student t criterion and so if a ratio is large enough (>2.00; because it roughly corresponds to 95% of confidence level for a *t*-test) then the variable is considered significant/important for the dimension. The bootstrap estimates a sampling distribution of a statistic by computing multiple instances of this statistic from bootstrapped samples obtained by sampling with replacement from the original sample (16, 19, 20).

TABLE 2 | Eight different radiomics feature categories along with a short explanation of each category is shown in this table.

Category	Number of radiomics features	Radiomics features
IBHF	9 features	Nine features are extracted from histogram of the pixel intensity values: 1-Mean, 2-Standard Deviation, 3-Skewness, 4-Kurtosis, 5-Entropy, 6-Central Moment of 3rd order, 7- Central Moment of 4th Order, 8- Central Moment of 5th Order, 9-Central Moment of 6th Order.
GLRL	7 features	Seven Gray Level Run Length texture descriptors were constructed based on the following emphasizes: Short Run Emphasis (SRE), Long Run Emphasis (LRE), Gray Level Non-Uniformity (GLN), Run Percentage (RP), Run Length Non-Uniformity (RLN), Low Gray Level Run Emphasis (LGRE), and High Gray Level Run Emphasis (HGRE).
LAWS	18 features	Nine textural maps were constructed by filtering the image data using the following convolution kernels: L5 = [1 4 6 4 1], E5 = [−1 −2 0 2 1], S5 = [−1 0 2 0 −1], R5 = [1 −4 6 −4 1] and then, 18 LAWS textural features were computed by applying and combining the energy and entropy operators on these maps as following: L5E5/E5L5, L5R5/R5L5, E5S5/S5E5, S5S5, R5R5, L5S5/S5L5, E5E5, E5R5/R5E5, and S5R5/R5S5.
DOST	18 features	The two-dimensional matrix of DOST coefficients was divided into nine equal segments and the energy and entropy of each segment was averaged over the tumor volume and eighteen features (nine energy along with nine entropy) were generated and used as the DOST radiomics features.
LBP	6 features	Local Binary Pattern algorithm with a radial filter (eight-neighborhood) was used to generate a two-dimensional LBP map and Entropy, Entropy, Mean, Standard Deviation, Skewness, and Kurtosis of the LBP maps were used as the six LBPF radiomics features.
2DWT	48 features	Two-dimensional Wavelet Transform with six decomposition levels for four different information attributes (Multi-resolution image, vertical, horizontal, and diagonal) was used to generate 24 maps of 2DWT information. Energy and entropy of the information maps were calculated and used as the 48 2DWT radiomics features.
2DGF	40 features	Two-dimensional Gabor (2DG) filter with five different scales for four different orientations generated 20 maps. Energy and entropy of the maps was averaged over the tumor volume and used as the 2DGT radiomics features.
GLCM	22 features	Gray-Level-Co-occurrence Matrix (GLCM) was generated and the following 22 features were measured from the GLCM using an 8-bit depth quantization: 1-Autocorrelation, 2-Contrast, 3-Correlation (2), 4-Correlation (1), 5-Cluster Prominence, 6-Cluster Shade, 7-Dissimilarity, 8-Energy, 9-Entropy, 10-Homogeneity (1), 11-Homogeneity (2), 12-Maximum probability, 13-Sum of squares(Variance), 14-Sum average, 15-Sum variance, 16-Sum entropy, 17-Difference variance, 18-Difference entropy, 19-Information measure of correlation (1), 20-Information measure of correlation (2), 21-Inverse difference normalized, and 22-Inverse difference moment normalized.

Artificial Neural Networks: Architecture Optimization, Training, and Validation Strategies

Eight latent variables constructed from the radiomics information were identified as the optimal feature set and were used as the input to an artificial neural networks (ANN) with a feed-forward multilayer perceptron (MLP) architecture and back-propagation training algorithm (21) for classification of DILs and NTs. In this type of ANN, the nodes are organized in multiple layers; The ANN used in our study had three layers: the input layer, single intermediate layer, and the output layer (21, 22). Nodes were interconnected by weights in such a way that information propagates from one layer to the next, passing through a sigmoid (bipolar) activation function (22). Learning rate and momentum factors were set to control the internode weight adjustments during training (learning rate: 0.01, and Momentum: 0.01). A back propagation learning strategy (21) was employed for training the ANN in a supervised mode. In this strategy, a trial set of weights (the weight vectors, one vector for each layer of the ANN) was proposed. The initial weights were assigned randomly, and the same set of initial weights was saved and used for different trial during the leave-one-out method. The weight vectors were then adjusted to minimize some measure of error (in this case the Mean Square Error, MSE) between the output of the ANN and the training set. This

procedure was performed iteratively across the entire data set using a batch processing mode to improve the convergence rate and the stability of training. The weight changes obtained from each training case were accumulated, and the weights updated after the entire set of training cases was evaluated. Batch processing improves stability, but with a tradeoff in reduction of the convergence (21–23).

Two different training and validation strategies were recruited and tested as follows:

Strategy 1: Leave-One-Out Cross-Validation (LOOCV) method, which is a particular case of the Leave-P-Out Cross Validation (called as Exhaustive Method) was employed for training, testing, and ANN architecture optimization (21, 22, 24–26). LOOCV was recruited to find the optimal structure, termination error, and validation of the ANN. As shown in **Figure 2**, this approach leaves one data point out of training data, i.e., if there are N data points in the original sample then, N-1 samples are used to train the model and 1 point is used as the validation. This is repeated for all combinations in which original sample can be separated this way, and then the error is averaged for all trials, to give overall effectiveness with less estimated bias (27). This method is generally preferred over the Leave-P-Out Cross Validation when the sample size is small since it does not suffer from the intensive computation, as number of possible combinations is equal to number of data
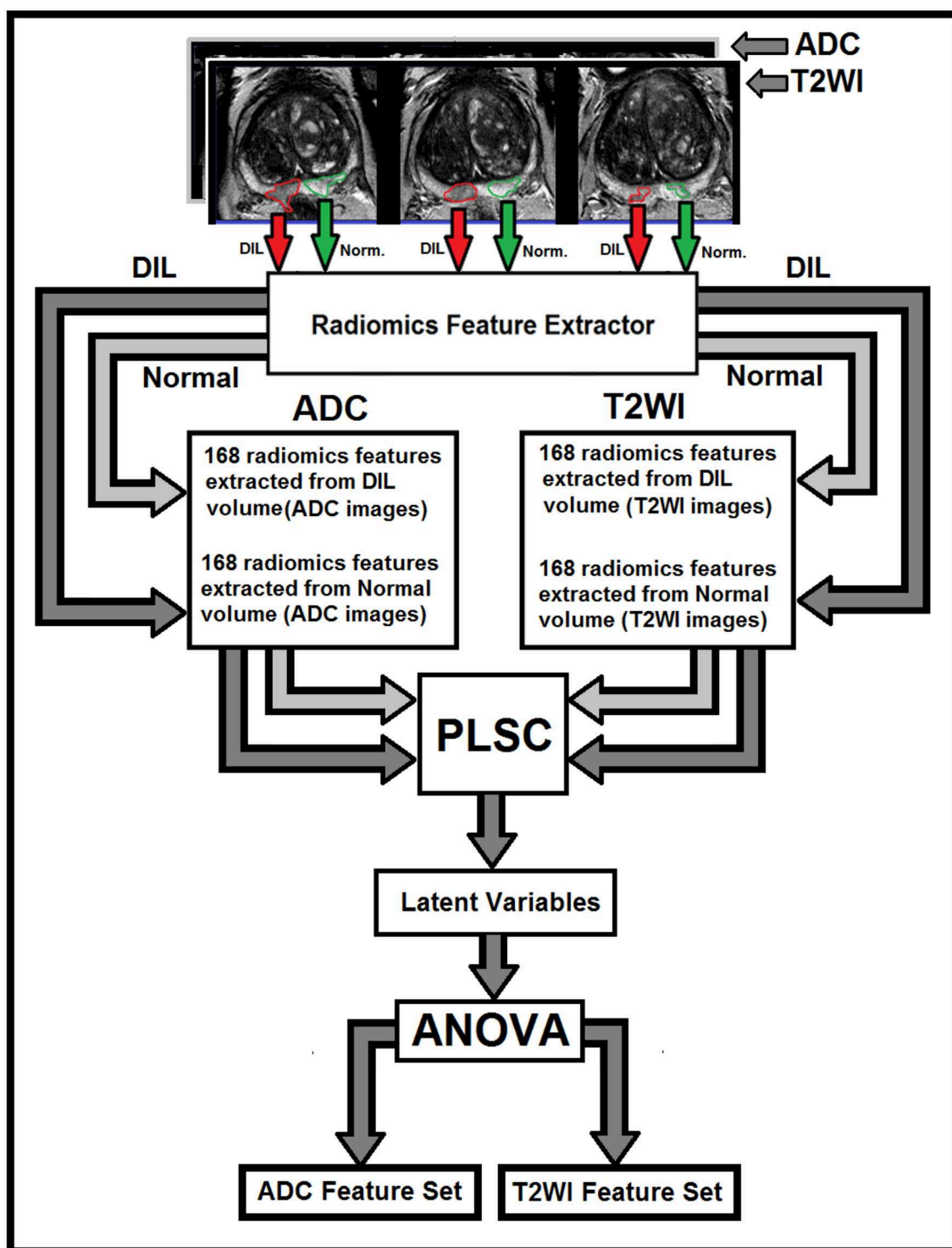
FIGURE 1 | The flowchart demonstrates different steps for the extraction of radiomics features from T2W images and ADC maps for DILs and normal tissues. As shown in this figure, for each MR modality, 168 radiomics features are extracted from normal and DIL volumes. The optimal feature set for the two MR modalities are identified using ANOVA applied on the latent variables generated by the PLSC technique for features with Silhouette coefficient of 0.5 and greater.

points in original sample or N (28). Finally, to evaluate the stability of the optimal ANN against optimal number of training epochs, a series of ROC curves were generated by applying a

threshold at the output of the randomly (100 times) trained ANN. The, the optimal cut-point which is the point closest-to-corner in the ROC plane was calculated. The optimal cut-point
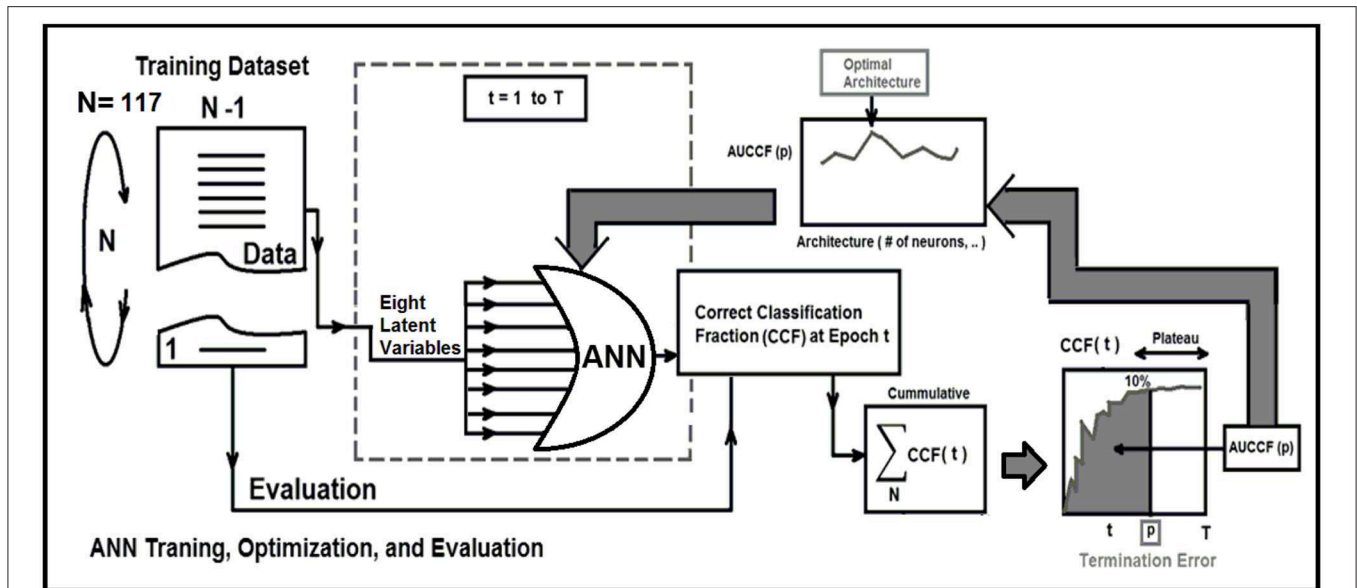
FIGURE 2 | This figure demonstrate three major phases as follows: Training, optimization, and evaluation phases for the ANN using the leave-one-out technique and area under correct classification fraction.

defines as the point minimizing the Euclidean distance between the ROC curve and the (0, 1) point (29). As the sensitivity (true positives) increases, the ANN can identify more cases with DIL, while the accuracy on identifying NTs (specificity) are sacrificed. Cut-points dichotomize the test values, so this provides the classification (DIL or not). Simultaneous assessment of sensitivity and specificity is used to estimate the cut-point value which is considered as optimal when the point classifies most of the individuals correctly (29, 30).

To measure how accurately the ANN matched the whole input dataset with the entire identifier set, the ANN's Correct-Classification-Fraction (CCF: True Positive plus True Negative, TP+TN) curve was generated at different levels of epochs during the LOOCV procedure. The area under Receiver-operating characteristic (AUROC, *Az-value*) curves (21, 22, 24, 25) for the ANN that is an index of predictive performance, was used to compare the ANN's performance in determining the optimal architecture of the ANN, and also finding the termination error (avoid overfitting) for training the optimal ANN.

Strategy 2: For each discriminant latent variable, the data of the patient group *A* (96 patients) was split 100 times into training and validation components. In each data split, two-thirds (67%) of the entire dataset was randomly sampled and used as a training set and the remaining one-third (33%) was used as the unseen cohort or validation dataset (31). Using the training and validation sets for each of the 100 iterations, the ANN was trained and validated separately for each discriminant latent variable. The same procedure was repeated for the set of eight latent variables. The AUROC, Positive Predictive value (PPV) and Negative Predictive value (NPV) were computed for each trial and were averaged to evaluate ANN classification performance for each discriminant latent variable and the set of eight latent variables.

All data processing and classifier implementation were performed using a series of in-house codes developed in the MATLAB environment.

Testing of Data Harmonization, Feature Consistency, and Generalization Error

Data harmonization refers to all efforts to combine different datasets collected by different scanners in different institutions. Finally, in order to test the consistency of the identified discriminant latent variables against the data harmonization and also testing the performance of the classifiers against prospective/unseen datasets (ANN generalization error), the following sub-analysis was conducted: An ANN was trained using the eight discriminant latent variables (constructed from radiomics information) extracted from patients information of group *A*. The trained ANN was then applied on the eight discriminant latent variables (constructed from radiomics information) extracted from patient information of group *B* (as test set or unseen patient cohorts). Ultimately, a ROC analysis was performed on the predictions of the trained ANN and AUROC, NP, and PP values for the unseen testing cohort (group *B*) were calculated.

Contour Variability Test

Nineteen patients were randomly chosen from hundred-seventeen patients and their DIL and NT contours were modified by scaling the contours by a factor of 1.2 in all directions followed by a 1 voxel shift in all directions and their modified contours were used to repeat the radiomics and PLSC analyses and ANOVA method was used to test the sensitivity of the latent variables against contour variability.

Feature no.	Image modality	Radiomics feature	Radiomics feature category	Bootstrapping ratio (mean/std)
1	T2WI	LBP_Energy	LBP	21412.01
2	ADC Map	LBP_Energy	LBP	410.83
3	T2WI	RLN	GLRL	159.69
4	ADC Map	RLN	GLRL	70.32
5	T2WI	GLN	GLRL	35.29
6	T2WI	HGRE	GLRL	35.28
7	ADC Map	DOST_ENTROPY_22	DOST	25.13
8	ADC Map	ENG_GAB02	2DGF	22.99
9	ADC Map	LBP_KURTOSIS	LBP	22.95
10	T2WI	LBP -KURTOSIS	LBP	22.92

Tumor Probability Map

The trained ANN and a two dimension (2D) convolutional sampling method window size = 25 × 25) were combined and used to estimate DIL-NT probability map for two test cases. Dice coefficients between the DIL contours and the DIL patch estimated from the probability maps ($P_{thr} > 0.001$) for the two cases were calculated and compared.

RESULTS

A flowchart demonstrating different steps for extracting radiomics features from T2W images and ADC maps for DILs and NTs are shown in **Figure 1**. As shown in the figure, for each MR modality, 168 radiomics features were extracted from each of the NTs and DILs and finally, the optimal discriminant latent feature set for the two MR modalities were identified using a PLSC technique and ANOVA. **Table 3** shows feature ranking results based on the PLSC and bootstrapping ratio techniques for the first 10 significant radiomic features of two MR modalities. **Figures 3A,B** demonstrate the scatter plots of the first three PLSC latent variables for T2WI and ADC, respectively. **Figures 3C,D** demonstrate the permutation tests for the inertia explained by the PLSC of the T2WI and ADC map along with their observed inertia for the 10,000 permutations.

Figure 4A shows correct classification fraction (CCF = TP + TN) of the optimal ANN at different training epochs for LOOCV technique. The epoch corresponding to 10% change in plateau for the optimum architecture (8:5:1) was used as the stopping epoch (epoch = 17) of the ANN. **Figure 4B** shows TP, TN, false positive (FP), and false negative (FN), of the optimal ANN at different training epochs.

The AUCCF values for different ANN structures for LOOCV technique are shown in **Figure 4C**. As shown in this figure, the ANN with five neurons in its only hidden layer shows the highest performance ($A_z = 0.95$) and is chosen as the ANN with optimal structure. **Figure 4D** shows the average AUROC of the ANN generated for randomly (100 times) trained ANNs along with the optimal cut-point (OCP = 0.96). Given the average AUROC (A_z test ~ 0.96), the optimal cut-point of the ANN, and the

eigen value distributions for the randomly permuted (10,000 permutations) radiomics features, the generalization error of the ANN was about 4% with a very strong permutation-invariant efficiency, $p < 0.0001$) against the order of the latent variables.

AUROC, PPV, and NPV for the conventional method were 94%, 0.95, and 0.92, respectively. ROC analyses for eight individual latent variables (4 for T2WI and 4 for ADC) are shown in **Figure 5**. **Figures 5A–D** demonstrate the ROC analyses of the ANN for the first 4 latent variables constructed from T2WI for 100 random iteration corresponding to a different division of training and validation data of group *A* while **Figures 5E–H** depict the corresponding information for the ADC map. **Table 4** shows AUROC, NPV, and PPV values along with their confidence intervals measured for each individual latent variable for 100 iterations (each corresponding to a different division of training and validation datasets).

As shown in **Figure 5I**, for the conventional training and validation method, the average AUROC, PPV and NPV were 95%, 0.96, and 0.93, respectively. **Figure 5J** shows the response of the trained ANN (group *A*) when it was applied on group B. The performance of the trained ANN (using group *A* dataset) when it was applied on the unseen data cohort (group *B*) was: Sensitivity/Specificity = 0.95/0.94. The radiomic-based latent variables of the NTs and DILs showed no statistically significant differences ($F_{statistic}$ for all 8 latent variables were smaller than $F_{critical} = 4.11$, with Confidence Level of 95%) against contour variability. **Figures 6A–F**, illustrate T2WI, ADC map, and lesion probability map for a slice of prostate gland of two different patients estimated by the trained ANN using a 2D-convolutional sampling method (window size = 25 × 25). Dice coefficients between DIL-NT probability map and physician contours for the two test cases were 0.82 and 0.71, respectively.

DISCUSSION

Recent studies have shown that cancerous tissues are spatially heterogeneous due to factors, such as cell structures, genes, protein contents, cell morphologies, tumor microenvironment, and physiology (32). Indeed, the main purpose of using radiomics is to reveal and extract additional information from medical imaging modalities, associated with *macroscopic* and *microscopic* image-based features that have the potential to serve as surrogates for pathophysiological and radiological parameters, such as tumor heterogeneity level, pathology, response to a given therapy, decoration and distribution of information in images, and structural and image-based patterns in digital images. In our study, given the variation and nature of the radiomics features, we extracted multi scale information in form of features from the prostate gland to characterize normal prostatic tissue and tumor phenotypes from multi model MRI.

The PLSC technique used in this study allowed the finding of shared information between the two image modalities (T2WI and ADC). This approach is equivalent to a correlation problem (16, 17, 33) and provided descriptive features from multivariate information in form of latent variables which are optimal linear combinations of the variables extracted
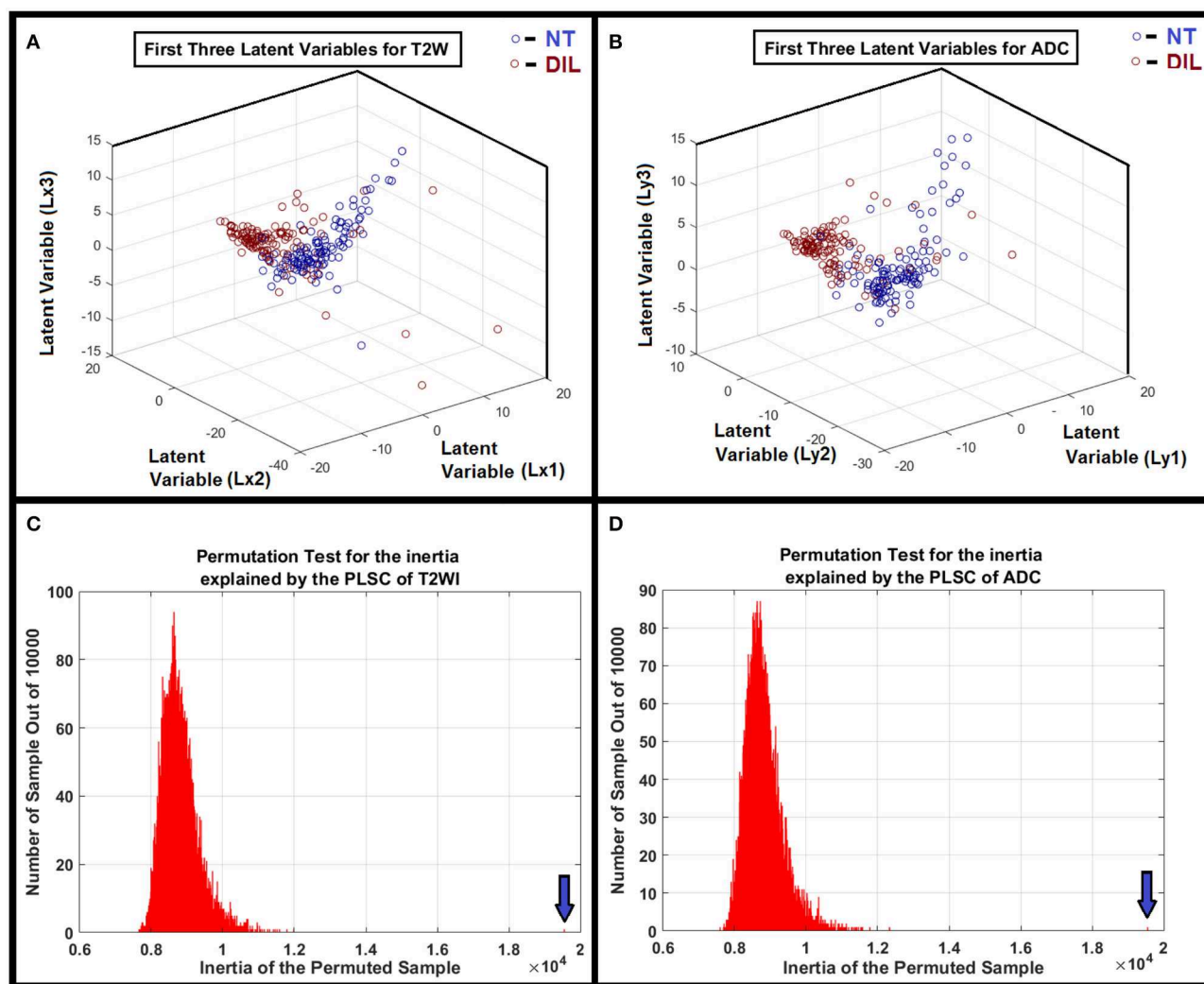
FIGURE 3 | (A,B) Clusters of NTs and DILs for each latent variable are well-separated with less diffusivity. It confirms that the distribution of the identified latent variable (PLSC-ANOVA) in the feature space is well-matched to its own cluster (less scattered) and poorly diffused to its neighboring clusters for the MR modalities. **(C,D)** Show the results of the permutation tests for the inertia explained by the PLSC of T2WI and ADC map for 10,000 permutations. As shown in the subfigures, the observed value (shown by vertical arrows) were never obtained in the 10,000 permutations for both modalities. Therefore, it is concluded that PLSC extracted a significant amount of common variance between these two modalities with $P < 0.0001$.

from the two image modalities. Partial least square (PLS) method that benefits from projecting feature information on latent structures, relates the information present in two data tables (modalities) that collect measurements on the same set of observations (16). PLSC latent variables constructed on the basis of radiomics information extracted from DIL and NT consists of all radiomics features and can help reveal variations of descriptive features or discriminant parameters for classification of DIL from NT. An adaptive classifier (such as ANN) provides capability of implicitly detecting complex non-linear relationships between dependent and independent radiomics variables (already found as optimal feature set in latent variable space) and their variations, modeling their non-linear changes as well as detecting all possible interactions between

the predictor variables. As shown in **Figures 3A,B**, clusters of NTs and DILs for each latent variable are well-separated with less diffused marginal points in the feature space. It confirms that the distribution of the identified latent variable (PLSC-ANOVA) in the PLSC space is well-matched to its own cluster (less scattered) and poorly diffused to its neighboring clusters. **Figures 3C,D** show the results of the permutation tests for the inertia explained by the PLSC of T2WI and ADC map for 10,000 permutations. The observed value (shown by vertical arrows) were never obtained in the 10,000 permutations for both modalities. Therefore, it is concluded that PLSC technique was able to successfully extract significant amount of common variances between these two modalities with p-value smaller than 0.0001.
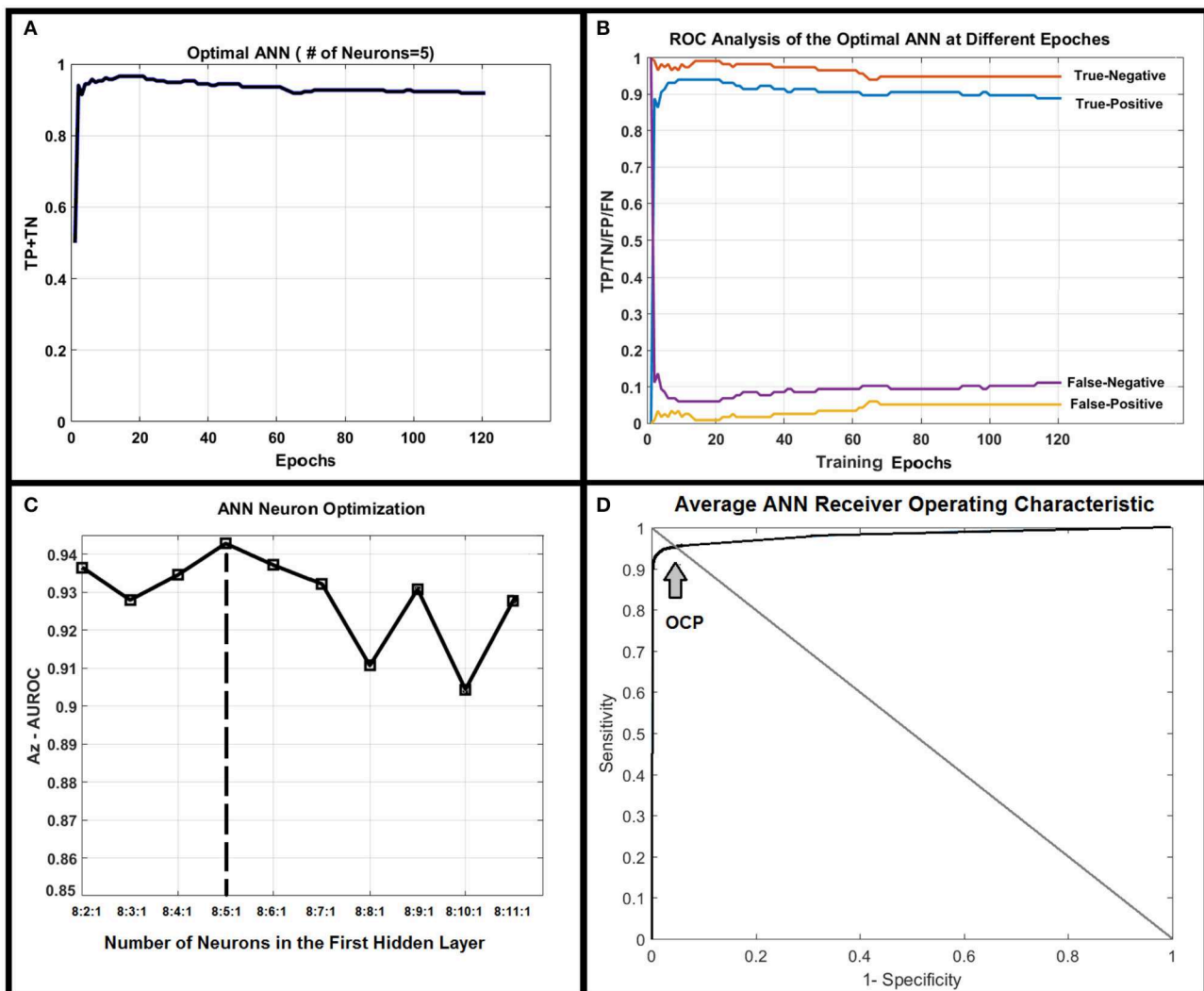
FIGURE 4 | (A) Shows true positive plus true negative (TP+TN) of the optimal ANN (8:5:1) at different training epochs. **(B)** Shows true positive, true negative, false positive, and false negative of the optimal ANN at different training epochs. **(C)** Demonstrates the area under receiver operating characteristic (AUROC, *Az* test) value for different ANN structures. As shown in this figure, the ANN with five neurons in its only hidden layer shows the highest performance and is chosen as the optimal ANN. **(D)** Shows the average ROC of the optimal ANN along with optimal-cut-point of the ANN.

Recruitment of PLSC technique and ANOVA in this study allowed robust comparison and revealing of the correlation and descriptive power of different radiomics features extracted from the two MR modalities, while providing more predictive accuracy and a much lower chance of risk for the two sets of features affecting each other. The major limitations could be the sensitivity to the relative scaling of the descriptor variables that was addressed by the standardization and harmonization steps prior to the feature extraction.

Recent studies (34–39) have shown that ADC measurements are affected by the user selected repetition time (T_R) values, especially if it is comparable to the relaxation time. The degree of T_R dependence is also codependent on another parameter

called number of diffusion preparation pulses. Similar to T_R dependence of ADC values, it is expected that there could be an echo time (T_E) dependence on ADC values. In fact, Wang et al. (39) found a modest correlation between T_E and ADC values in the prostate. It has been shown that tissue specific relaxation time parameters such as T1 and T2 and imaging parameters such as T_R and T_E affects the optimum b-value for different anatomies, tissues, and even lesion types within the same organ. Therefore, since the ADC value could be highly and "*non-linearly*" affected by the MR imaging parameters (34–39), in this study, as part of data harmonization, normalization to normal volume was performed to suppress the effect of the MR imaging parameters on the ADC values. Such normalization made the ANN independent and less sensitive to the MR imaging
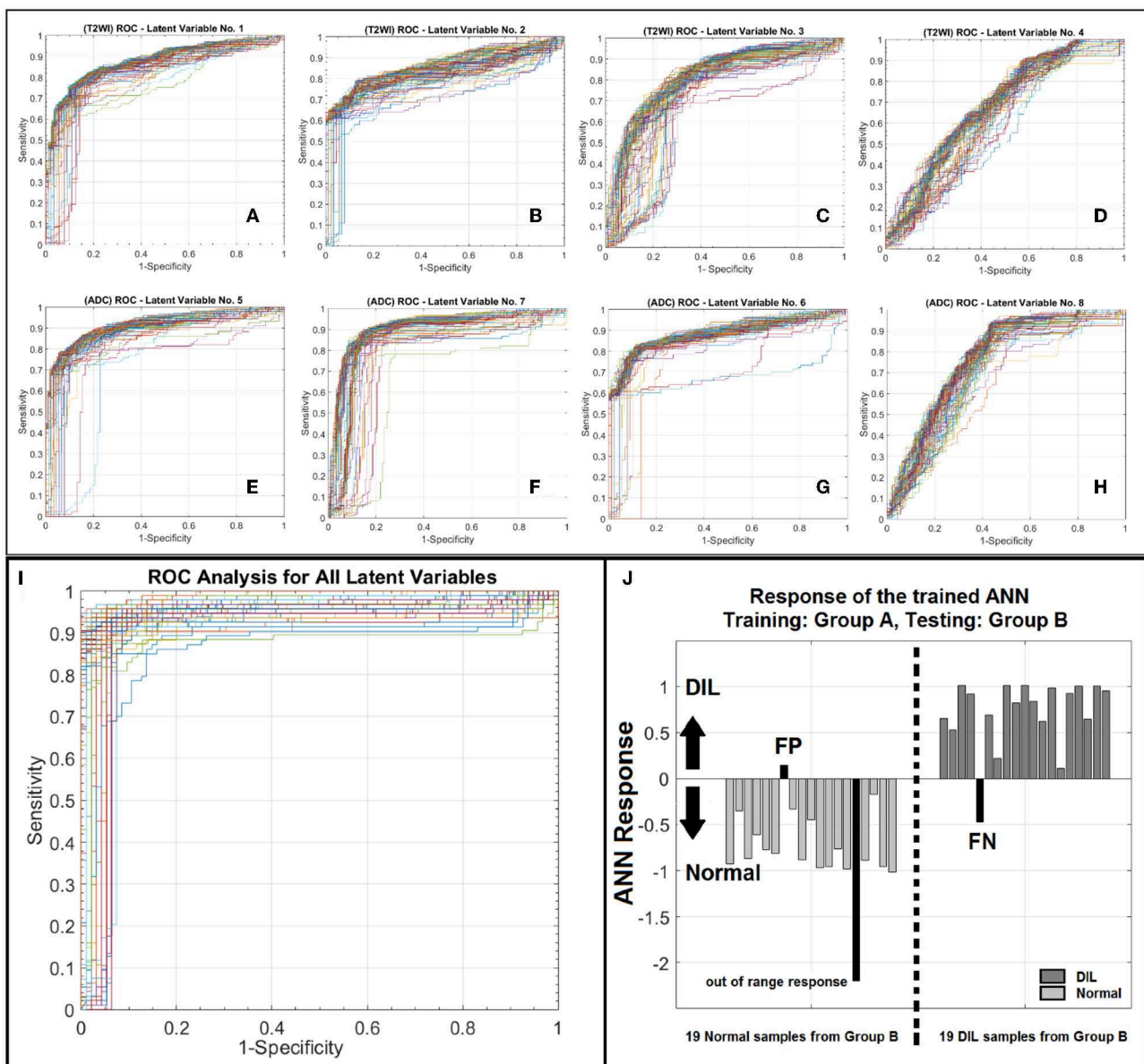
FIGURE 5 | (A–D) Depict ROC curves corresponding to 100 iterations each corresponding to a different division of training and validation datasets for ANN for the T2WI latent variables number 1–4. **(E–H)** depict ROC curves corresponding to 100 iterations each corresponding to a different division of training and validation datasets for ANN for the ADC latent variables number 1–4. As shown in this figure for each modality, from left to right as the order of latent variable increases the information content or discrimination power of the variable for classification deceases. **(I)** illustrates a family of ROC curves for 100 iterations, each corresponding to a different division of training and validation datasets for ANN for all 8 latent variables. **(J)** shows the response of the trained ANN against an unseen/prospective dataset (trained with group A and tested with group B).

parameters for prospective patients whom could be scanned with different scanners or different imaging parameters.

As shown in **Figure 5** and according to the statistical measures reported in **Table 4**, as it is expected, for each modality, from left to right (**Figures 5A–D** or **Figures 5E–H**), as the order of the latent variable increases the information content or discrimination power of the variable for DIL classification deceases. As shown in **Table 4** and **Figure 5**, the analysis results

strongly confirm that compared to T2WI modality, the ADC modality is more discriminative with higher information content for the classification of DILs and NTs.

The application of novel machine learning techniques such as Bayesian approach, Support vector machine (SVM) kernels: polynomial, radial base function (RBF) and Gaussian and Decision Tree for detecting prostate cancer have been proposed by several research groups (40–42). Moreover, different

TABLE 4 | This table shows AUROC, NPV, and PPV values along with their confidence intervals measured for each individual latent variable for 100 iterations (each corresponding to a different division of training and validation datasets).

Latent variable	AUROC	AUROC-CI	PPV	PPV-CI	NPV	NPV-CI
First latent variable (T2WI)	0.87	0.86–0.89	0.86	0.84–0.88	0.78	0.76–0.79
Second latent variable (T2WI)	0.79	0.72–0.85	0.84	0.78–0.90	0.71	0.68–0.74
Third latent variable (T2WI)	0.76	0.75–0.80	0.72	0.67 0.74	0.72	0.70–0.74
Fourth latent variable (T2WI)	0.66	0.64–0.68	0.58	0.57–0.60	0.69	0.65–0.73
First latent variable (ADC)	0.91	0.90–0.92	0.88	0.85–0.90	0.82	0.80–0.84
Second latent variable (ADC)	0.88	0.86–0.89	0.87	0.85–0.89	0.81	0.80–0.83
Third latent variable (ADC)	0.79	0.72–0.87	0.80	0.75–0.85	0.83	0.81–0.86
Fourth latent variable (ADC)	0.74	0.72–0.76	0.66	0.64–0.67	0.81	0.78–0.85
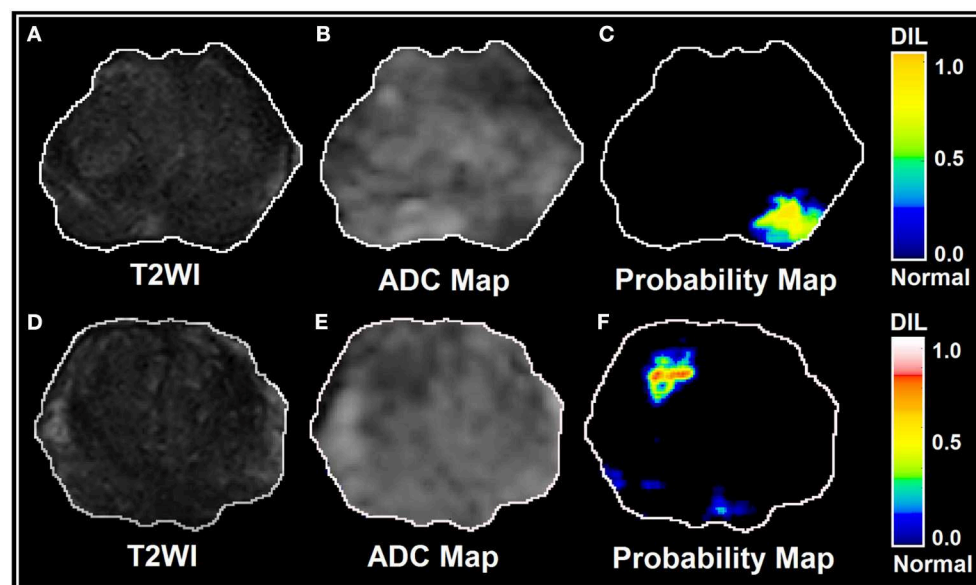
FIGURE 6 | (A–F) illustrate T2WI, ADC map, and lesion probability map for a slice of prostate gland for two different patients estimated by the trained PLSC-ANN using a 2D-convolutional sampling method (window size = 25 × 25).

features extracting strategies are proposed to improve the DIL detection performance (40). ANNs have been used in different fields on a variety of tasks such as computer vision, speech recognition, machine translation, social network filtering, medical diagnosis, and in many other domains. There have been numerous applications of ANNs within medical decision-making (26, 43, 44). It has been shown that ANNs have unique properties including robust performance in dealing with noisy or incomplete input patterns, high fault tolerance, and the ability to generalize from the training data (26, 43). The adaptive model constructed in this study can benefit from the ANN's properties stated above and can distinguish DILs from NTs with almost uniform sensitivity at different levels of specificities (see **Figures 4A,B, 5I**). The stability (lesions being non-patchy and uniform) of the predicted DILs and NTs in the probability maps (shown in **Figure 6**) clearly confirm the robustness of the PLSC-ANN technique in information extraction from the two MR modalities. The proposed ANN in this study was trained without any data augmentation. The results implied that

the trained ANN can also evaluate any suspicious lesion in different zones of the prostate gland (PZ or TZ) regardless of its Gleason score.

Our study also confirms that the most discriminant features are textural-based features and given the bootstrapping feature ranking results, it can be concluded that frequency or arrangement-based features (LBP, GLRL, DOST, and 2DGF, see **Table 3**, a measure of the decoration or disorder of information distribution within a region), that are associated with subtle and descriptive information content of the two image modalities, play a key role in discrimination of DIL from NT. Also, we did not include morphological features such as volume, shape, solidity, convexity, eccentricity, and etc. in order to eliminate any possible biasing result from the manual contouring of DILs and NTs.

In this study, DIL and the NT contours were separately drawn on each image modality. While such a process could increase the chance of contour variability and negatively increase the variation of the data, it had an advantage that the two image modalities (T2W images and ADC map) did not necessarily need

to be co-registered to each other prior to the radiomic analysis and adaptive modeling and therefore, the analysis results were not negatively affected by any possible co-registration errors. DILs and NTs contoured on unregistered image modalities were directly used for training and testing of the ANN. We only co-registered the two image modalities (T2WI and ADC map) using rigid co-registration [affine transform (45)] method for the two test cases (see **Figure 6**) to predict DIL-NT probability map using the trained ANN and 2D-convolutional sampling method.

The current major computer aided diagnosis systems recorded AUROC performance ranging from 0.77 to 0.89 and the focus was to detect lesions in the peripheral zone. Most image features, either individually or in combination that were effective in the differentiation of prostate cancer, are volume averaged quantities such as the 10th percentile of the ADC, T2W signal intensity skewness (46). Niaf et al. studied texture features extracted from MP-MRI on 30 fully annotated patients using four different feature selection and classification methods (47). They could achieve a diagnostic performance of 0.89 but the study was limited to the peripheral zone only. The performance was poorer due to the overfitting problem when all features were used for classification.

In this study, despite using 117 subjects (two cohorts: 96, and 19) with two different training and validation strategies, there are still several challenges as follows: Compared to the number of radiomics features, the study is limited by the number of patients, which will impact the optimal features selected, and also might render a predictive model susceptible to Type II errors. A larger sample size will also allow the construction of a more reliable ANN in order to draw a reliable and unequivocal conclusion.

In this study, two different training and validation strategies were recruited and the strong agreement between the analysis results confirmed the robustness of the identified features. In the first strategy, employing the LOOCV method in this study, allowed us to use a high proportion of the available training data fraction ($1 - 1/K = 0.99$ for $K = 117$), for training, while making use of all the data in estimating the generalization error or agreement. The cost is that the process can be lengthy, since we need to train and evaluate the network K times. Typically, according to the literatures, $K \approx 10$ is considered reasonable (48). In this study, K was set to 117 for 117 patients (one case with DIL and NT in each fold) and the ANN had a single output, to predict the outcome. The radiomics features selected might be impacted by the intensities, size of the contour, and contrast of the NT. Since the region of interests were delineated manually, the accuracy and variability of the ROIs could impact on the optimal feature selection and the training results.

The Az-test for the average ROC analysis of the ANN is 1% higher than the Az-test of the optimal ANN (see **Figures 3C,D**). This is due to the difference between the way the two tests are conducted: for average AUROC, each NT or DIL from each subject is considered as a sample (thus the total samples are equal to 234) while in the ordinary Az-test for the optimal ANN, pair of NT and DIL for each subject is considered as a sample (thus the total samples are equal to 117). Strong agreement between the statistical measures of the LOOCV and conventional methods and also the high predictive power of the trained ANN (group A)

when it was applied on group B (as prospective or unseen data cohort), confirm the consistency and high information content of the discriminant features identified in this study.

The 2D-convolutional sampling analysis results presented in **Figure 6**, imply that the trained-ANN is capable of estimating the DIL and normal tissue probabilities when the target contour (the 2D window) consists of a mixed radiomic information extracted from DIL and normal tissue.

ANN was implanted as a classifier since it has high tolerance against variation of input feature components and contours (according to the contour variability test results) while they are less sensitive to random noise (49), which allows the construction of a variation- and noise-insensitive adaptive classifier with higher accuracy and speed. Most importantly, ANN considers non-linear relationships among input data that cannot always be recognized by conventional analyses. Results of the permutation test also imply that the discriminant features used for training, are reliable and efficient for classification.

CONCLUSION

In conclusion, this study demonstrates the high performance of combining radiomics analysis, PLSC technique and adaptive model for extracting and ranking features from multimodal MR information such as T2WI and ADC maps to detect DILs and NTs in patients with PCa. The radiomics information of ADC modality was proved to have higher discrimination power compared to the corresponding features extracted from T2WI modality. Results are suggestive that the integration of quantitative image analysis methods such as radiomics analysis and PLSC technique when combined with an adaptive model can help identify imaging biomarkers and show great potential to help clinicians improve the classification of clinically significant prostate lesions for therapy of prostate cancer.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this manuscript will be made available by the authors, without undue reservation, to any qualified researcher.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by The Internal Review Board at Henry Ford Health System. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

HB-E and NW designed the research and methodology. HB-E, CL, BJ, and NW performed the research. HB-E and CL contributed to the data pre-processing. HB-E developed statistical analysis, PLSC, ANN training, and validation as well as the development of analytical tools. HB-E and NW wrote the

REFERENCES

1. Dinh CV, Steenbergen P, Ghobadi G, Heijmink SW, Pos FJ, Haustermans K, et al. Magnetic resonance imaging for prostate cancer radiotherapy. *Phys Med.* (2016) 32:446–51. doi: 10.1016/j.ejmp.2016.01.484

2. Moghanaki D, Turkbey B, Vapiwala N, Ehdaie B, Frank SJ, McLaughlin PW, et al. Advances in prostate cancer magnetic resonance imaging and positron emission tomography-computed tomography for staging and radiotherapy treatment planning. *Semin Radiat Oncol.* (2017) 27:21–33. doi: 10.1016/j.semradonc.2016.08.008

3. Dearnaley DP, Jovic G, Syndikus I, Khoo V, Cowan RA, Graham JD, et al. Escalated-dose versus control-dose conformal radiotherapy for prostate cancer: long-term results from the mrc rt01 randomised controlled trial. *Lancet Oncol.* (2014) 15:464–73. doi: 10.1016/S1470-2045(14)70040-3

4. Li B, Du Y, Yang H, Huang Y, Meng J, Xiao D. Magnetic resonance imaging for prostate cancer clinical application. *Chin J Cancer Res.* (2013) 25:240–9. doi: 10.3978/j.issn.1000-9604.2013.06.04

5. Anderson ES, Margolis DJ, Mesko S, Banerjee R, Wang PC, Demanes DJ, et al. Multiparametric MRI identifies and stratifies prostate cancer lesions: implications for targeting intraprostatic targets. *Brachytherapy.* (2014) 13:292–8. doi: 10.1016/j.brachy.2014.01.011

6. Rischke HC, Nestle U, Fechter T, Doll C, Volegova-Neher N, Henne K, et al. 3 Tesla multiparametric MRI for gtv-definition of dominant intraprostatic lesions in patients with prostate cancer–an interobserver variability study. *Radiat Oncol.* (2013) 8:183. doi: 10.1186/1748-717X-8-183

7. Turkbey B, Mani H, Shah V, Rastinehad AR, Bernardo M, Pohida T, et al. Multiparametric 3T prostate magnetic resonance imaging to detect cancer: histopathological correlation using prostatectomy specimens processed in customized magnetic resonance imaging based molds. *J Urol.* (2011) 186:1818–24. doi: 10.1016/j.juro.2011.07.013

8. Delongchamps NB, Beuvon F, Eiss D, Flam T, Muradyan N, Zerbib M, et al. Multiparametric MRI is helpful to predict tumor focality, stage, and size in patients diagnosed with unilateral low-risk prostate cancer. *Prostate Cancer Prostatic Dis.* (2011) 14:232–7. doi: 10.1038/pcan.2011.9

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) 30:1234–48. doi: 10.1016/j.mri.2012.06.010

10. 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/ncomms5644

11. Litjens G, Debats O, Barentsz J, Karssemeijer N, Huisman H. Computer-aided detection of prostate cancer in MRI. *IEEE Trans Med Imaging.* (2014) 33:1083–92. doi: 10.1109/TMI.2014.2303821

12. Fei B. Computer-aided diagnosis of prostate cancer with MRI. *Curr Opin Biomed Eng.* (2017) 3:20–7. doi: 10.1016/j.cobme.2017.09.009

13. Lemaître G, Martí R, Freixenet J, Vilanova JC, Walker PM, Meriaudeau F. Computer-aided detection and diagnosis for prostate cancer based on mono and multi-parametric MRI: a review.

Comput Biol Med. (2015) 60:8–31. doi: 10.1016/j.compbiomed.2015.02.009

14. Barentsz JO, Richenberg J, Clements R, Choyke P, Verma S, Villeirs G, et al. ESUR prostate mr guidelines 2012. *Eur Radiol.* (2012) 22:746–57. doi: 10.1007/s00330-011-2377-y

15. Bagher-Ebadian H, Siddiqui F, Liu C, Movsas B, Chetty IJ. On the impact of smoothing and noise on robustness of CT and CBCT radiomics features for patients with head and neck cancers. *Med Phys.* (2017) 44:1755–70. doi: 10.1002/mp.12188

16. Abdi H, Williams LJ. Partial least squares methods: partial least squares correlation and partial least square regression. *Methods Mol Biol.* (2013) 930:549–79. doi: 10.1007/978-1-62703-059-5_23

17. Grellmann C, Bitzer S, Neumann J, Westlye LT, Andreassen OA, Villringer A, et al. Comparison of variants of canonical correlation analysis and partial least squares for combined analysis of MRI and genetic data. *Neuroimage.* (2015) 107:289–310. doi: 10.1016/j.neuroimage.2014.12.025

18. Holm S. A simple sequentially rejective multiple test procedure. *Scand J Stat.* (1979) 6:65–70.

19. Abdi H, Dunlop JP, Williams LJ. How to compute reliability estimates and display confidence and tolerance intervals for pattern classifiers using the bootstrap and 3-way multidimensional scaling (distatis). *Neuroimage.* (2009) 45:89–95. doi: 10.1016/j.neuroimage.2008.11.008

20. McIntosh AR, Chau WK, Protzner AB. Spatiotemporal analysis of event-related fmri data using partial least squares. *Neuroimage.* (2004) 23:764–75. doi: 10.1016/j.neuroimage.2004.05.018

21. Freeman JA, Skapura DM. *Neural Networks: Algorithms, Applications, and Programming Techniques.* Reading, MA: Addison-Wesley (1991).

22. Looney CG. *Pattern Recognition Using Neural Networks : Theory and Algorithms for Engineers and Scientists.* New York, NY: Oxford University Press (1997).

23. Begg R, Kamruzzaman J, Sarkar R. *Neural Networks in Healthcare: Potential and Challenges.* Hershey, PA: Idea Group Pub (2006).

24. Bagher-Ebadian H, Jafari-Khouzani K, Mitsias PD, Lu M, Soltanian-Zadeh H, Chopp M, et al. Predicting final extent of ischemic infarction using artificial neural network analysis of multi-parametric mri in patients with stroke. *PLoS ONE.* (2011) 6:e22626. doi: 10.1371/journal.pone.0022626

25. Bagher-Ebadian H, Nagaraja TN, Paudyal R, Whitton P, Panda S, Fenstermacher JD, et al. MRI estimation of contrast agent concentration in tissue using a neural network approach. *Magn Reson Med.* (2007) 58:290–7. doi: 10.1002/mrm.21332

26. Pasini A. Artificial neural networks for small dataset analysis. *J Thorac Dis.* (2015) 7:953–60. doi: 10.3978/j.issn.2072-1439.2015.04.61

27. Beau-Faller M, Degeorges A, Rolland E, Mounawar M, Antoine M, Poulot V, et al. Cross-validation study for epidermal growth factor receptor and KRAS mutation detection in 74 blinded non-small cell lung carcinoma samples: a total of 5550 exons sequenced by 15 molecular French laboratories (evaluation of the EGFR mutation status for the administration of EGFR-TKIs in non-small cell lung carcinoma [ERMETIC] project–part 1). *J Thorac Oncol.* (2011) 6:1006–15. doi: 10.1097/JTO.0b013e318211dcee

28. Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence.* San Mateo, CA: Morgan Kaufmann (1995).

29. Perkins NJ, Schisterman EF. The inconsistency of "optimal" cutpoints obtained using two criteria based on the receiver operating characteristic curve. *Am J Epidemiol.* (2006) 163:670–5. doi: 10.1093/aje/kwj063

30. Unal I. Defining an optimal cut-point value in roc analysis: an alternative approach. *Comput Math Methods Med.* (2017) 2017:3762651. doi: 10.1155/2017/3762651

31. Steyerberg EW, Harrell FE Jr. Prediction models need appropriate internal, internal-external, and external validation. *J Clin Epidemiol.* (2016) 69:245–7. doi: 10.1016/j.jclinepi.2015.04.005

32. Gevaert O, Xu J, Hoang CD, Leung AN, Xu Y, Quon A, et al. Non-small cell lung cancer: identifying prognostic imaging biomarkers by leveraging public gene expression microarray data–methods and preliminary results. *Radiology.* (2012) 264:387–96. doi: 10.1148/radiol.12111607

33. Ziegler G, Dahnke R, Winkler AD, Gaser C. Partial least squares correlation of multivariate cognitive abilities and local brain structure in children and adolescents. *Neuroimage.* (2013) 82:284–94. doi: 10.1016/j.neuroimage.2013.05.088

34. Celik A. Effect of imaging parameters on the accuracy of apparent diffusion coefficient and optimization strategies. *Diagn Interv Radiol.* (2016) 22:101–7. doi: 10.5152/dir.2015.14440

35. Ogura A, Hayakawa K, Miyati T, Maeda F. Imaging parameter effects in apparent diffusion coefficient determination of magnetic resonance imaging. *Eur J Radiol.* (2011) 77:185–8. doi: 10.1016/j.ejrad.2009.06.031

36. Qin W, Yu CS, Zhang F, Du XY, Jiang H, Yan YX, et al. Effects of echo time on diffusion quantification of brain white matter at 1.5 T and 3.0 T. *Magn Reson Med.* (2009) 61:755–60. doi: 10.1002/mrm.21920

37. Saritas EU, Lee JH, Nishimura DG. Snr dependence of optimal parameters for apparent diffusion coefficient measurements. *IEEE Trans Med Imaging.* (2011) 30:424–37. doi: 10.1109/TMI.2010.2084583

38. Soman S, Holdsworth SJ, Skare S, Andre JB, Van AT, Aksoy M, et al. Effect of number of acquisitions in diffusion tensor imaging of the pediatric brain: optimizing scan time and diagnostic experience. *J Neuroimaging.* (2015) 25:296–302. doi: 10.1111/jon.12093

39. Wang S, Peng Y, Medved M, Yousuf AN, Ivancevic MK, Karademir I, et al. Hybrid multidimensional T(2) and diffusion-weighted mri for prostate cancer detection. *J Magn Reson Imaging.* (2014) 39:781–8. doi: 10.1002/jmri.24212

40. Hussain L, Ahmed A, Saeed S, Rathore S, Awan IA, Shah SA, et al. Prostate cancer detection using machine learning techniques by employing combination of features extracting strategies. *Cancer Biomark.* (2018) 21:393–413. doi: 10.3233/CBM-170643

41. Lv D, Guo X, Wang X, Zhang J, Fang J. Computerized characterization of prostate cancer by fractal analysis in MR images. *J Magn Reson Imaging.* (2009) 30:161–8. doi: 10.1002/jmri.21819

42. Tiwari P, Kurhanewicz J, Madabhushi A. Multi-kernel graph embedding for detection, gleason grading of prostate cancer via MRI/MRS. *Med Image Anal.* (2013) 17:219–35. doi: 10.1016/j.media.2012.10.004

43. Hu X, Cammann H, Meyer HA, Miller K, Jung K, Stephan C. Artificial neural networks and prostate cancer–tools for diagnosis and management. *Nat Rev Urol.* (2013) 10:174–82. doi: 10.1038/nrurol.2013.9

44. Schwarzer G, Schumacher M. Artificial neural networks for diagnosis and prognosis in prostate cancer. *Semin Urol Oncol.* (2002) 20:89–95. doi: 10.1053/suro.2002.32492

45. Berger M. *Geometry.* Berlin: Springer (1977).

46. Peng Y, Jiang Y, Yang C, Brown JB, Antic T, Sethi I, et al. Quantitative analysis of multiparametric prostate MR images: differentiation between prostate cancer and normal tissue and correlation with gleason score–a computer-aided diagnosis development study. *Radiology.* (2013) 267:787–96. doi: 10.1148/radiol.13121454

47. Niaf E, Rouvière O, Mège-Lechevallier F, Bratan F, Lartizien C. Computer-aided diagnosis of prostate cancer in the peripheral zone using multiparametric MRI. *Phys Med Biol.* (2012) 57:3833–51. doi: 10.1088/0031-9155/57/12/3833

48. Bishop CM. *Neural Networks for Pattern Recognition.* London: Oxford University Press (1997).

49. Miloš J. Madi VJM. Assessing the sensitivity of the artificial neural network to experimental noise: a case study. *FME Trans.* (2010) 38:189–195. Available online at: https://pdfs.semanticscholar.org/985a/96ee91f2418baf32544500eb7e7ddaf0077d.pdf?_ga=2.269201969.1397016782.1574085333-336839267.1574085333

Synthetic CT Generation Based on T2 Weighted MRI of Nasopharyngeal Carcinoma (NPC) Using a Deep Convolutional Neural Network (DCNN)

Yuenan Wang[1]*, Chenbin Liu[1], Xiao Zhang[2] and Weiwei Deng[2]

[1] Department of Radiation Oncology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital & Shenzhen Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Shenzhen, China,
[2] Department of Mechanics and Aerospace Engineering, Southern University of Science and Technology, Shenzhen, China

Purpose: There is an emerging interest of applying magnetic resonance imaging (MRI) to radiotherapy (RT) due to its superior soft tissue contrast for accurate target delineation as well as functional information for evaluating treatment response. MRI-based RT planning has great potential to enable dose escalation to tumors while reducing toxicities to surrounding normal tissues in RT treatments of nasopharyngeal carcinoma (NPC). Our study aims to generate synthetic CT from T2-weighted MRI using a deep learning algorithm.

Methods: Thirty-three NPC patients were retrospectively selected for this study with local IRB's approval. All patients underwent clinical CT simulation and 1.5T MRI within the same week in our hospital. Prior to CT/MRI image registration, we had to normalize two different modalities to a similar intensity scale using the histogram matching method. Then CT and T2 weighted MRI were rigidly and deformably registered using intensity-based registration toolbox *elastix* (version 4.9). A U-net deep learning algorithm with 23 convolutional layers was developed to generate synthetic CT (sCT) using 23 NPC patients' images as the training set. The rest 10 NPC patients were used as the test set (∼1/3 of all datasets). Mean absolute error (MAE) and mean error (ME) were calculated to evaluate HU differences between true CT and sCT in bone, soft tissue and overall region.

Results: The proposed U-net algorithm was able to create sCT based on T2-weighted MRI in NPC patients, which took 7 s per patient on average. Compared to true CT, MAE of sCT in all tested patients was 97 ± 13 Hounsfield Unit (HU) in soft tissue, 131 ± 24 HU in overall region, and 357 ± 44 HU in bone, respectively. ME was −48 ± 10 HU in soft tissue, −6 ± 13 HU in overall region, and 247 ± 44 HU in bone, respectively. The majority soft tissue and bone region was reconstructed accurately except the interface between soft tissue and bone and some delicate structures in nasal cavity, where the inaccuracy was induced by imperfect deformable registration. One patient example was shown with almost no difference in dose distribution using true CT vs. sCT in the PTV regions in the sinus area with fine bone structures.

Conclusion: Our study indicates that it is feasible to generate high quality sCT images based on T2-weighted MRI using the deep learning algorithm in patients with nasopharyngeal carcinoma, which may have great clinical potential for MRI-only treatment planning in the future.

INTRODUCTION

There is an emerging interest in applying magnetic resonance imaging (MRI) during radiation treatment (RT) (1, 2). This is mainly because MRI can provide superior soft tissue contrast without ionizing radiation. MRI offers more consistent and accurate target delineation in head and neck cancers, brain tumors, sarcomas, and tumor sites in the abdomen and pelvis (3–6). It has been reported that applying MRI to RT has great benefits to improve radiation dosimetry and to increase therapeutic ratio, such as reducing toxicity to critical organs and enabling dose escalation to tumor sites to achieve survival gains (4, 7). In addition, not only anatomical but also functional information can be obtained non-invasively using MRI, which makes MRI suitable for quantitative and longitudinal evaluation of treatment response (8–10). Therefore, MRI integrated with the conventional CT-sim in RT planning has become an essential step in modern RT process (1–3).

As we know, nasopharyngeal carcinoma (NPC) is a common malignancy in Southeast Asia. Integrating MRI to RT in patients with NPC can be especially helpful due to its relatively complicated target structures and surrounding critical normal tissues. Accurate delineation of critical structures and tumors in NPC may not only help patients gain survival but also improve life quality. However, there are multiple challenges in integrating MRI to clinical RT. The acquisition time of MRI pulse sequences is typically much longer than CT, since the MRI scanning protocol generally includes not only localizer, T1 weighted, T2 weighted, diffusion weighted imaging (DWI) but also dynamic contrast enhanced (DCE) sequences. Also, parameters of MRI pulse sequences such as bandwidth, TR, TE and the receiver coils need to be manipulated based on patients' anatomical sites or pathological examinations. MRI in general is more technically challenging to radiation physicists and physicians compared to CT. Hence, MRI technologists may need more time to adjust complex parameters or to optimize coils during an MRI scan (11). Secondly, MRI is inherently susceptible to motion artifact and geometric distortion (1, 2, 11, 12). For example, the geometrical uncertainty of ∼2 and 2–3 mm was observed for the brain and pelvic sites, respectively (13, 14). Such systematic errors can lead to RT target miss and compromise local control.

Another well-known challenge lies in the conversion of electron density or HU values in synthetic CT based on MR images. CT images can be used for radiation treatment planning is because they can be directly scaled to photon attenuation map. However, MRI does not provide such information (11, 12). Currently there are three methods of mapping HU based on the intensity of MR images (15, 16): atlas-based (17), voxel-based (18), and hybrid type (19). The atlas-based method of producing synthetic CT images may require CT to MRI registration where CT and MRI atlas scan pair can correspond anatomically (17). In contrast, voxel-based method is focused on using voxel by voxel mapping based on intensity or spatial location of the MRI images acquired from different MRI pulse sequences (18). The hybrid method combines atlas- and voxel-based methods, where deformable registration from the atlas-based method and local pattern recognition from the voxel-based method are applied to obtain attenuation information in the MR images. From this point of view, our proposed deep learning method where both registration and voxel-by-voxel patterns are learned through U-net, can be considered as the hybrid method.

In fact, machine learning and deep learning have been applied to many medical fields including radiation oncology (20), which has main components of data, model, cost or loss of the model, and model optimizer. Topics of how to apply and what are the challenges of machine learning, neural networks, and artificial intelligence (AI) to the clinical RT process have been discussed previously on the red journal (21). Here we aim to apply deep learning algorithms such as the U-net convolution neural network (CNN) approach to convert T2-weighted MRI to synthetic CT.

MATERIALS AND METHODS

To convert the T2-weighted MRI to synthetic CT images, there were four major steps in our method illustrated in **Figure 1**: (1) MR image normalization into the similar intensity scale; (2) voxel-based rigid and deformable registration for CT and MRI; (3) U-net model training with 2/3 datasets; (4) U-net model testing with the rest 1/3 datasets and evaluation of the synthetic CT images.

Data Acquisition

Thirty-three nasopharyngeal carcinoma (NPC) patients were retrospectively selected for this study with the approval of our hospital's internal review board (IRB). All patients underwent CT simulation in the head-first supine position with the Civco 5-point head, neck and shoulder mask on a GE Discovery CT scanner (Milwaukee, WI, USA) prior to RT planning with resolution of 512 × 512, slice thickness of 2.5 mm, 120 kVp and 300 mAs. Within the same week of CT acquisitions, diagnostic MRI was obtained using 1.5 T Siemens Avanto MRI scanner (Erlangen, Germany) in our hospital, where T2 weighted MRI
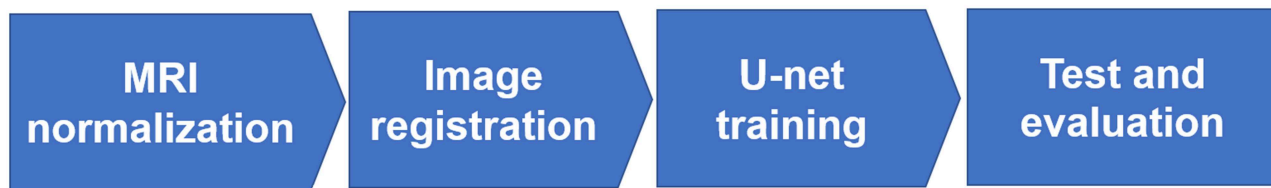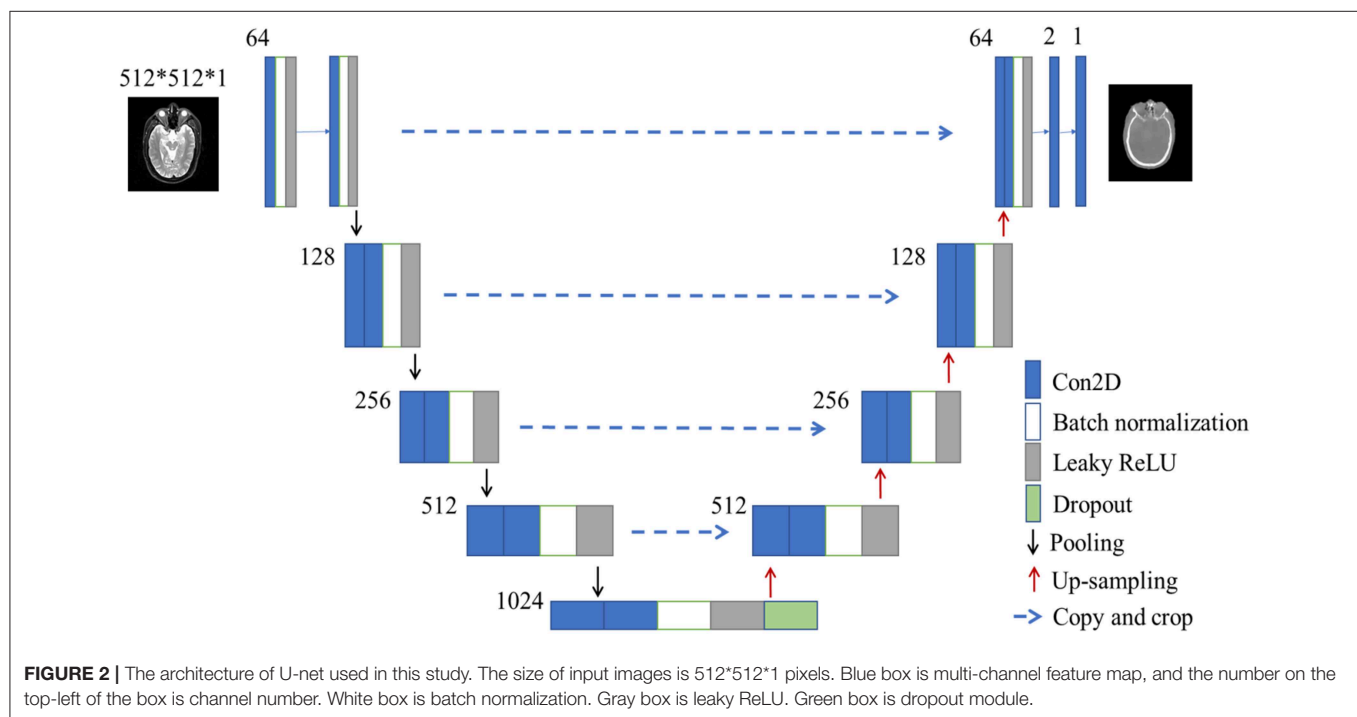
FIGURE 1 | The workflow diagram of generating synthetic CT from T2-weighted MRI using U-net.
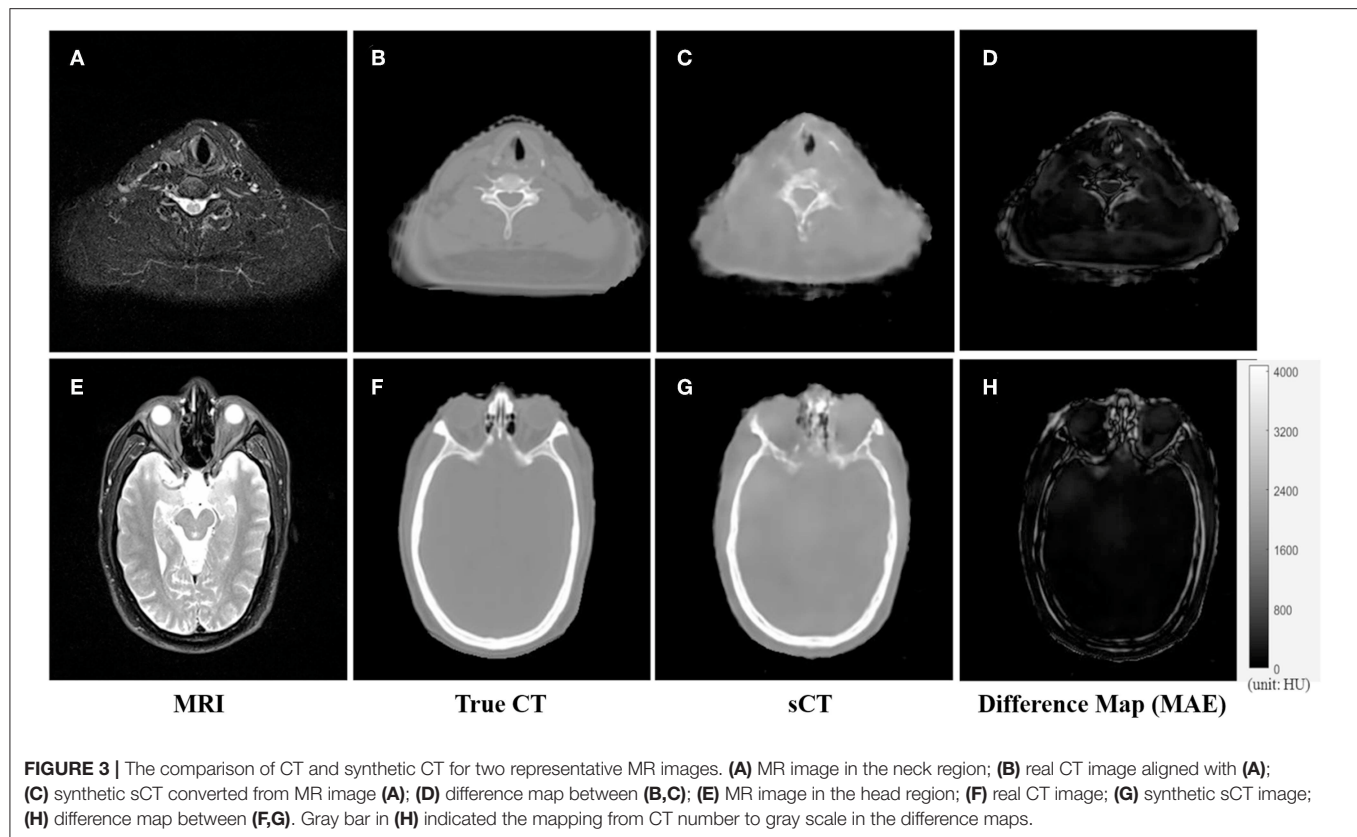

FIGURE 2 | The architecture of U-net used in this study. The size of input images is 512*512*1 pixels. Blue box is multi-channel feature map, and the number on the top-left of the box is channel number. White box is batch normalization. Gray box is leaky ReLU. Green box is dropout module.

was acquired using fat-saturated (FS) turbo spin echo (TSE) with resolution of 256 × 256 and slice thickness of 5 mm.

Image Preprocessing

Prior to the rigid and deformable registration between T2 weighted MRI and CT images, we had to normalize the two imaging sets of different modalities to a similar intensity scale using the histogram matching method (**Figure 1**'s first step: MRI normalization). Although lacking of a normalized intensity scale of MRI had no impact on clinical diagnosis provided by radiologists, it would influence the quality of image registration and deep learning, which highly depended on the similarity of image intensity between MRI and CT to achieve high-quality results. We used histogram matching method, which was independent of patients' image sets and specific brands of the MRI scanner used (22). In our study, the normalization process took account of all the NPC patients' samples by identifying 10 decile landmarks in the histogram of each MR image and calculated the mean values of each landmark as the standard scale. It was used to transform the MR images of the same protocol and body region to the standard scale (23).

To conduct rigid and deformable registration of the MRI and CT imaging modalities, we used an open source image registration package called *elastix* (version 4.9) (24, 25), where the traditional iterative intensity-based image registration method was applied. For all NPC patients, the rigid image registration was performed followed by deformable registration. In the rigid registration, multi-resolution registration method was used, and the optimizer was adaptive stochastic gradient descent. In the deformable registration, multi-metric and multi-resolution registration method was used with advanced Mattes mutual information as the similarity metrics and transform bending energy penalty for smooth displacement (26) (**Figure 1**'s second step: Image registration).

After image normalization and image registration steps described as the above, a U-net deep learning method was developed to generate synthetic CT from T2-weighted MRI using 23 convolutional layers of CNN, shown in **Figure 2**. To train and evaluate the U-net model, the 33 patients' dataset were randomly divided into two groups: 23 were used as the training set (∼2/3 of the total datasets) and the rest 10 were used as the test set (∼1/3 of the total datasets) (**Figure 1**'s third and fourth steps).

FIGURE 3 | The comparison of CT and synthetic CT for two representative MR images. **(A)** MR image in the neck region; **(B)** real CT image aligned with **(A)**; **(C)** synthetic sCT converted from MR image **(A)**; **(D)** difference map between **(B,C)**; **(E)** MR image in the head region; **(F)** real CT image; **(G)** synthetic sCT image; **(H)** difference map between **(F,G)**. Gray bar in **(H)** indicated the mapping from CT number to gray scale in the difference maps.

U-Net as a Deep Learning Algorithm

The U-net CNN structure consists of a contracting path and an expansive path (27), shown in **Figure 2**. The contracting path follows the typical architecture of a convolutional network. It consists of the repeated application of two 3 × 3 convolutions (unpadded convolutions), each followed by a rectified linear unit (ReLU) and a 2 × 2 max pooling operation with a stride of 2 for down-sampling. At each down-sampling step, we doubled the number of feature channels.

In contrast to the contracting path, the expansive path is composed of an up-sampling of the feature map followed by a 2 × 2 convolution (i.e., "up-convolution") that halved the number of feature channels, a concatenation with the correspondingly cropped feature map from the contracting path, and two 3 × 3 convolutions, each followed by a ReLU. The cropping is necessary due to the loss of border pixels in every convolution. At the final layer, a 1 × 1 convolution was used to map each 64-component feature vector to the desired number of classes. In the final layer, a convolution was used to map the feature to the desired value, which was the intensity of the synthetic CT. Therefore, in the expansive path, a large amount of image features was used to reconstruct a new image of the same size as the input one. The implementation of our U-net was shown in **Figure 2**.
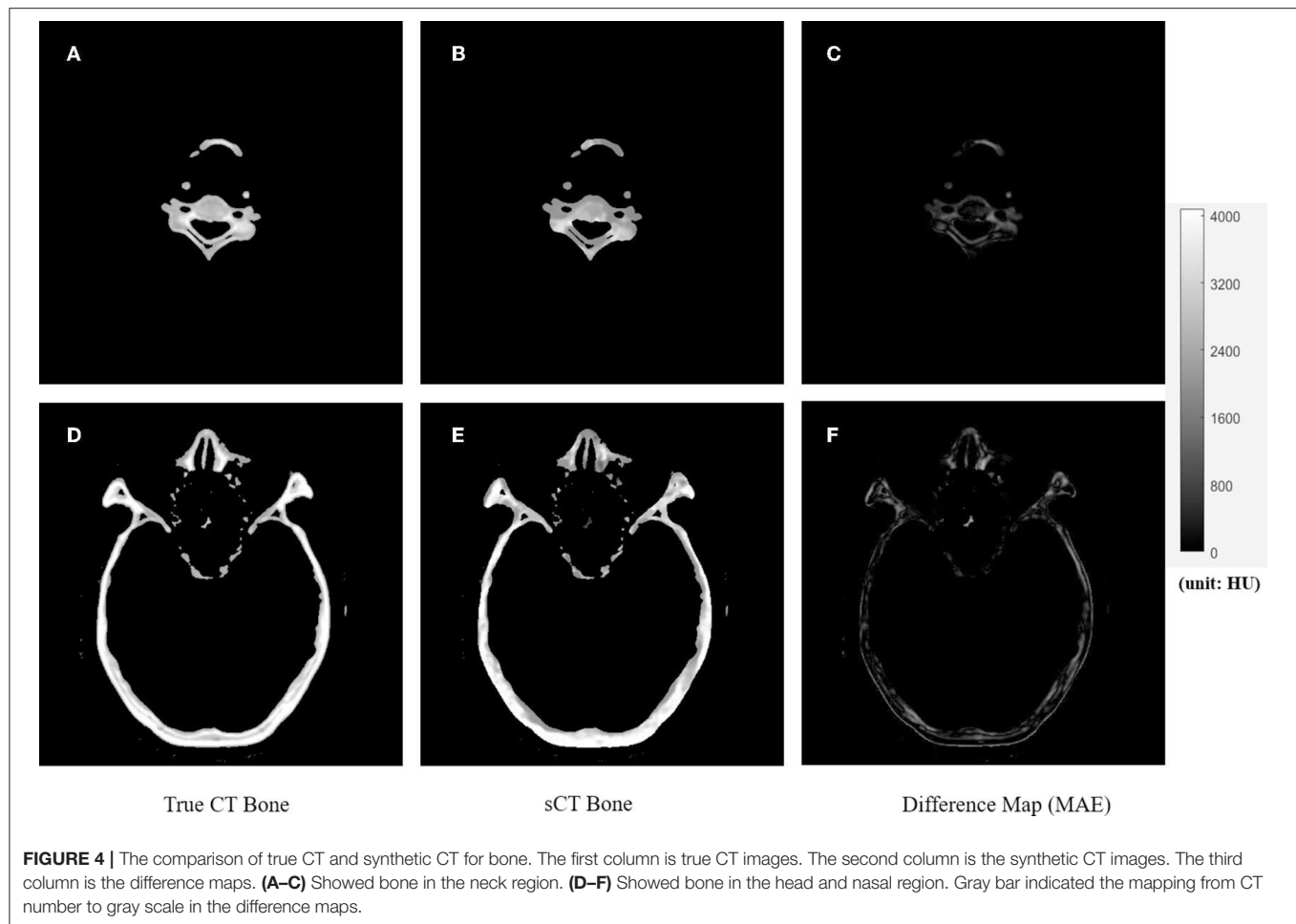
Here we used batch normalization and leaky ReLU in our network, which was different from the classical U-net (27). Our U-net was developed in Keras framework which was a high-level neural network API with Tensorflow as the backend. In total, the U-net network in our study had 23 convolutional layers. To allow a seamless tiling of the output segmentation map, we also selected the input tile size such that all 2 × 2 max-pooling operations were applied to a layer with an even x- and y-size.

Evaluation

The 33 NPC patients were randomly divided into two groups: 23 as the training set and 10 as the test set. The U-net model described in the previous section was trained through feeding MRI and CT images from the training set into the neural network. The synthetic CTs were generated using the trained model for the test set.

To visually inspect the difference between true CT and synthetic CT, difference maps were generated. The pixel intensity of the difference map was the absolute difference between real CT and synthetic CT. Darker region in the difference map indicated smaller errors of CT values or HU number in the region of synthetic CT, and vice versa.

The mean absolute error (MAE) and mean error (ME) were used to quantify the absolute difference and mean difference within the body, respectively. The body masks were

FIGURE 4 | The comparison of true CT and synthetic CT for bone. The first column is true CT images. The second column is the synthetic CT images. The third column is the difference maps. **(A–C)** Showed bone in the neck region. **(D–F)** Showed bone in the head and nasal region. Gray bar indicated the mapping from CT number to gray scale in the difference maps.

generated using OTSU's thresholding method and morphological operations (28, 29).

$$MAE = \frac{1}{n} \sum_{i=1}^{n} \left| CT(i) - sCT(i) \right| \tag{1}$$

$$ME = \frac{1}{n} \sum_{i=1}^{n} \left(CT(i) - sCT(i) \right) \tag{2}$$

where n is the total number of pixels within the body outline. $CT(i)$ is the ith pixel in real CT image, and $sCT(i)$ is the ith pixel in the synthetic sCT.

To further evaluate the accuracy of synthetic CTs in different tissues, the threshold of 300 HU was used on the true CT images to separate the bone and soft tissues. The MAEs and MEs in bone and soft tissues were calculated, respectively.

RESULTS

Comparison of True CT and Synthetic CT Images

An example of the T2-weighted MRI, true CT-sim, synthetic CT, and MAE differences in the axial view of two representative slices was shown in the first to fourth column in **Figure 3**. Soft

tissues in the synthetic CT (**Figures 3C,G**) had similar intensities as the true CT (**Figures 3B,F**). The major difference between true CTs and synthetic CTs was in the air-bone and bone-soft tissue interface (**Figures 3D,H**: the MAE map).

Figures 4, **5** showed the axial view for bone and soft tissues, respectively. The bone structures in synthetic CTs was well-reconstructed by our model, such as the nasal bone (**Figure 4E**) and bone marrow (**Figures 4B,E**). The soft tissues in synthetic CTs had the similar intensity as the real ones (**Figures 5B,E**). However, the interface between bone and soft tissues had higher deviation, and the delicate structures in nasal cavity were blurred in the synthetic CTs (**Figure 5B**). The majority soft tissue and bone region was reconstructed accurately except the interface between soft tissue and bone and some delicate structures in nasal cavity, where the inaccuracy might be induced by imperfect deformable registration.

Quantitative Analysis

The summary of HU difference between the true CT and synthetic CT images was listed in **Table 1**. Compared to true CT, MAE of sCT in the 10 tested patients was 97 ± 13 HU in soft tissue, 131 ± 24 HU in overall region, and 357 ± 44 HU in bone, respectively. ME was −48 ± 10 HU in soft tissue, −6 ± 13 HU in overall region, and 247 ± 44 HU in
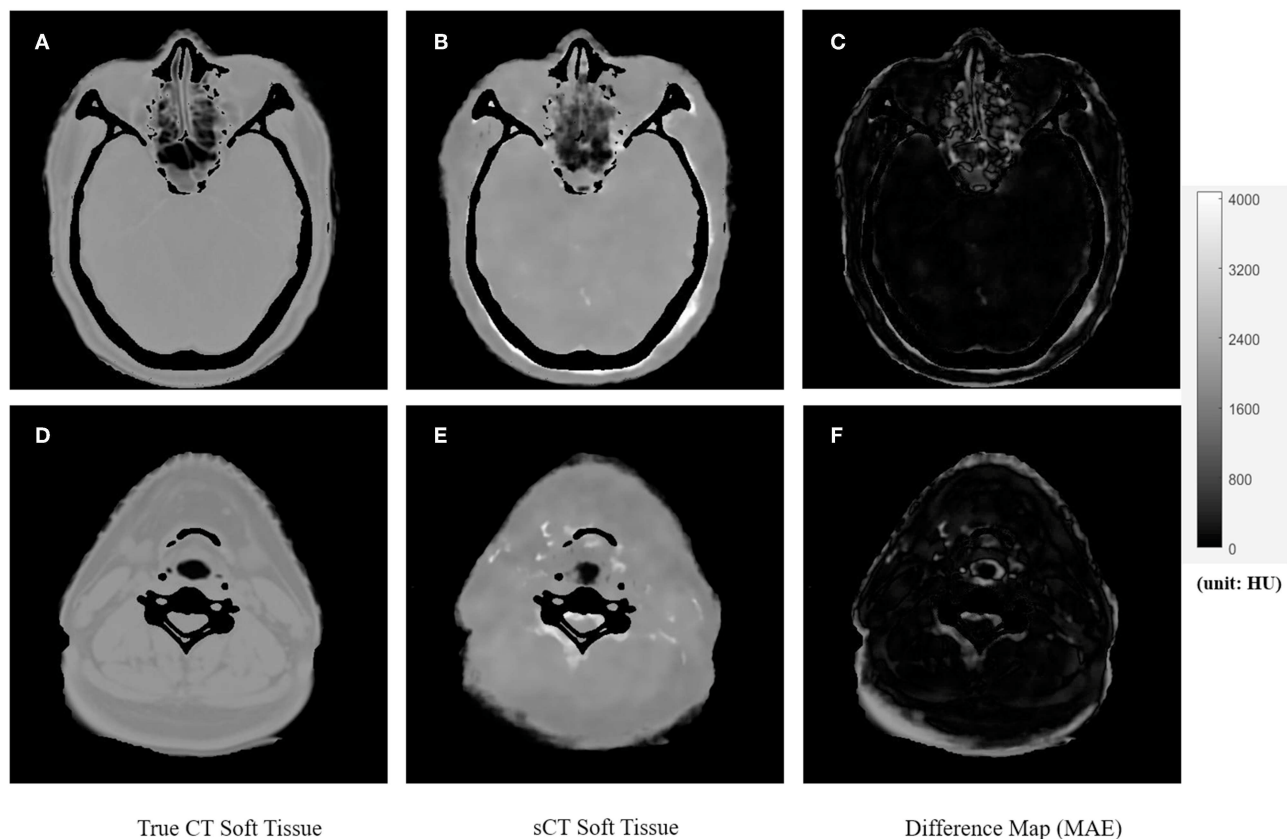
FIGURE 5 | The comparison of true CT and synthetic CT for soft tissues. The first column is true CT images. The second column is the synthetic CT images. The third column is the difference maps. **(A–C)** Showed soft tissue in the head and nasal region. **(D–F)** Showed soft tissues in the neck region. Gray bar indicated the mapping from CT number to gray scale in the difference maps.

bone, respectively. As shown in **Table 1**, MAE and ME varied in different patients. The synthetic CTs of Patient #1 had the lowest deviation in overall body, bone, and soft tissues (overall body: MAE = 91; bone: MAE = 300; soft tissue: MAE = 75; unit: HU). The synthetic CTs of Patient #3 had the largest deviations in overall body, bone, and soft tissues (overall body: MAE = 170; bone: MAE = 430; soft tissue: MAE = 118; unit: HU).

We also calculated ME to evaluate the average errors of each patient. In most patients (patient #1, 2, 4, 5, 6, 7, 9, 10), the synthetic CTs overestimated the CT number in the overall body region. Only in 2 patients (patient #3, 8), the CT number in synthetic CTs were underestimated, especially in the bone region. We noted that the CT number of bones in synthetic CTs was underestimated, while CT number of the soft tissues was overestimated using our U-net algorithm.

The GPU-based U-net model was trained with 23 patients' datasets using 20 h. The average time for each test patient was only 7 s. The total time of converting T2-weighted MRI to sCT for all 10 test patients using our deep learning algorithm was <1 min.

DISCUSSION

We have developed a feasible deep learning algorithm for converting MRI to HU maps to facilitate the MR-only treatment planning in the future. Based on the performance metrics such as MAE and ME, our soft tissue and overall region had acceptable HU differences. However, the bone region had larger errors due to less pixels of bone area compared to those of soft tissue and hence much less samples for training. In addition, bone regions have a large range of HU values, typically from several hundreds to several thousand HU numbers, which makes the training more difficult than the narrower range of HU numbers in soft tissue. One way to improve the results in the bone region is to separately train soft tissue and bone (30); another approach is to acquire ultrashort TE (UTE) MRI sequence to obtain better labeling of the bone region in MR images (31).

As mentioned in the previous review articles by Edmund et al. (15), there is no obvious favorable method among different types of MRI contrast(s) in the generation of synthetic CT to increase the accuracy. The reason we use the 2D images of T2-weighted MRI to generate synthetic CT images is simply due to its popularity in the existing radiotherapy workflow for target

TABLE 1 | Summary of all 10 test patients.

	MAE: soft tissue	MAE: bone	MAE: overall region	ME: soft tissue	ME: bone	ME: overall region
Patient 1	75	300	91	−32	191	−15
Patient 2	98	323	126	−58	178	−26
Patient 3	118	430	170	−49	312	15
Patient 4	109	370	145	−55	279	−8
Patient 5	79	303	100	−34	227	−6
Patient 6	104	369	136	−39	258	−1
Patient 7	93	343	120	−43	242	−9
Patient 8	93	421	158	−61	293	12
Patient 9	105	363	137	−50	274	−5
Patient 10	95	352	128	−57	217	−20
Mean ± SD	97 ± 13	357 ± 44	131 ± 24	−48 ± 10	247 ± 44	−6 ± 13

delineation. In our study, it took 20 h to train the U-net model with 23 patients' MRI and CT datasets. The average time for each test patient was only 7 s. The total time of converting T2-weighted MRI to sCT for all 10 test patients using our deep learning algorithm was <1 min, which has great clinical potential for online MRI conversion in the future.

It has been noticed by Edmund et al. (15) that the current performance metrics such as MAE and Dice do not reflect the corresponding dosimetric and geometrical agreement between the true CT and synthetic CT. Therefore, more unambiguous metrics should be developed, where the results should not depend on the selected CT number threshold (for example, our study used HU = 300 as the threshold for bone and soft tissue). Another concern of the synthetic CT methods is about the clinical implementation to the existing RT workflow. For the brain, it has been shown that a bulk density assignment may be sufficient for RT treatment planning (32). However, the head and neck region is more challenging in planning with many close-orientated organs at risk (OAR). Therefore, we may need more accurate HU maps in the conversion using the pixel-based deep learning method. We noticed there were underestimations in bones and overestimations in soft tissues. The use of L2 distance (mean squared error) as the loss function could cause the image blurring, which tended to predict an average CT value of both bone and soft tissues. The low prediction accuracy in the interface could be due to the errors of image registration and suboptimal prediction model. To encourage less blurring and improve the prediction accuracy, the L1 distance and a more complicated neural network with more fitting parameters could be introduced.

We have noticed several limitations in this study. First, the co-registration of MRI and CT-sim images may introduce systematic errors. It has been reported that MRI-CT co-registration may introduce geometrical uncertainties of ~2 mm for the brain and neck region (13) and of 2–3 mm for prostate and gynecological patients (14). Although our MRI and CT were acquired within the same week and similar scan position, the T2-weighted MRI was acquired in the department of diagnostic radiology without head and neck masks and without the flat couch top, the patients' chin position of CT-sim was still slightly different from that of MRI. Therefore, the rigid and deformable registration using

the open source software could introduce geometrical errors, which makes the U-net downstream more difficult to accurately map HU values pixel-by-pixel. Furthermore, MRI has more geometrical distortion inherently compared to CT due to its gradient non-linearity and magnetic field inhomogeneity (33). In addition, patients inside the MRI bore can introduce geometrical distortion from susceptibility effect and chemical shift, which is difficult to correct. The traditional way of applying MRI to radiation treatment planning (RTP) is to acquire diagnostic MRI and then to conduct deformable image registration of MRI to the planning CT. The patient position of diagnostic MRI scans may be different from that of CT-sim or treatment position, which can introduce systematic errors (3, 34). Therefore, in order to minimize error and increase accuracy of deep learning-based MRI conversion to CT, we should use the MRI simulation with exactly the same immobilization devices as the CT simulation, which will be possible in 6 months when we have an in-department new MRI simulator.

The second limitation lies in the U-net deep learning algorithm. Deep learning algorithms are widely available, such as deep convolutional network (what we used), recurrent neural network (RNN), deep residual network (DRN), generative adversarial network (GAN), long/short term memory (LSTM). However, they may be susceptible to overfitting, difficult to interpret, or issues of accuracy. It has been reported that the deep CNN method competed favorably compared to the atlas-based method in the MRI conversion process (29). Here we used U-net CNN in the synthetic CT generation from T2-weighted MRI. However, U-net only interprets the non-linear mapping between MR and CT images through the training process. GAN, for example, has great potential to develop a better understanding of the non-linear relationship by generating images and improving the output through the discriminative algorithm (35). In the future, the structure of deep learning networks can be optimized to enhance accuracy and reduce the non-linear mapping error in the MRI conversion of CT numbers.

The third limitation is the sample size. It has been observed in our study that increasing the sample size can significantly improve the image quality and accuracy of the synthetic CT. For example, we started with 13 patients as the training set and later increased the sample size to 23 patients in the training set. The
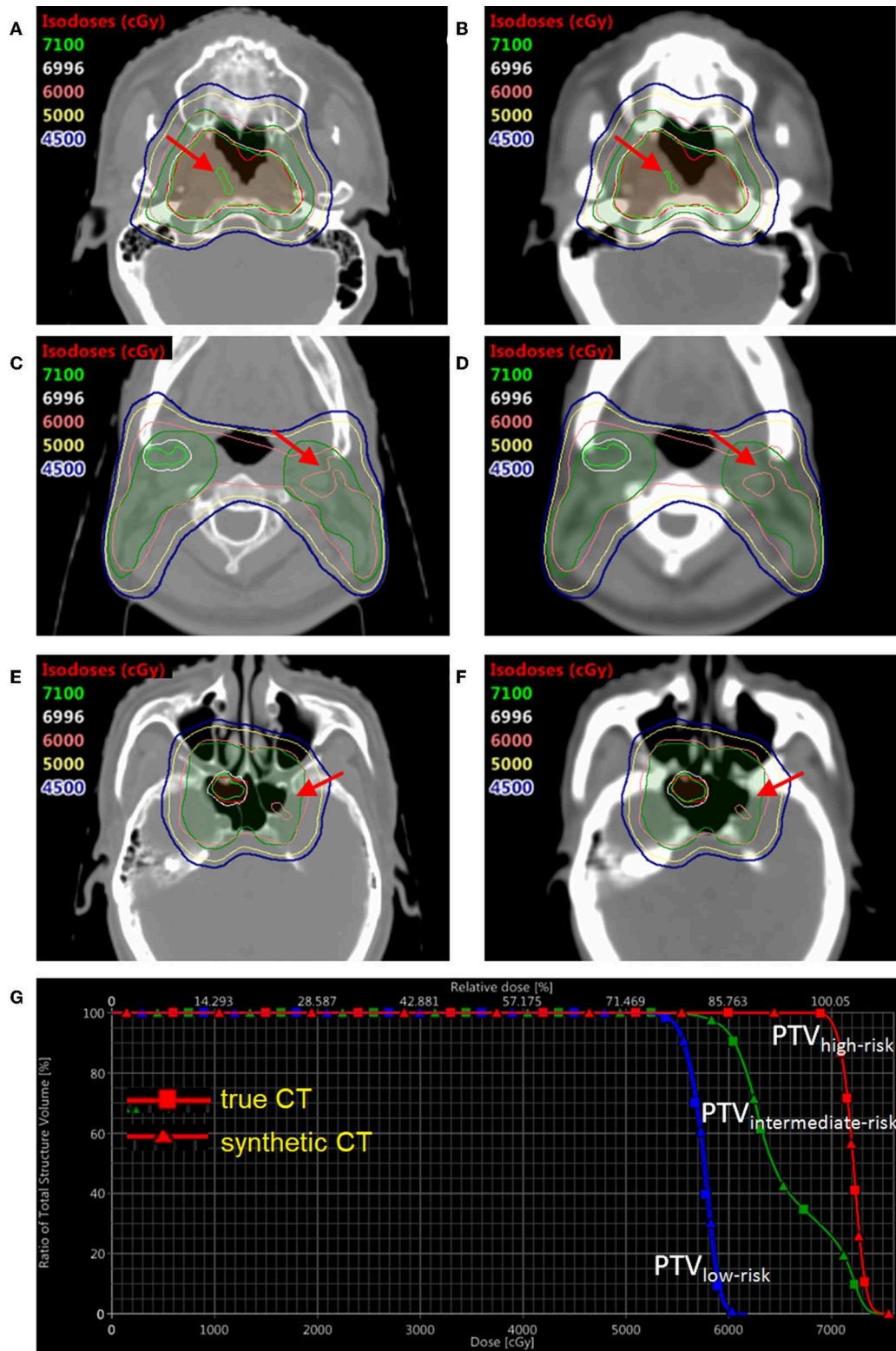
FIGURE 6 | The dose distribution and DVH comparison between true CT and synthetic CT in an example NPC patient. Three slices of true CT are shown on the left (A,C,E); Three corresponding slices of synthetic CT are shown on the right (B,D,F). The DVH comparison based on true CT and synthetic CT is shown in (G). The translucent red region is high-risk PTV with prescription dose of 69.96 Gy; the translucent green region is intermediate-risk PTV with prescription dose of 60 Gy; the translucent blue region (not shown in axial views here) is low-risk PTV with prescription dose of 54.4 Gy. DVH with squares and triangles is based on true and synthetic CT, respectively.

MAE of HU difference map has been decreased significantly. Also, we didn't use image augmentation to increase data sets in our study, which may help to improve the accuracy of MRI conversion to synthetic CT.

In order to compare the dose distribution using true vs. synthetic CT, one patient example was selected with the tumor in the sinus area and nearby fine bone structure (**Figure 6**). The mean HU value difference between true CT and sCT in the bone region was 191. The mean difference between true CT and sCT in the soft tissue region was 32. The treatment plan using the true CT was constructed with two full RapidArc in the Eclipse TPS v13.5 (Varian Medical Systems) and clinically approved by radiation oncologists. The dose distribution was subsequently recalculated based on sCT in the same treatment planning system. The three PTV regions, which were high-risk, intermediate-risk, and low-risk PTVs, as shown in DVH and isodose lines in **Figure 6**, had almost no difference between true and synthetic CT. For instance, the difference of D98% between the high-risk, intermediate-risk, and low-risk PTVs using true CT and sCT was <1%.

In summary, a promising method of synthetic CT generated from MRI has been proposed. Our pixel-based U-net deep learning algorithm of converting T2-weighted 2D MRI to HU mapping shows clinical potential of feasibility and simplicity with acceptable accuracy in soft tissue and overall region in the nasopharyngeal cancer site, which can be improved in the future by increasing the sample size of training data, acquiring same setup position of CT-sim vs. MRI-sim, and applying advanced neural networks such as GAN for better non-linear mapping.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/supplementary material.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by IRB, CAMS Shenzhen Cancer Hospital. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

YW conceived the idea, collected data, and wrote the manuscript. CL developed the U-net coding. XZ debugged the program and analyzed results. WD checked results and revised the manuscript.

REFERENCES

1. McGee KP. MRI in radiation oncology: underserved needs. *Magn Reson Med*. (2016) 75:11–4. doi: 10.1002/mrm.25826

2. Dirix P, Haustermans K, Vandecaveye V. The value of magnetic resonance imaging for radiotherapy planning. *Semin Radiat Oncol*. (2014) 24:151–9. doi: 10.1016/j.semradonc.2014.02.003

3. Dawson LA, Menard C. Imaging in radiation oncology: a perspective. *Oncologist*. (2010) 15:338–49. doi: 10.1634/theoncologist.2009-S106

4. Haie-Meder C, Pötter R, Van Limbergen E, Briot E, De Brabandere M, Dimopoulos J, et al. Recommendations from Gynaecological (GYN) GEC-ESTRO Working Group (I): concepts and terms in 3D image-based 3D treatment planning in cervix cancer brachytherapy with emphasis on MRI assessment of GTV and CTV. *Radiother Oncol*. (2005) 74:235–45. doi: 10.1016/j.radonc.2004.12.015

5. Voroney JP, Brock KK, Eccles C, Haider M, Dawson LA. Prospective comparison of computed tomography and magnetic resonance imaging for liver cancer delineation using deformable image registration. *Int J Radiat Oncol Biol Phys*. (2006) 66:780–91. doi: 10.1016/j.ijrobp.2006.05.035

6. Emami B, Sethi A, Petruzzelli GJ. Influence of MRI on target volume delineation and IMRT planning in nasopharyngeal carcinoma. *Int J Radiat Oncol Biol Phys*. (2003) 57:481–8. doi: 10.1016/S0360-3016(03)00570-4

7. Steenbakkers RJ, Deurloo KE, Nowak PJ, Lebesque JV, van Herk M, Rasch CR. Reduction of dose delivered to the rectum and bulb of the penis using MRI delineation for radiotherapy of the prostate. *Int J Radiat Oncol Biol Phys*. (2003) 57:1269–79. doi: 10.1016/S0360-3016(03)01446-9

8. Muruganandham M, Clerkin PP, Smith BJ, Anderson CM, Morris A, Capizzano AA, et al. 3-Dimensional magnetic resonance spectroscopic imaging at 3 Tesla for early response assessment of glioblastoma patients during external beam radiation therapy. *Int J Radiat Oncol Biol Phys*. (2014) 90:181–9. doi: 10.1016/j.ijrobp.2014.05.014

9. Moffat BA, Chenevert TL, Lawrence TS, Meyer CR, Johnson TD, Dong Q, et al. Functional diffusion map: a noninvasive MRI biomarker for early stratification of clinical brain tumor response. *Proc Natl Acad Sci USA*. (2005) 102:5524–9. doi: 10.1073/pnas.0501532102

10. Padhani AR. Diffusion magnetic resonance imaging in cancer management. *Semin Radiat Oncol*. (2011). 21:119–40. doi: 10.1016/j.semradonc.2010.10.004

11. Plewes DB, Kucharczyk W. Physics of MRI: a primer. *J Magn Reson Imaging*. (2012) 35:1038–54. doi: 10.1002/jmri.23642

12. Owrangi AM, Greer PB, Glide-Hurst CK. MRI-only treatment planning: benefits and challenges. *Phys Med Biol*. (2018) 63:05TR01. doi: 10.1088/1361-6560/aaaca4

13. Ulin K, Urie MM, Cherlow JM. Results of a multi-institutional benchmark test for cranial CT/MR image registration. *Int J Radiat Oncol Biol Phys*. (2010) 77:1584–9. doi: 10.1016/j.ijrobp.2009.10.017

14. Wang D, Doddrell D. Geometric distortion in structural magnetic resonance imaging. *Curr Med Imaging Rev*. (2005) 1:49–60. doi: 10.2174/1573405052953029

15. Edmund JM, Nyholm T. A review of substitute CT generation for MRI-only radiation therapy. *Radiat Oncol*. (2017) 12:28. doi: 10.1186/s13014-016-0747-y

16. Johnstone E, Wyatt JJ, Henry AM, Short SC, Sebag-Montefiore D, Murray L, et al. Systematic review of synthetic computed tomography generation methodologies for use in magnetic resonance imaging-only radiation therapy. *Int J Radiat Oncol Biol Phys*. (2018) 100:199–217. doi: 10.1016/j.ijrobp.2017.08.043

17. Dowling JA, Lambert J, Parker J, Salvado O, Fripp J, Capp A, et al. An atlas-based electron density mapping method for magnetic resonance imaging (MRI)-alone treatment planning and adaptive MRI-based prostate

radiation therapy. *Int J Radiat Oncol Biol Phys.* (2012) 83: e5–11. doi: 10.1016/j.ijrobp.2011.11.056

18. Kim J, Glide-Hurst C, Doemer A, Wen N, Movsas B, Chetty IJ. Implementation of a novel algorithm for generating synthetic CT images from magnetic resonance imaging data sets for prostate cancer radiation therapy. *Int J Radiat Oncol Biol Phys.* (2015) 91:39–47. doi: 10.1016/j.ijrobp.2014.09.015

19. Tyagi N, Fontenla S, Zhang J, Cloutier M, Kadbi M, Mechalakos J, et al. Dosimetric and workflow evaluation of first commercial synthetic CT software for clinical use in pelvis. *Phys Med Biol.* (2017) 62:2961–75. doi: 10.1088/1361-6560/aa5452

20. Bibault JE, Giraud P, Burgun A. Big Data and machine learning in radiation oncology: state of the art and future prospects. *Cancer Lett.* (2016) 382:110–7. doi: 10.1016/j.canlet.2016.05.033

21. El Naqa I, Brock K, Yu Y, Langen K, Klein EE. On the fuzziness of machine learning, neural networks, and artificial intelligence in radiation oncology. *Int J Radiat Oncol Biol Phys.* (2018) 100:1–4. doi: 10.1016/j.ijrobp.2017.06.011

22. Nyúl LG, Udupa JK Zhang X. New variants of a method of mri scale standardization. *IEEE Trans Med Imaging.* (2000)19:143–50. doi: 10.1109/42.836373

23. Shah M, Xiao Y, Subbanna N, Francis S, Arnold DL, Collins DL, et al. Evaluating intensity normalization on mris of human brain with multiple sclerosis. *Med Image Anal.* (2011) 15:267–82. doi: 10.1016/j.media.2010.12.003

24. Klein S, Staring M, Murphy K, Viergever MA, Pluim JPW. Elastix: a toolbox for intensity-based medical image registration. *IEEE Trans Med Imaging.* (2010) 29:196–205. doi: 10.1109/TMI.2009.2035616

25. Shamonin DP, Bron EE, Lelieveldt BP, Smits M, Klein S, Staring M, et al. Fast parallel image registration on CPU and GPU for diagnostic classification of Alzheimer's disease. *Front Neuroinform.* (2014) 7:50. doi: 10.3389/fninf.2013.00050

26. Leibfarth S, Mönnich D, Welz S, Siegel C, Schwenzer N, Schmidt H, et al. A strategy for multimodal deformable image registration to integrate PET/MR into radiotherapy treatment planning. *Acta Oncol.* (2013) 52:1353–9. doi: 10.3109/0284186X.2013.813964

27. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention.* Munich: Springer (2015). p. 234–41.

28. Otsu N. A threshold selection method from gray-level histograms. *IEEE Trans Syst Man Cybern.* (1979) 9:62–6. doi: 10.1109/TSMC.1979.4310076

29. Han X. MR-based synthetic CT generation using a deep convolutional neural network method. *Med Phys.* (2017) 44:1408–19. doi: 10.1002/mp.12155

30. Korhonen J, Kapanen M, Keyriläinen J, Seppälä T, Tenhunen M. A dual model HU conversion from MRI intensity values within and outside of bone segment for MRI-based radiotherapy treatment planning of prostate cancer. *Med Phys.* (2014) 41:011704. doi: 10.1118/1.4842575

31. Jang H, Liu F, Zhao G, Bradshaw T, McMillan AB. Technical Note: deep learning based MRAC using rapid ultra-short echo time imaging. *Med Phys.* (2018) 45:3697–704. doi: 10.1002/mp.12964

32. Jonsson JH, Karlsson MG, Karlsson M, Nyholm T. Treatment planning using MRI data: an analysis of the dose calculation accuracy for different treatment regions. *Radiat Oncol.* (2010) 5:62. doi: 10.1186/1748-717X-5-62

33. Devic S. MRI simulation for radiotherapy treatment planning. *Med Phys.* (2012) 39:6701–11. doi: 10.1118/1.4758068

34. Schmidt MA, Payne GS. Radiotherapy planning using MRI. *Phys Med Biol.* (2015) 60: R323–61. doi: 10.1088/0031-9155/60/22/R323

35. Kazemifar S, McGuire S, Timmerman R, Wardak Z, Nguyen D, Park Y, et al. MRI-only brain radiotherapy: assessing the dosimetric accuracy of synthetic CT images generated using a deep learning approach. *Radiother Oncol.* (2019) 136:56–63. doi: 10.1016/j.radonc.2019.03.026

ORIGINAL RESEARCH

Quantitative MRI Changes During Weekly Ultra-Hypofractionated Prostate Cancer Radiotherapy With Integrated Boost

Marcel A. van Schie[1]*, Petra J. van Houdt[1], Ghazaleh Ghobadi[1], Floris J. Pos[1], Iris Walraven[1], Hans C. J. de Boer[2], Cornelis A. T. van den Berg[2], Robert Jan Smeenk[3], Linda G. W. Kerkmeijer[2,3] and Uulke A. van der Heide[1]

[1] Department of Radiation Oncology, The Netherlands Cancer Institute, Amsterdam, Netherlands, [2] Department of Radiation Oncology, University Medical Center Utrecht, Utrecht, Netherlands, [3] Department of Radiation Oncology, Radboud University Medical Center, Nijmegen, Netherlands

Purpose: Quantitative MRI reflects tissue characteristics. As possible changes during radiotherapy may lead to treatment adaptation based on response, we here assessed if such changes during treatment can be detected.

Methods and Materials: In the hypoFLAME trial patients received ultra-hypofractionated prostate radiotherapy with an integrated boost to the tumor in 5 weekly fractions. We analyzed T2 and ADC maps of 47 patients that were acquired in MRI exams prior to and during radiotherapy, and performed rigid registrations based on the prostate contour on anatomical T2-weighted images. We analyzed median T2 and ADC values in three regions of interest (ROIs): the central gland (CG), peripheral zone (PZ), and tumor. We analyzed T2 and ADC changes during treatment and compared patients with and without hormonal therapy. We tested changes during treatment for statistical significance with Wilcoxon signed rank tests. Using confidence intervals as recommended from test-retest measurements, we identified persistent T2 and ADC changes during treatment.

Results: In the CG, median T2 and ADC values significantly decreased 12 and 8%, respectively, in patients that received hormonal therapy, while in the PZ these values decreased 17 and 18%. In the tumor no statistically significant change was observed. In patients that did not receive hormonal therapy, median ADC values in the tumor increased with 20%, while in the CG and PZ no changes were observed. Persistent T2 changes in the tumor were found in 2 out of 24 patients, while none of the 47 patients had persistent ADC changes.

Conclusions: Weekly quantitative MRI could identify statistically significant ADC changes in the tumor in patients without hormonal therapy. On a patient level few persistent T2 changes in the tumor were observed. Long-term follow-up is required to relate the persistent T2 and ADC changes to outcome and evaluate the applicability of quantitative MRI for response based treatment adaptation.

Keywords: quantitative MRI, ultra-hypofractionated prostate radiotherapy, MRI changes, T2 mapping, ADC mapping, hormonal therapy

INTRODUCTION

Whole gland dose escalation for prostate cancer has shown to result in increased biochemical control rates, but is associated with increased toxicity (1). Focal dose escalation may benefit patient outcome without compromising toxicity levels compared to conventional treatment. This hypothesis is currently tested in the FLAME trial (2) where patients received an integrated boost up to 95 Gy to the visible tumor in addition to a whole gland dose of 77 Gy in 35 treatment fractions. With advancing insight in prostate tumor radiobiology, hypofractionated prostate radiotherapy is increasingly performed (3, 4). With ultra-hypofractionation, the therapeutic ratio between tumor control and toxicity increases even further due to the low α/β ratio of prostate cancer. Several ultra-hypofractionation trials have demonstrated similar toxicity as compared to standard fractionation, with reduced treatment time (5–8). Also, non-inferiority has already been demonstrated (7, 8). For intermediate to high-risk disease, the combination of ultra-hypofractionation with a focal dose escalation to the tumor as conducted in the FLAME trial may even result in better outcomes. Therefore, ultra-hypofractionation was combined with a focal boost to the tumor to treat intermediate to high-risk prostate cancer in the hypoFLAME trial.

In prostate cancer long term follow-up of at least 5 years is required to evaluate treatment outcome. If changes in the prostate occur at an early stage during treatment and are related to outcome, treatment adaptation for prostate cancer could be considered.

Quantitative MRI is known to reflect tissue characteristics. Diffusion weighted imaging (DWI) and T2 mapping are suitable quantitative MRI techniques to investigate tissue properties in the prostate (9, 10). Through DWI a quantitative apparent diffusion coefficient (ADC) map can be obtained that represents water diffusion between cells and allows to discriminate between malignant and benign prostate tissue. Furthermore, the ADC value of tumor tissue was found to relate to aggressiveness of the disease (11). With T2 mapping a spatial distribution of T2 values can be calculated that are unique to biological tissues. T2 was for example found to correlate with hypoxia (12, 13). Since prostate tumors have different properties from benign prostate tissue, T2 mapping has the potential to discriminate between benign and malignant tissue.

Since quantitative MRI reflects tissue characteristics, tissue changes due to treatment may be visible on quantitative MRI as well. Therefore, quantitative MRI has the potential to generate imaging biomarkers for treatment response assessment. Before investigating this potential role for quantitative MRI, the first step is to identify if any changes in the tumor during treatment can be detected on quantitative MRI.

To identify changes in the prostate during treatment, in the hypoFLAME trial we acquired quantitative MRI data at each weekly fraction of radiation and tracked quantitative MRI values during the course of treatment. Since concurrent hormonal therapy may affect these MRI values (14), we also investigated the influence of hormonal therapy on tissue changes during radiotherapy.

METHODS AND MATERIALS

Patient Characteristics

We collected data of 73 patients from two institutions who participated in the hypoFLAME trial (clinicaltrials.gov NCT02853110). All patients had biopsy-proven, clinically localized, intermediate to high-risk prostate cancer (15). Patients were excluded if they had a contraindication for performing an MRI examination, if no tumor nodule was visible on MRI or if placement of fiducial markers was unsafe. Other exclusion criteria were ≥5 mm seminal vesicle invasion, lymph node or distant metastasis, or an iPSA of more than 30 ng/mL. Also patients that received previous pelvic irradiation or underwent transurethral resection of the prostate (TURP), or patients with an International Prostate Symptom Score (IPSS) > 15 or an World Health Organization (WHO) >2 were not included in the trial. We obtained approval from the institutional review boards and written informed consent from all included patients.

Treatment Delivery

Patients were treated in the University Medical Center in Utrecht (UMCU, $n = 36$) and the Netherlands Cancer Institute in Amsterdam (NKI, $n = 37$). Dual-arc VMAT treatment was delivered once per week with 35 Gy in five fractions to the prostate, with an integrated focal boost up to 50 Gy to the visible tumor on MRI. Position verification of the prostate was performed prior to each radiation fraction using gold fiducial markers visible on cone-beam CT. In the UMCU 10 out of 36 patients received concurrent hormonal therapy for a period of 6–36 months, in the NKI these were 31 out of 37 patients. Hormonal therapy was typically started 2–6 weeks prior to the start of radiotherapy.

Scanning Protocol

Prior to treatment patients received a planning CT scan and MRI exam, including a T2-weighted scan and a diffusion weighted imaging (DWI) scan. In the NKI also a T2 mapping sequence was performed. In both institutions patients were scanned on a 3T Philips Ingenia MRI scanner. Specifications of the scanned MRI sequences are listed in **Table 1**. To track changes in the prostate and tumor during treatment, a weekly repeat MRI exam was scanned at each treatment fraction that included the same image sequences as the pretreatment MRI exam.

Calculation of T2 and ADC Maps

The DWI scans were acquired with different protocols as described in **Table 1**. For consistency between institutions we only considered b-values between 200 and 800 s/mm^2. In the NKI cohort we calculated the ADC maps using b = 200 and 800 s/mm^2, in the UMCU cohort we calculated the ADC maps using b = 300, 500, and 800 s/mm^2.

In the NKI cohort we derived quantitative T2 maps from the T2 mapping sequence. For calculation of the T2 map we applied an in-house developed weighted logarithmic fitting algorithm to determine the T2 value per voxel in the image (16).

TABLE 1 | Specifications of MRI sequences in the UMCU and NKI.

	UMCU	NKI
T2-WEIGHTED (TSE)		
Voxel size (mm^3)		
Acquired	0.6 × 0.7 × 3	0.7 × 0.7 × 3
Reconstructed	0.5 × 0.5 × 3	0.4 × 0.4 × 3
FOV (mm^3)	200 × 200 × 90 /	282 × 282 × 75–90
	230 × 230 × 141–150	
TE / TR (ms)	90–100 / 3,770–8,620	120 / 3,690–7,930
T2 MAPPING (MULTI-ECHO SPIN-ECHO)		
Voxel size (mm^3)		
Acquired		0.8 × 0.8 × 3 / 1.0 × 1.0 × 3
Reconstructed		0.4 × 0.4 × 3 / 0.6 × 0.6 × 3
FOV (mm^3)		170 × 170 × 60
TE / TR (ms)		32 / 2,470–4,150
Echo spacing (ms)		16
Echoes (n)		12
DWI (SINGLE-SHOT EPI)		
Voxel size (mm^3)		
Acquired	3.0 × 3.0 × 4	2.3 × 2.4 × 3
Reconstructed	2.5 × 2.5 × 4	1.1 × 1.1 × 3
FOV (mm^3)	256 × 256 × 66	256 × 256 × 60–66
TE / TR (ms)	62–93 / 3,400–4,940	62 / 2,860–5,410
b-values (s/mm^2)	0, 100, 300, 500, 800, 1,000	0, 200, 800

FOV, field of view; TE, echo time; TR, repetition time. For T2 mapping patients were consistently scanned with one of the reported voxel sizes.

Image Registration

We registered all images to the pretreatment images to allow for tracking of prostate and tumor changes during treatment. All registrations were performed rigidly with in-house developed software using mutual information as the cost function, and registrations were manually adapted whenever required. Within each MRI exam the b = 0 s/mm^2 image from the DWI was selected, since it contained most anatomical information, and registered to the T2-weighted image. We applied the transformation matrix obtained from registration to the ADC map to register it to the T2-weighted image. From the T2 echo image series the image with echo time closest to the echo time of the T2-weighted image (TE = 120 ms) was selected and registered to the T2-weighted image. We applied the transformation matrix to the T2 map to register it to the T2-weighted image. From each repeat MRI exam we registered the T2-weighted image to the pretreatment T2-weighted image.

Delineations

We delineated the prostate and the peripheral zone on T2-weighted MRI and labeled the remaining part of the prostate as central gland (CG). The delineation of the tumor was based on multi-parametric MRI. CG, PZ, and tumor together are referred to as ROIs throughout this study.

Image Analysis

We resampled the registered images to 1 mm isotropic voxels. This allowed for exclusion of an isotropic margin of 2 mm around

TABLE 2 | Number of patients per institution from which T2 and ADC maps were available for analysis, separated by hormonal therapy (HT or No HT).

	UMCU	NKI	All
T2			
HT		21	21
No HT		3	3
All		24	24
ADC			
HT	4	24	28
No HT	15	4	19
All	19	28	47

each ROI that was considered to minimize the impact of residual registration errors. We extracted the median value within each ROI on T2 and ADC. We determined the population median value for each time point during treatment. Per patient we normalized the values to the pretreatment value to examine the relative behavior over time. We stratified by patients with and without hormonal therapy to investigate the influence on T2 and ADC changes during hypofractionated radiotherapy.

On a patient level we identified significant trends using confidence intervals for T2 and ADC defined by literature values. These confidence intervals were derived from test-retest measurements. For T2 we used a confidence interval of 11% as found by van Houdt et al. (17). For ADC we used a value of 47% as recommended by the Quantitative Imaging Biomarkers Alliance (QIBA) (18). These confidence intervals separate real changes in T2 and ADC values from measurement imprecision with 95% confidence. We subsequently determined the number of patients in which T2 and ADC changes were outside the confidence intervals at any time point during treatment and were persistent until week 5.

Statistics

We performed Wilcoxon signed rank tests to identify if changes per ROI were statistically significant during treatment. We applied a Bonferroni correction to account for multiple testing (nine tests), considering $p < 0.0056$ as significance level. All image analysis and statistical tests were performed using MATLAB (MathWorks, Natick, MA, USA).

RESULTS

We did not perform analysis on 15 patients for whom <3 out of 6 MRI exams were scanned. Eleven patients were not analyzed since they were scanned with two different DWI scanning protocols during acquisition of pretreatment and repeat MRI. We could not analyze T2 values of four patients since pretreatment T2 maps were not acquired. **Table 2** summarizes the number of patients per institution available for analysis.

The T2-weighted images, T2 and ADC maps from one patient are shown in **Figure 1** for all time points. A decrease in contrast within the prostate can be observed in all three image sequences over the course of treatment, which reduces the conspicuity of the tumor from the surrounding prostate tissue.
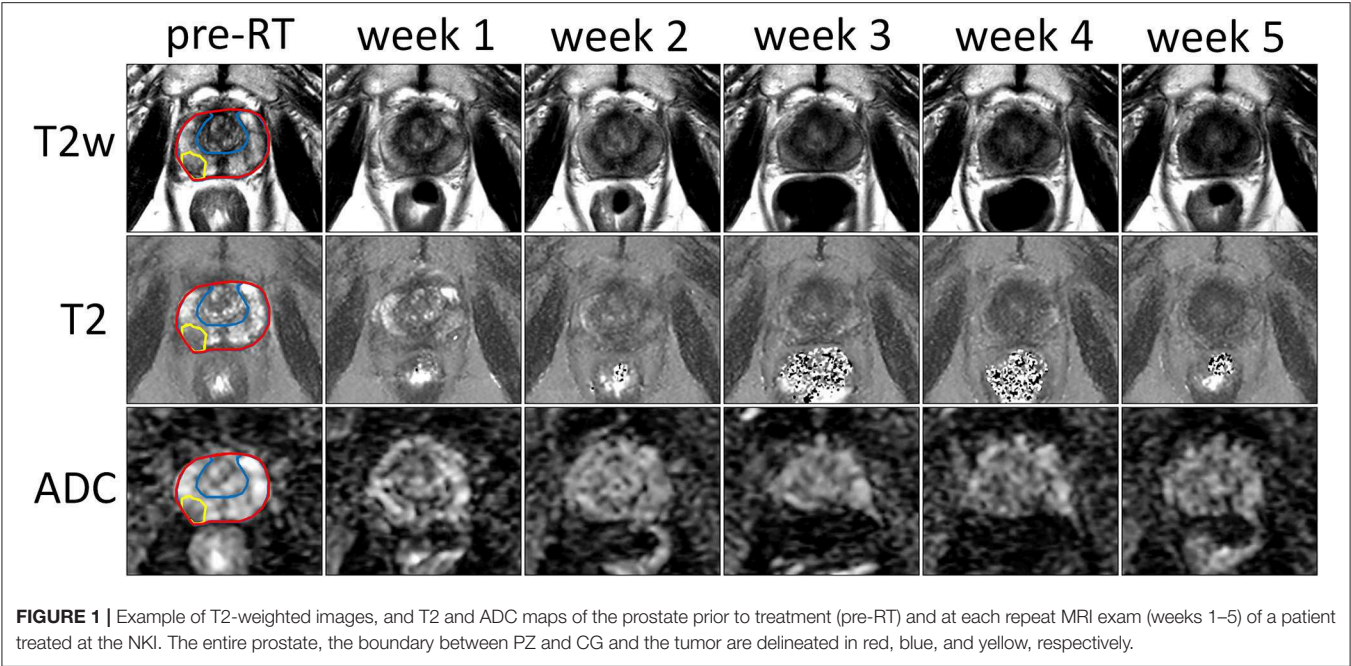
FIGURE 1 | Example of T2-weighted images, and T2 and ADC maps of the prostate prior to treatment (pre-RT) and at each repeat MRI exam (weeks 1–5) of a patient treated at the NKI. The entire prostate, the boundary between PZ and CG and the tumor are delineated in red, blue, and yellow, respectively.

TABLE 3 | Population median and interquartile range (between brackets) of median T2 (in ms) and ADC values (in 10^{-3} mm^2/s) in the CG, PZ, and tumor on pretreatment quantitative MRI.

	T2 (ms)	ADC (10^{-3} mm^2/s)	
	NKI	**UMCU**	**NKI**
CG	93 (19)	**1.30** (0.13)	**1.09** (0.18)
PZ	110 (26)	**1.37** (0.19)	**1.24** (0.24)
Tumor	80 (9)	1.07 (0.20)	0.90 (0.28)

Statistically significant differences between institutions are indicated in bold.

Median values of T2 and ADC in the CG, PZ, and tumor during pretreatment imaging are shown in **Table 3**. We observed statistically significant differences in the CG and PZ between the ADC values in the UMCU cohort and the NKI cohort.

T2 and ADC values normalized to the pretreatment values are shown in **Figure 2**. In the CG we observed a median decrease of 12% on T2 and 8% on ADC in patients that received hormonal therapy. T2 and ADC values at week 5 were significantly lower compared to pretreatment values. For patients that received no hormonal therapy, the median ADC value decreased 4% and this was not statistically significant.

In the PZ we observed similar behavior. In patients with hormonal therapy the median T2 and ADC value decreased significantly with 17 and 18%, respectively, while in patients without hormonal therapy we observed a non-significant decrease in ADC of 5%.
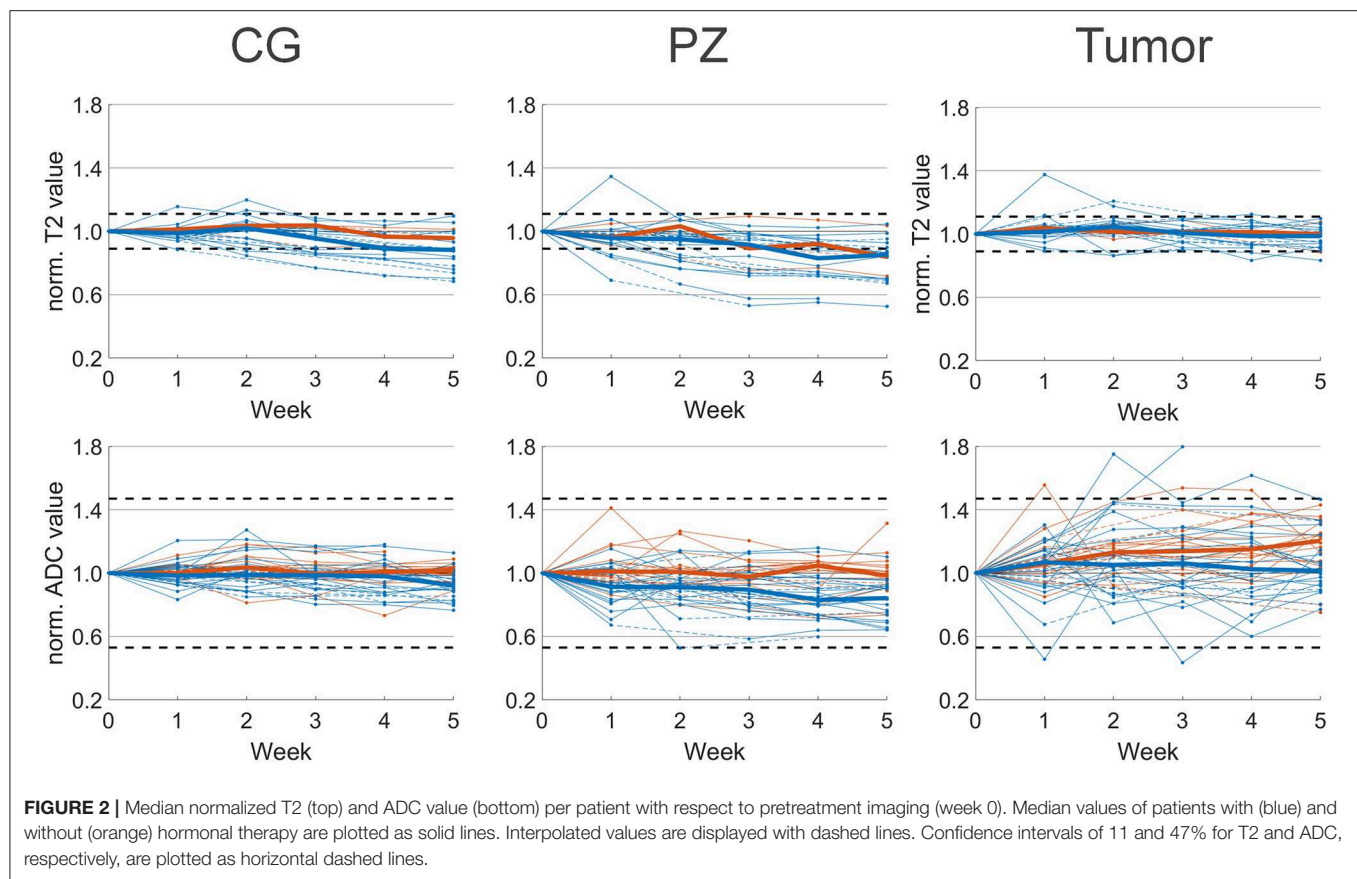
In the tumor the behavior was different from CG and PZ. Median increases of 5 and 7% on T2 and ADC maps were found for patients with hormonal therapy, and these were not statistically significant. For patients without hormonal therapy,

on ADC we observed a median increase of 20% that was statistically significant.

Due to the low number of patients that were scanned with a T2 mapping sequence and received no hormonal therapy, we did not test statistical significance of T2 changes in these patients. On an individual patient level we found that 14 out of 21 patients who received hormonal therapy, showed persistent T2 changes larger than 11% during treatment. These were 11 patients with persistent changes in the CG, 12 in the PZ and one in the tumor. For the three patients without hormonal therapy, two patients had persistent T2 changes, from which one showed changes in the CG, two in the PZ, and one in the tumor. In total 67% of the 23 patients showed persistent T2 changes during treatment. In contrast, on ADC maps for both patients with and without hormonal therapy we observed no changes outside the confidence interval of 47%.

DISCUSSION

In this study we analyzed changes in the prostate as observed on quantitative MRI during hypofractionated radiotherapy with an integrated boost to the tumor. Using repeated imaging we observed changes in median T2 and ADC values that depended on the use of hormonal therapy. The changes we observed can explain the reduced tumor conspicuity that is observed after primary radiotherapy. However, depending on hormonal therapy this can be explained by either normalization of tumor characteristics or by a decrease of normal prostate tissue values. For patients who received hormonal therapy, we observed a reduction of T2 and ADC values in the PZ, while values in the tumor did not change significantly. However, for patients who did not receive hormonal therapy, we found that ADC values increased significantly in the tumor, but not in the PZ.

FIGURE 2 | Median normalized T2 (top) and ADC value (bottom) per patient with respect to pretreatment imaging (week 0). Median values of patients with (blue) and without (orange) hormonal therapy are plotted as solid lines. Interpolated values are displayed with dashed lines. Confidence intervals of 11 and 47% for T2 and ADC, respectively, are plotted as horizontal dashed lines.

The pretreatment ADC values were significantly different between the two institutions. This may be a consequence of the DWI scanning protocols. The b-values in both protocols were similar with b = 200 and 800 s/mm^2 in the NKI and b = 300, 500, and 800 s/mm^2 in the UMCU. However, the acquisition voxel size in the UMCU protocol was 2.2 times larger than in the NKI protocol. This resulted in a different signal to noise ratio and could contribute to differences in ADC values [19].

In the literature a similar variation between ADC values was found. In one study median ADC values in the tumor of $1.08 \pm 0.39 \cdot 10^{-3}$ mm^2/s (mean $\pm$ SD) prior to treatment are reported [20]. ADC values in the untreated healthy PZ were $1.8 \pm 0.4 \cdot 10^{-3}$ mm^2/s. Other studies found values of $1.6 \pm 0.2 \cdot 10^{-3}$ mm^2/s in the healthy prostate of untreated patients [21, 22]. Again differences in DWI protocol as well as image reconstruction methods may have contributed to the existing variation.

We observed different trends in patients that did and did not receive hormonal therapy. Hormonal therapy however correlated with the institution where patients were treated. In the UMCU 4 out of the 19 patients received hormonal therapy, while in the NKI this was 24 out of the 28 patients. Because of this unbalanced distribution we could not separate hormonal therapy from institution to explain the differences in normalized ADC value behavior during treatment. This was also the reason we did not compare the T2 values

for patients with and without hormonal therapy in the NKI cohort.

One study describes prostate and tumor changes on MRI during treatment. Foltz et al. [23] reported an early treatment response in the entire prostate and CG, plus a progressive response in the PZ and tumor toward the end of treatment. A statistically significant change in the tumor was found after 6 weeks on ADC. Early treatment response in the tumor was not observed on either T2 or ADC. While there were differences in the overall treatment duration, the frequency of imaging and the time between radiotherapy fractions compared to our study. Our quantitative MRI results indicate similar behavior. We found progressive T2 changes in the PZ and late ADC changes in the tumor. This qualitative comparison is only indicative though, since the use of hormonal therapy was not reported in Foltz et al. [23].

Here we analyzed the T2 and ADC changes in prostate and tumor only during treatment. Dinis Fernandes et al. [16] reported late changes on quantitative MRI in recurrent prostate cancer patients that were scanned at least 2 years after primary treatment. Adjuvant hormonal therapy was given in 82% of the patients but ended at least 1 year before the MRI examination. Changes in CG and PZ regions on both T2 and ADC maps were found and reduced contrast between PZ and tumor on T2 maps was observed. Median T2 values in the CG, PZ and tumor decreased by 29, 19, and 5%, while we observed statistically

significant decreased values of 12 and 17% in the CG and PZ and no statistically significant change in the tumor. For ADC values a reduction of 5–9% in CG, PZ and tumor was observed 2 year after treatment. In our study we observed a decrease of 8 and 18% in the CG and PZ in case of hormonal therapy, while an increase of 20% was found in the tumor in absence of hormonal therapy. Based on these findings we expect further reduction of T2 values in the CG and PZ after treatment, as well as post-treatment changes in ADC. Also the treatment fractionation and both timing and duration of hormonal therapy may contribute to the discrepancies between both studies. Follow-up of patients in our study will be required to confirm if changes on T2 and ADC correlate with long term biochemical recurrence free survival.

We implemented a rigid registration method to align all images to the pretreatment T2-weighted image. More accurate registration methods like deformable registration could be more appropriate when registering between MRI exams. Deformable registration would account for possible deformations of the prostate between MRI exams and allow for voxel-level analysis. However, as a result of treatment we experienced intensity changes on T2-weighted images that lead to incorrect deformations we were unable to manually adapt. Therefore, we applied rigid registrations instead and minimized registration inaccuracy via removal of an isotropic margin around each ROI, which required resampling of all images. Since we performed our analysis on ROI level, we expect limited impact of both the registration method and the image resampling on our results.

Using quantitative MRI, on a population level we were able to find significant ADC changes in the intraprostatic tumors of patients that did not receive hormonal therapy during hypofractionated radiotherapy. However, early during treatment, when treatment adaptation could be considered, no significant change was identified in the tumor. We did observe only two individual patients that showed persistent T2 changes in the tumor, while no individual patients showed persistent ADC changes in the tumor. On ADC we did observe several patients with early and progressive trends in the tumor although these trends were within the confidence intervals. If these trends are continued after treatment and exceed the confidence intervals, a possible relation between early treatment response and clinical outcome could be established. Follow-up is therefore desired for assessing the potential role of quantitative MRI for adaptation of hypofractionated radiotherapy based on early treatment response.

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 the Medical Ethics Review Committee, University Medical Center, Utrecht, The Netherlands. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

MS, PH, RS, LK, and UH contributed to the conception and design of the study. PH, FP, HB, CB, LK, and UH contributed to the acquisition of data for the study. MS, PH, GG, IW, and UH contributed to the analysis of data for the study. MS and IW performed statistical analysis. MS wrote the first draft of the manuscript. MS and UH wrote sections of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

REFERENCES

1. Viani GA, Stefano EJ, Afonso SL. Higher-than-conventional radiation doses in localized prostate cancer treatment: a meta-analysis of randomized, controlled trials. *Int J Radiat Oncol Biol Phys.* (2009) 74:1405–18. doi: 10.1016/j.ijrobp.2008.10.091

2. Lips IM, van der Heide UA, Haustermans K, van Lin EN, Pos F, Franken SP, et al. Single blind randomized phase III trial to investigate the benefit of a focal lesion ablative microboost in prostate cancer (FLAME-trial): study protocol for a randomized controlled trial. *Trials.* (2011) 12:255. doi: 10.1186/1745-6215-12-255

3. Brenner DJ, Hall EJ. Fractionation and protraction for radiotherapy of prostate cancinoma. *Int J Radiat Oncol Biol Phys.* (1999) 43:1095–101. doi: 10.1016/S0360-3016(98)00438-6

4. Miralbell R, Roberts SA, Zubizarreta E, Hendry JH. Dose-fractionation sensitivity of prostate cancer deduced from radiotherapy outcomes of 5,969 patients in seven international institutional datasets: α/β = 1.4 (0.9-2.2) Gy. *Int J Radiat Oncol Biol Phys.* (2012) 82:e17–24. doi: 10.1016/j.ijrobp.2010.10.075

5. Morgan SC, Hoffman K, Loblaw DA, Buyyounouski MK, Patton C, Barocas D, et al. Hypofractionated radiation therapy for localized prostate cancer: executive summary of an ASTRO, ASCO, and AUA evidence-based guideline. *Pract Radiat Oncol.* (2018) 8:354–60. doi: 10.1016/j.prro.2018.08.002

6. Van As NJ, Brand D, Tree A, Ostler PJ, Chu W, Loblaw A, et al. PACE: analysis of acute toxicity in PACE-B, an international phase III randomized controlled trial comparing stereotactic body radiotherapy (SBRT) to conventionally fractionated or moderately hypofractionated external beam radiotherapy (CFMHRT) for localized prostate cancer (LPCa). *J Clin Oncol.* (2019) 37:1–1. doi: 10.1200/JCO.2019.37.7_suppl.1

7. Widmark A, Gunnlaugsson A, Beckman L, Thellenberg-Karlsson C, Hoyer M, Lagerlund M, et al. Extreme hypofractionation versus conventionally fractionated radiotherapy for intermediate risk prostate cancer: early toxicity results from the Scandinavian randomized phase III trial "HYPO-RT-PC". *Int J Radiat Oncol Biol Phys.* (2016) 96:938–9. doi: 10.1016/j.ijrobp.2016.09.049

8. Widmark A, Gunnlaugsson A, Beckman L, Thellenberg-Karlsson C, Hoyer M, Lagerlund M, et al. Ultra-hypofractionated versus conventionally fractionated radiotherapy for prostate cancer: 5-year outcomes of the HYPO-RT-PC randomised, non-inferiority, phase 3 trial. *Lancet.* (2019) 394:385–95. doi: 10.1016/S0140-6736(19)31131-6

9. Bonekamp D, Kohl S, Wiesenfarth M, Schelb P, Radtke JP, Götz M, et al. Radiomic machine learning for characterization of prostate lesions with MRI: comparison to ADC values. *Radiology.* (2018) 289:128–37. doi: 10.1148/radiol.2018173064

10. Mai J, Abubrig M, Lehmann T, Hilbert T, Weiland E, Grimm MO, et al. T2 mapping in prostate cancer. *Invest Radiol.* (2019) 54:146–52. doi: 10.1097/RLI.0000000000000520

11. Manetta R, Palumbo P, Gianneramo C, Bruno F, Arrigoni F, Natella R, et al. Correlation between ADC values and Gleason score in evaluation of prostate cancer: multicentre experience and review of the literature. *Gland Surg.* (2019) 8:S216–22. doi: 10.21037/gs.2019.05.02

12. Hoskin PJ, Carnell DM, Taylor NJ, Smith RE, Stirling JJ, Daley FM, et al. Hypoxia in prostate cancer: correlation of BOLD-MRI with pimonidazole immunohistochemistry-initial observations. *Int J Radiat Oncol Biol Phys.* (2007) 68:1065–71. doi: 10.1016/j.ijrobp.2007.01.018

13. Luttje MP, van Buuren LD, Luijten PR, van Vulpen M, van der Heide UA, Klomp DWJ. Towards intrinsic R2* imaging in the prostate at 3 and 7 tesla. *Magn Reson Imaging.* (2017) 42:16–21. doi: 10.1016/j.mri.2017.04.014

14. Hötker AM, Mazaheri Y, Zheng J, Moskowitz CS, Berkowitz J, Lantos JE, et al. Prostate cancer: assessing the effects of androgen-deprivation therapy using quantitative diffusion-weighted and dynamic contrast-enhanced MRI. *Eur Radiol.* (2015) 25:2665–72. doi: 10.1007/s00330-015-3688-1

15. Ash D, Flynn A, Battermann J, de Reijke T, Lavagnini P, Blank L, et al. ESTRO/EAU/EORTC recommendations on permanent seed implantation for localized prostate cancer. *Radiother Oncol.* (2000) 57:315–21. doi: 10.1016/S0167-8140(00)00306-6

16. Dinis Fernandes C, van Houdt PJ, Heijmink SWTPJ, Walraven I, Keesman R, Smolic M, et al. Quantitative 3T multiparametric MRI of benign and malignant prostatic tissue in patients with and without local recurrent prostate cancer after external-beam radiation therapy. *J Magn Reson Imaging.* (2019) 50:269–78. doi: 10.1002/jmri.26581

17. van Houdt PJ, Agarwal HK, van Buuren LD, Heijmink SWTPJ, Haack S, van der Poel HG, et al. Performance of a fast and high-resolution multi-echo spin-echo sequence for prostate T2 mapping across multiple systems. *Magn Reson Med.* (2018) 79:1586–94. doi: 10.1002/mrm.26816

18. 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:e101–21. doi: 10.1002/jmri.26518

19. Scherrer B, Gholipour A, Warfield SK. Super-resolution in diffusion-weighted imaging. *Med Image Comput Comput Assist Interv.* (2011) 14(Pt 2):124–32. doi: 10.1007/978-3-642-23629-7_16

20. Sato C, Naganawa S, Nakamura T, Kumada H, Miura S, Takizawa O, et al. Differentiation of noncancerous tissue and cancer lesions by apparent diffusion coefficient values in transition and peripheral zones of the prostate. *J Magn Reson Imaging.* (2005) 21:258–62. doi: 10.1002/jmri.20251

21. Westphalen AC, Reed GD, Vinh PP, Sotto C, Vigneron DB, Kurhanewicz J. Multiparametric 3T endorectal MRI after external beam radiation therapy for prostate cancer. *J Magn Reson Imaging.* (2012) 36:430–7. doi: 10.1002/jmri.23672

22. Kim CK, Park BK, Lee HM. Prediction of locally recurrent prostate cancer after radiation therapy: incremental value of 3T diffusion-weighted MRI. *J Magn Reson Imaging.* (2009) 29:391–7. doi: 10.1002/jmri.21645

23. Foltz WD, Wu A, Chung P, Catton C, Bayley A, Milosevic M, et al. Changes in apparent diffusion coefficient and T2 relaxation during radiotherapy for prostate cancer. *J Magn Reson Imaging.* (2013) 37:909–16. doi: 10.1002/jmri.23885

Check for updates

Dosimetric Optimization and Commissioning of a High Field Inline MRI-Linac

Urszula Jelen[1]*, Bin Dong[1], Jarrad Begg[1,2,3], Natalia Roberts[1,4], Brendan Whelan[5], Paul Keall[1,5] and Gary Liney[1,2,3,4]

[1] Department of Medical Physics, Ingham Institute for Applied Medical Research, Liverpool, NSW, Australia, [2] Liverpool Cancer Therapy Centre, Radiation Physics, Liverpool, NSW, Australia, [3] School of Medicine, University of New South Wales, Sydney, NSW, Australia, [4] Centre for Medical Radiation Physics, University of Wollongong, Wollongong, NSW, Australia, [5] Sydney Medical School, ACRF Image X Institute, University of Sydney, Sydney, NSW, Australia

Purpose: Unique characteristics of MRI-linac systems and mutual interactions between their components pose specific challenges for their commissioning and quality assurance. The Australian MRI-linac is a prototype system which explores the inline orientation, with radiation beam parallel to the main magnetic field. The aim of this work was to commission the radiation-related aspects of this system for its application in clinical treatments.

Methods: Physical alignment of the radiation beam to the magnetic field was fine-tuned and magnetic shielding of the radiation head was designed to achieve optimal beam characteristics. These steps were guided by investigative measurements of the beam properties. Subsequently, machine performance was benchmarked against the requirements of the IEC60976/77 standards. Finally, the geometric and dosimetric data was acquired, following the AAPM Task Group 106 recommendations, to characterize the beam for modeling in the treatment planning system and with Monte Carlo simulations. The magnetic field effects on the dose deposition and on the detector response have been taken into account and issues specific to the inline design have been highlighted.

Results: Alignment of the radiation beam axis and the imaging isocentre within 2 mm tolerance was obtained. The system was commissioned at two source-to-isocentre distances (SIDs): 2.4 and 1.8 m. Reproducibility and proportionality of the dose monitoring system met IEC criteria at the larger SID but slightly exceeded it at the shorter SID. Profile symmetry remained under 103% for the fields up to ~34 × 34 and 21 × 21 cm^2 at the larger and shorter SID, respectively. No penumbra asymmetry, characteristic for transverse systems, was observed. The electron focusing effect, which results in high entrance doses on central axis, was quantified and methods to minimize it have been investigated.

Conclusion: Methods were developed and employed to investigate and quantify the dosimetric properties of an inline MRI-Linac system. The Australian MRI-linac system has been fine-tuned in terms of beam properties and commissioned, constituting a key step toward the application of inline MRI-linacs for patient treatments.

Keywords: MRI-linac, commissioning, beam characterization, dosimetry, magnetic field

INTRODUCTION

Limitations of image-guidance based on MV and kV radiation beams prompted development of systems combining linear accelerators and Magnetic Resonance Imaging (MRI) scanners (1). These hybrid systems, MRI-linacs, offer superior soft tissue contrast for visualization of the tumor and of the organs at risk which can be used for daily plan adaptation and/or real-time imaging during the treatment dose delivery. Four MRI-linac designs exist to date, employing a range of magnetic field strengths and two beam-to-magnetic-field orientations: *perpendicular* (or *transverse*) and *parallel* (or *inline*) and have been recently reviewed by Liney et al. (2). Two transverse systems: Unity (Elekta, UK) (3) and MRIdian (Viewray, USA) (4) are now available commercially and used clinically, while the two inline designs: Aurora RT (MagnetTx Oncology Solutions, Canada) (5) and Australian MRI-Linac (6) are at the research prototype stage.

Unique characteristics of these systems and mutual interactions between their components pose specific challenges for their commissioning and quality assurance. The foremost is the compatibility of the dosimetric equipment with the magnetic field due to the presence of ferrous materials or unscreened mechanical or electrical components. Furthermore, the hybrid nature of the MRI-linac treatment units requires the assessment of their concurrent functionalities, for instance dose deposition during imaging, congruence of the imaging and radiation isocentres, RF interference or gantry movement effect on the magnetic field homogeneity (7). And finally, the presence of the magnetic field also affects the radiation beam generation (8, 9) and the dose deposition (10, 11).

Magnetic field influence on dose deposition is dependent on the radiation beam orientation relative to the magnetic field and on its strength. In brief, the trajectories of both the contaminant electrons as well as of the secondary electrons are altered by the Lorentz force. In transverse MRI-linacs, this causes the electron paths between collisions to become curved and results in: (1) shifted and asymmetric beam penumbra (10), (2) decreased build-up distance (10), (3) skin dose reduction within and possible increase outside the primary beam (12, 13), and (4) localized dose increase at high-to-low density interfaces due to the electron-return effect (ERE) (14). Inline MRI-linacs instead minimize or even exploit some of these effects. The Lorenz force causes the electrons to spiral around the magnetic field direction and successive energy losses in collisions lead to the shrinkage of their helical orbits (11) which results in: (1) reduction of the beam penumbra (11), (2) dose enhancement on the beam central axis (CAX), especially in low density materials (15), (3) reduction in the dose deposition perturbations due to density heterogeneities (11) and focusing of the contaminate electrons around the radiation beam axis (16, 17). In both perpendicular and parallel orientations these effects, unfamiliar in conventional radiation therapy, require characterization during commissioning.

It should be emphasized that, both the dose deposition in matter as well as the response of the dosimeters are affected by the magnetic field. The trajectories of electrons traversing their active volume change, however this change may be different in the materials constituting the detector (e.g., air cavities in ion chambers, silicon wafers in diode detectors etc.) than in the surrounding medium. As a result, the reading of the detector may not represent the dose that would be deposited in the medium in its absence. Furthermore, many detectors are not symmetric; therefore the change in their response is dependent on their orientation in the magnetic field. These effects have been observed for various types of detectors (18–22) and must be considered both in absolute (21, 23, 24) as well as in relative dosimetry (25). Additionally, air gaps present between the dosimeter and the surrounding material have been shown to influence the detector response (26, 27).

The interaction of the radiation beam and magnetic field renders the commissioning of a MRI-linac a custom task requiring adaptation of existing methods and considerate choice of dosimetric equipment. To date no guidelines on this new type of technology are available (7). For a commercial transverse MRI-linac, dosimetric (28), and imaging-oriented (7) commissioning have been described recently. Inline systems, owing to the fundamental difference in their design, have a set of specific properties which have to be addressed. For the Australian MRI-linac, a high field prototype exploring the inline configuration, the imaging performance, also in the presence of the radiation beam, has been investigated previously (29). The aim of this work was 2-fold:

(i) to demonstrate the dosimetric properties characteristic to an inline configuration; and

(ii) to commission the radiation-related aspects of the Australian MRI-linac for its application in clinical treatments, in particular:

   (a) to fine-tune the system for optimal characteristics of the radiation beam;
   (b) to characterize the its dosimetric components and the radiation beam according to international standards for medical linear accelerators; and
   (c) to acquire base data for beam modeling in the treatment planning system (TPS) and with Monte Carlo (MC) method.

MATERIALS AND METHODS

The Australian MRI-Linac

The Australian MRI-Linac consists of a dedicated open bore 1 T magnet (Agilent, UK) and a linear accelerator Linatron-MP (Varex, USA) with a stand-alone, clinical multi-leaf collimator (MLC) Millennium (Varian Medical Systems, USA). While the design permits radiation beam entry (and patient positioning) in either orientation, the current system employs inline orientation with a fixed horizontal beam and patient entry through the magnets gap. The system is not equipped with secondary collimators (jaws) and the MLC leaves travel in a horizontal (x) direction with no collimator rotation possible. Uniquely, the linac and the MLC are mounted on rails with a docking system, allowing variation of the source-to-isocentre distance (SID) between 3.2 and 1.8 m and enabling measurements at different magnetic field strengths (2, 29).

The Linatron-MP generates flattening filter free (FFF) photon beams at two nominal energies (4 and 6 MV) and the pulse repetition frequency between 50 and 400 for the 4 MV beam and 50 and 200 for the 6 MV beam. For clinical application, only the 6 MV beam and trigger rate 200 will be used and all the measurements reported in this manuscript were performed

FIGURE 1 | Layout of the Australian MRI-linac with a coordinate system originating at the system's isocentre overlaid.

with these settings. Machine output is calibrated by the vendor to deliver 1 Gy at d_{max} at 1 m distance per monitoring unit (MU) for an open field (as the linac is equipped with primary 30° conical collimator only) in ~0 T magnetic field.
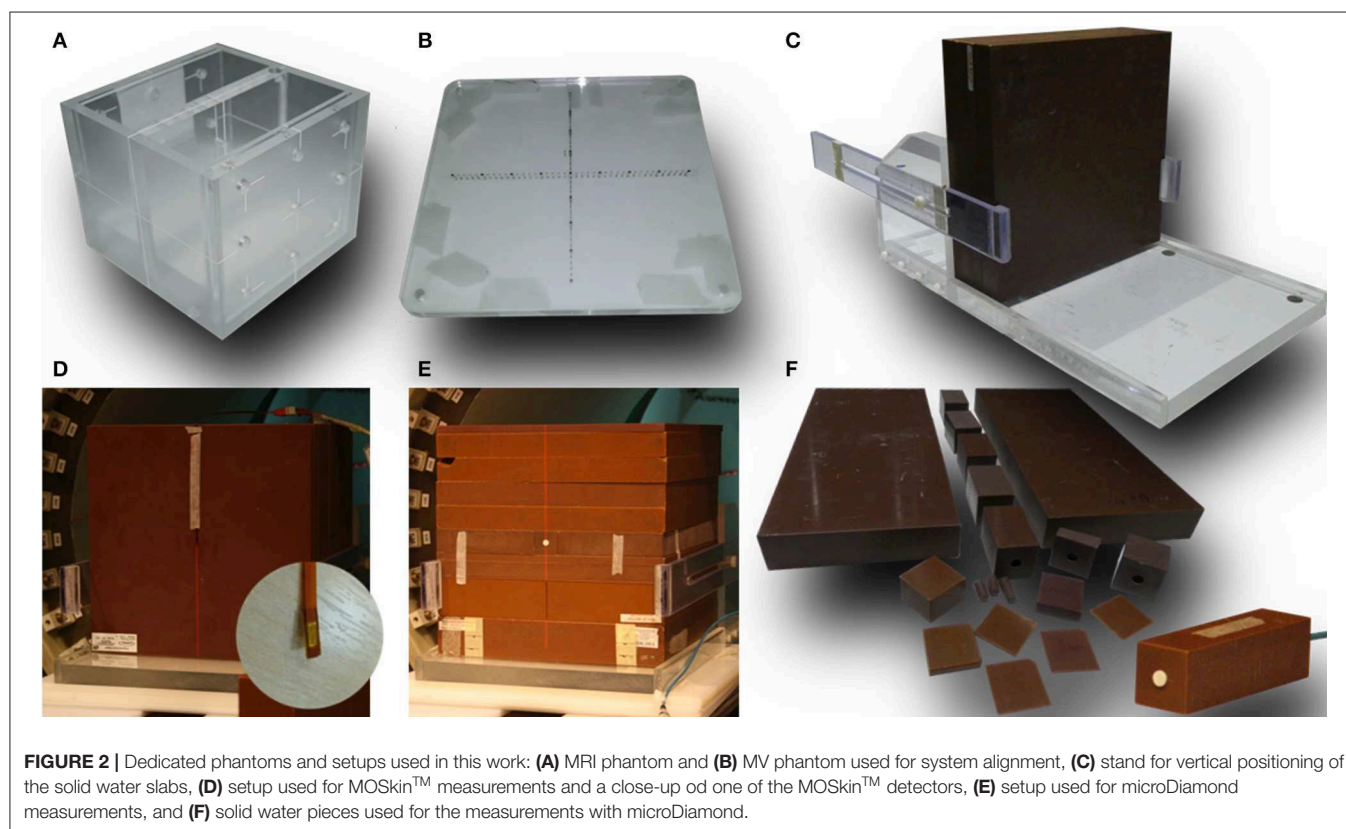
The conceptual design of the Australian MRI-linac and the coordinate system, originating at the isocentre, used in this work are shown in **Figure 1** and further details of the system can be found elsewhere (2, 16).
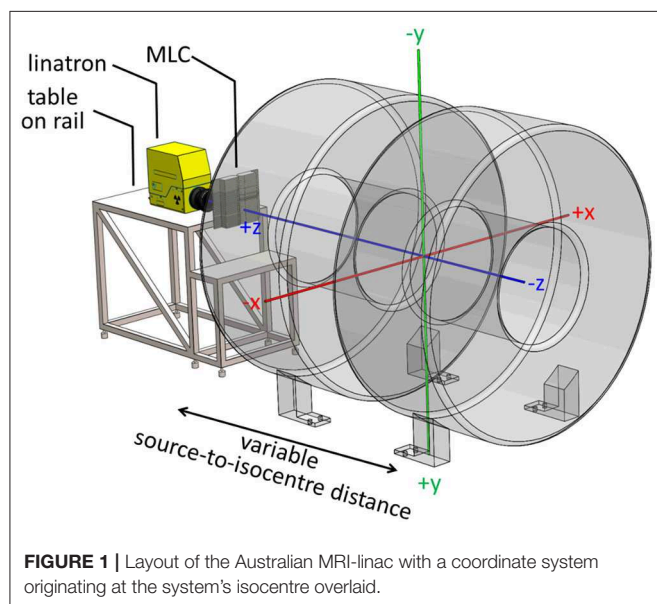
Phantoms and Detectors

For geometrical tests a combination of MRI and MV visible phantoms was used: (1) a dedicated MRI phantom (Leeds Test Objects, UK) consisting two chambers separated by 2 cm thick wall with five narrow bore holes connecting them and filled with MRI visible solution (**Figure 2A**) and (2) two acrylic plates with embedded fiducial markers for MV visibility (**Figure 2B**). A stand-alone EPID XRD 1640 AL7-M (PerkinElmer, USA) with a pixel matrix of 1,024 × 1,024 and pixel size 0.4 mm was used for the tests which involved imaging of the measurement setup components using the radiation beam. EPID images were processed using ImageJ software (National Institutes of Health, USA).

For point dose measurements a Farmer-type chamber FC65-G (Scanditronix-Wellhöfer, USA), positioned vertically either in a manual 2D water tank (for absolute does measurements) or in solid water blocks (for relative dose measurements), connected with a bias of 300 V CEP to a Unidos (PTW, Freiburg, Germany) electrometer was used. Whenever the solid water setup was used,

FIGURE 2 | Dedicated phantoms and setups used in this work: **(A)** MRI phantom and **(B)** MV phantom used for system alignment, **(C)** stand for vertical positioning of the solid water slabs, **(D)** setup used for MOSkin™ measurements and a close-up od one of the MOSkin™ detectors, **(E)** setup used for microDiamond measurements, and **(F)** solid water pieces used for the measurements with microDiamond.

the chamber holder was filled with water to avoid the presence of air gaps. The chamber and the electrometer were independently characterized for reproducibility and linearity using a well type strontium source prior to measurements. Finally, the chamber is traceable to the National Physical Laboratory (NPL, UK) and has been calibrated in a 6 MV FFF beam both in 0 T and in 1 T field (30).

For electron contamination characterization and entrance dose measurements a synthetic microDiamond 60019 (PTW, Germany) was used connected to a Unidos electrometer with a bias of 0 V was oriented with the long axis parallel to the beam. The detector's sensitive volume is 2.2 mm in diameter and 1 µm thick and the effective point of measurement (EPOM) is at 1 mm depth and has been determined to be unaffected by the magnetic field (31). While an increased angular dependence for the diamond detector response in a transverse 1.5 T field has been observed (31), which was deemed relevant for relative dosimetry at distant off-axis positions or at different gantry angles, Monte Carlo simulations indicate that this effect is minimized in inline orientation (22). For higher resolution information, these measurements were complemented with the data acquired using MOSkin^{TM} detectors (**Figure 2D**), developed at Centre for Medical Radiation Physics (CMRP) of the University of Wollongong, which feature an EPOM of 0.07 mm. These detectors were used with their own readout system measuring their gate threshold voltage. As their sensitivity to radiation dose decreases over large voltage ranges, the readout was corrected by taking a reference reading at the beginning and end of each set of measurements. The MOSkin^{TM} detectors have been recently shown to agree with EBT3 films in 1 T inline magnetic field and were deemed suitable for relative dose measurements (32). For both, microDiamond and MOSkin^{TM} measurements customized solid water blocks were used adapted to host the detectors and to enable measurements depth variation (**Figures 2D–F**).

For beam symmetry and flatness assessment as well as for some profile measurements, the Starcheck maxi MR array (PTW, Germany) was used. The array consists of 707 vented ionization chambers arranged, with 3 mm resolution, along the principal axes and the diagonals of a 40 × 40 cm^2 area and designed for use in magnetic fields of up to 1.5 T. It was characterized for reproducibility, linearity, sensitivity to misalignment, and geometrical fidelity, based on the IEC60731 (33), both in a 1 T field on the Australian MRI-linac and in a 0 T field on a 6 MV Elekta (Elekta, UK) clinical linear accelerator at the Liverpool Cancer Therapy Centre. Different orientation of individual detectors in detector arrays leads to non-negligible artifacts in profile measurements (34) for transverse MRI-linacs, however these effects have not been observed for the Australian MRI-linac employing an inline configuration. The profiles were analyzed with the Mephysto (PTW, Germany) software accompanying the detector.

Beam depth and cross profiles were acquired using Gafchromic^{TM} EBT3 films (Ashland, USA) placed in solid water blocks, as standard scanning water tanks are not compatible with the MRI-linac systems, due to the presence of metallic components and size restrictions. A dedicated stand (**Figure 2C**) was constructed to keep the solid water blocks tightly together

for the profile measurements in order to eliminate the presence of air gaps. The relative response of the EBT3 films has been shown to be unaffected by the magnetic field (35, 36). The batch of films used was calibrated using a 6 MV beam on an Elekta linear accelerator at the Liverpool Cancer Therapy Centre. The film handling and analysis followed published recommendations (37, 38). The films were scanned using a Perfection V700 Photo (Epson, Japan) flatbed scanner with resolution of 72 dpi and in 48-bit RGB format and all scanner color corrections turned off. A black paper frame was used to position the films at a consistent area of the scanner bed. The orientation of all films was kept constant and aligned to within ±5°. A thin glass plate was placed on top of the films during digitization in order to keep the films flat on the scanner bed. Films were processed and profiles were extracted using ImageJ software (National Institutes of Health, USA).

System Optimization
Magnetic Shielding Optimization
The fringe field affects the radiation beam generation and transport in the linac (8, 9, 39). In particular, (1) it may deflect the electrons produced by the electron gun and reducing the stream of electrons injected to the waveguide and hence the beam output and (2) it may shift the incidence of the electron beam on the target and lead to the deformation of the resulting photon beam profiles and output loss as the beam passes through the primary collimator. Initially, these effects have been reduced by magnetic shielding placed around the target area outside of the x-ray head housing. However, at shorter SIDs (i.e., in higher fringe field) magnetic shielding closer to the target was necessary. To ensure clinically acceptable beam characteristics at the shortest SID, a shield to be placed directly above the beam centerline has been prototyped first, using sheets of µ-metal (Magnetic Shield Corporation, USA), and later manufactured out of iron and fixed to ensure stability and reproducibility. The design of the shield was guided by measurements of the beam output and of the profile symmetry using the Starcheck$^{maxi-MR}$ array. For these measurements, the detector array was placed at the isocentre with build-up material equivalent to 10 cm of water and 10 cm of backscatter material behind it and profiles of varying field sizes were acquired at different SIDs.

System Alignment
Alignment of the radiation beam with the MRI scanner imaging isocentre was a two-step process. First, the MRI phantom was scanned using a T1-weighted spin-echo sequence in XZ (resolution: 0.9 × 0.9 × 5 mm) and XY (resolution: 0.8 × 0.8 × 5 mm) planes and realigned iteratively until its position matched the localization of the imaging isocentre. The in-room lasers were then set to indicate the imaging isocentre using the external markings on the phantom. Next, fiducial marker phantoms were added to the setup at the proximal and the distal end of the bore and the whole setup was imaged (at different SIDs) using the EPID placed behind the bore (**Figure 4A**). Based on the projections of the fiducial markers the position of the radiation source was calculated and the linatron was iteratively re-aligned to achieve the best congruence of the radiation

isocentre with the imaging isocentre over the range of SIDs. Finally, the half-blocked fields were imaged with EPID placed at two distances: behind the bore and in front of the bore and MLC center position at the isocentre was calculated based on these images to guide the iterative re-alignment of the MLC assembly and fine-tuning of the MLC central axis parameter in the MLC control software.

Field Size and Leaf Width Calibration

Due to a non-standard distance between the radiation source and the MLC, the magnification factor to apply on the field sizes set in the MLC control software had to be determined. It was obtained by the measurement of the actual field sizes produced by a set of square fields defined in the control software for standard clinical geometry for Millennium MLC, referred to in the reminder of this manuscript as nominal field sizes, using the EPID placed at the distance of 100 cm from the radiation source. The resulting calibration factors were then extrapolated to other SIDs and verified using films placed at the isocentre on the surface of the solid water phantom.

Functional Performance Characteristics

System characteristics have been benchmarked at commissioned SIDs against the applicable requirements either using or adapting the methods specified in the IEC 60976/977 standard (40, 41). Non-applicable tests included: electron radiation beams, dependencies on angular positions (collimator, gantry), moving beam RT, indicators (light field, front pointer, etc.) not present in the current system and patient support system constituting a separate development.

Dose Monitoring System

Reproducibility, proportionality, field size dependence and stability of the dose monitoring system were assessed using the chamber FC65-G placed in a solid water holder at the isocentre with 10 cm build-up and 10 cm backscatter material. 1 Gy irradiations with fields of $\sim$10 $\times$ 10 cm^2 were performed for the reproducibility measurements and with $\sim$20 $\times$ 5 and $\sim$5 $\times$ 20 cm^2 fields for the field size dependency measurements. Proportionality was assessed over a range of doses from 0.1 to 10 Gy. Stability after a high absorbed dose was assessed as the difference in measurements prior to and after a 30 min period of irradiation and stability throughout the week was assessed through measurements on 5 subsequent days following 3 h of stand-by mode. All stability tests have been performed at dose of 1 Gy for an $\sim$10 $\times$ 10 cm^2 field. Additionally, the magnitude of the magnetic field at the position of the monitoring chamber was recorded using a VGM gaussmeter (AlphaLab, USA).

Depth Dose Characteristics

Percentage depth dose (PDD) curves for $\sim$10 $\times$ 10 cm^2 and maximum field sizes were acquired using EBT3 films placed in solid water blocks and aligned parallel to the beam axis at SSD = SID$-$10 cm. PDDs were extracted along the beam central axis and at $\pm$3.5 cm off-axis in order to assess the depth of dose maximum (d$_{max}$) and the penetrative quality, defined as depth at which the dose amounts to 80% of the maximum dose (D$_{max}$) (40, 41). Additionally, films placed at the surface of the

phantom perpendicular to the beam direction were used for surface dose measurements.

Increased dose in the initial few centimeters around the beam central axis due to the contaminant electron focusing in the fringe field of the magnet has been previously both, modeled (11, 42) and observed experimentally on an earlier (16) and the current prototype (32). To characterize this effect in detail for the current system and to explore possible methods to mitigate it (off-axis irradiation, use of bolus), microDiamond and MOSkin^{TM} detectors were used. The detectors were placed in solid water blocks adapted to enable data acquisition at variable depths or covered with increasing layers of kapton tape.

Beam quality was characterized using tissue phantom ratio (TPR$_{20/10}$) measured using the FC65-G ion chamber in solid water using a field closest to 10 $\times$ 10 cm^2 and the dose of 1 Gy.

Beam Uniformity

Beam symmetry and flatness were measured using the Starcheck^{maxiMR} array for fields closest to 5 $\times$ 5, 10 $\times$ 10, 30 $\times$ 30 cm^2, and the maximum field size at commissioned SIDs. The detector array was placed at the isocentre with water equivalent material amounting to 10 cm of build-up and 10 cm of back scatter. Beam symmetry was determined as the maximum ratio of the doses at any two positions symmetrical to the beam axis and flatness as the ratio between the maximum and the minimum dose inside the flattened area (40, 41).

Beam penumbra (20–80%) was measured at the isocentre at 10 cm depth. For fields of $\sim$5 $\times$ 5 and 10 $\times$ 10 cm^2 it was extracted from EBT3 films placed perpendicular to the beam axis within solid water blocks. For the 30 $\times$ 30 cm^2 and the maximum commissioned field sizes, which exceed the size of the solid water blocks, penumbra measured with the Starcheck^{maxiMR} array and dedicated build-up plates is reported. In order to apply the standard 20–80% definition for penumbra evaluation, only the field edge sections of the FFF profile have to be considered (43). This was achieved by identifying the profile inflection points using a method which calculates the third derivative of the profile, proposed by Fogliata et al. (44), and renormalizing profiles to these points.

Isocentre

Congruence of the imaging and radiation isocentre, expressed as horizontal (x) and vertical (y) offset between the beam focal position and the in-room lasers, was measured using the setup and methods described in section System Alignment.

Geometry of the Beam Limiting System

Symmetry of the opening around the imaging isocentre and parallelism of the leaves to the in-room lasers have been tested here, since the current system is not equipped with diaphragms. These tests have been performed using the setup and methods described in section System Alignment. The latter was measured as the angle between the line defined by the projection of the fiducials and projection of the MLC CAX in EPID images (see **Figure 6**) using ImageJ software.

Base Data Acquisition

Beam characterization has been performed at commissioned SIDs and SSD = SID$-$10 cm following the AAPM Task Group

106 recommendations (45). Non-applicable procedures included elements not available in the current system: measurements of tray and wedge factors, tests of the light field and radiation field concurrence and characterization of the electron beams. In the initial clinical phase, the MRI-linac will be used only for static 3D conformal treatments, hence only the acquisition of the input data relevant for such treatments is reported in this work.

Depth Dose Characteristics and Surface Dose

PDDs were measured for square fields of up to ~18 × 18 cm^2 and two rectangular fields of ~18 × 6 and ~6 × 18 cm^2 using EBT3 films placed in solid water blocks and aligned parallel to the beam axis. Additionally, films placed at the surface of the phantom perpendicular to the beam direction were used for surface dose measurements. PDDs were extracted along the central axis.

Beam Profiles

Beam profiles were acquired for square fields of up to ~18 × 18 cm^2 and two rectangular fields of ~18 × 6 and ~6 × 18 cm^2 at depths of 1, 5, 10, and 20 cm. EBT3 films were placed perpendicular to the beam direction in solid water blocks.

Tissue Phantom Ratios

Beam quality measurements as required by IEC 60976/977 standard (40, 41) were addressed as part of functional performance characterization in section Depth Dose Characteristics.

Output Factors

Total scatter factors were measured in solid water using the FC65-G ion chamber and the microDiamond detector for field sizes from the smallest available to ~25 × 25 cm^2. Collimator scatter factors were measured at 10 cm depth using the FC65-G ion chamber in a GEC-ESTRO mini phantom placed horizontally (46). Results were normalized to the field closest to 10 × 10 cm^2.

Beam Output Calibration

Absolute dosimetry was performed following the TRS-398 protocol (47). Output was measured in a manual 2D water tank in 10 cm depth under isocentric conditions for square fields closest to 10 × 10 cm^2 using FC65-G ion chamber calibrated in the magnetic field and traceable to NPL as described above. Corrections for polarity, recombination, ambient room conditions and magnetic field were applied. Polarity was measured via acquisition of output with opposite polarizing potentials (-300 and $+300$ V) applied to the chamber and yielded $k_{pol} = 1.0005$. Recombination was measured via the two voltage method using polarizing potentials of -300 and -100 V and yielded $k_s = 1.0015$. Calibration in 1 T yielded a k_B factor of 0.99 (30). Correction of $k_{FFF} = 1.003$ was used for FFF beam volume averaging effects. A CC13 chamber was used to normalize the output between measurements.

MLC Characterization

Positional accuracy of the MLC (symmetry around the isocentre and tilt) has been assessed as described in section Geometry of the Beam Limiting System and beam penumbra was measured with EBT3 films as described in section Beam Uniformity.

The MLC transmission measurements were performed using a method similar to one described by Arnfield et al. (48) and Patel et al. (49) with leaves fully closed and the gap between the opposing leaf pairs displaced to one side. Average MLC transmission was measured using the FC65-G ion chamber placed in solid water perpendicular to the leaf travel direction at the isocentre at depth of 11 cm. Simultaneously, the intra- and interleaf leakage was measured using an EBT3 film placed at 10 cm depth.

RESULTS

System Optimization
Magnetic Shielding Optimization

Figure 3 shows the profiles of a nominal 11 × 11 cm^2 field acquired at different SIDs with only external magnetic shielding (dotted lines) and with optimized internal magnetic shielding (solid lines). The shielding significantly improved the profile symmetry even at shortest SID and reduced the beam output loss in the target area. It should be noted however that, located relatively far from the electron gun, it was not effective in reducing the beam loss occurring there. Based on these results the system has been commissioned at two SIDs: at 2.4 m, where full range of fields (up to 34.3 × 34.0 cm^2 at the isocentre) fulfilled the symmetry criteria, and at 1.8 m, where fields of up to 21.4 × 21.2 cm^2 fulfilled the symmetry criteria. The measurements reported in the reminder of this manuscript have been performed at these two SIDs, unless stated otherwise.

System Alignment

The MRI scans of the alignment phantom with the grid indicating the imaging isocentre superimposed are shown in **Figure 4B** and the measured agreement was better than 1 mm.

The physical change in position of the linatron relative to the origin axis over the rail length was 2 mm horizontally (x) and 3 mm vertically (y).

Example composite portal images, showing the projections of the fiducial markers (as indicators of the laser positions) and of the edges of half blocked fields formed by the MLC, are shown in **Figure 4C**. Composite images were created as: |image$_{negativexblocked}$ − image$_{positivexblocked}$| + |image$_{negativeyblocked}$ − image$_{positiveyblocked}$| allowing visualization of the MLC axes. Alignment of radiation isocentre and the MLC center relative to the positioning lasers for different SIDs is summarized in **Figure 5.** The results indicate a variation of the radiation focal spot offset (circles in **Figure 5**) from to the laser with changing SID of up to 1.5 ± 2.5 mm in the horizontal (x) direction and 1.4 ± 1.8 mm in the vertical (y) direction, except at SID of 1.8 m where it reached 2.1 ± 1.6 mm. This leads to the variation of the position of the MLC CAX projection with changing SID, which in the horizontal (x) direction could be compensated for using an SID specific parameter setting in the MLC control software.

Offsets of the MLC CAX projection position with respect to the lasers measured in EPID images acquired in front of the magnet bore ($z = -167.6$ cm) and behind the magnet bore ($z = 130.1$ cm) (bars in **Figure 5**) were used to interpolate the MLC CAX projection at the isocentre ($z = 0$ cm) (×'s in

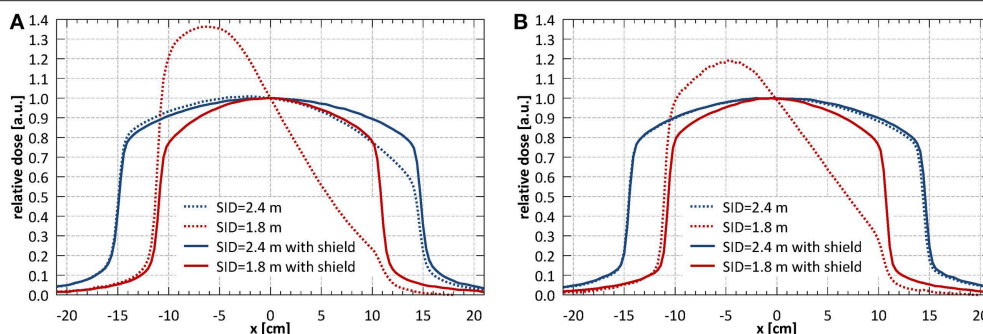**FIGURE 3 | (A)** Horizontal (x) and **(B)** vertical (y) profiles of a nominal 11 × 11 cm² field acquired at different SIDs with only external magnetic shielding (dotted lines) and with optimized internal magnetic shielding (solid lines).
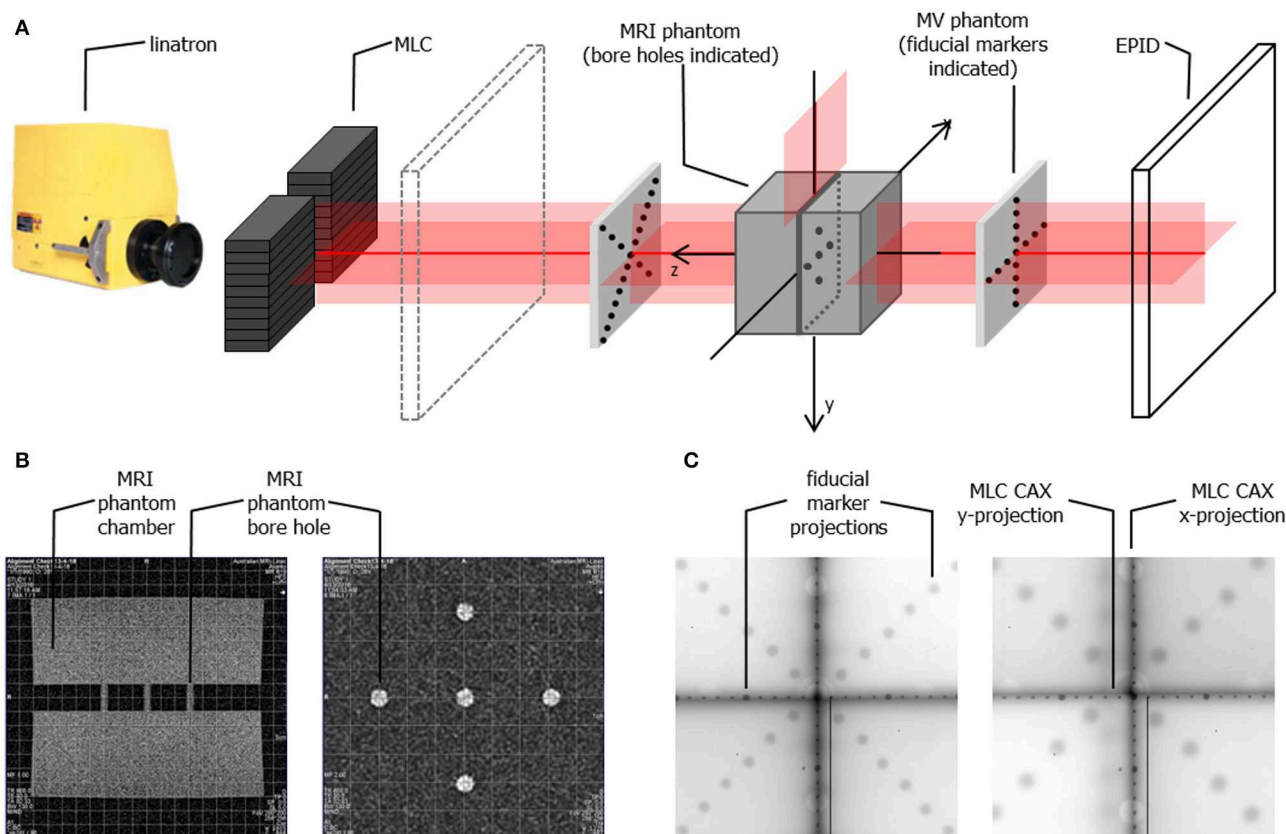

FIGURE 4 | (A) Schematic representation of the phantom setup used for geometrical alignment of the system **(B)** MR scans of the alignment phantom in x-z plane (left) and x-y plane (right) showing the bore holes aligned to the imaging isocentre (indicated by the superimposed grid) **(C)** example composite portal images showing the projections of the fiducial markers (aligned to the in-room lasers) and of the edges of half blocked fields formed by the MLC for SID of 2.4 m (left) and SID of 1.8 m (right). Composite images are created as: $|image_{negativexblocked} - image_{positivexblocked}| + |image_{negativeyblocked} - image_{positiveyblocked}|$ allowing visualization of the MLC axes.

Figure 5). In the horizontal (x) direction it was within 0.4 ± 1.9 mm at all SIDs. In the vertical (y) direction it was within 1.8 ± 2.2 mm at all SIDs except the largest two (3.2 and 3.0 m).

Field Size and Leaf Width Calibration

The measured field sizes were 7.2% larger in horizontal (x) and 6.1% larger in vertical (y) direction than the nominal field sizes. Extrapolated to the two commissioned

SIDs this yielded magnification factors of 2.638 and 2.612 in horizontal and vertical direction for SID of 2.4 m and 1.944 and 1.924 in horizontal and vertical direction for SID of 1.8 m. This was incorporated in the TPS in the definition of the MLC leaf projection widths in the isocentre plane.

The full width half maximum of the surface profiles acquired using EBT3 film for fields closest to 10 × 10 cm²: 10.6 × 10.5 cm²
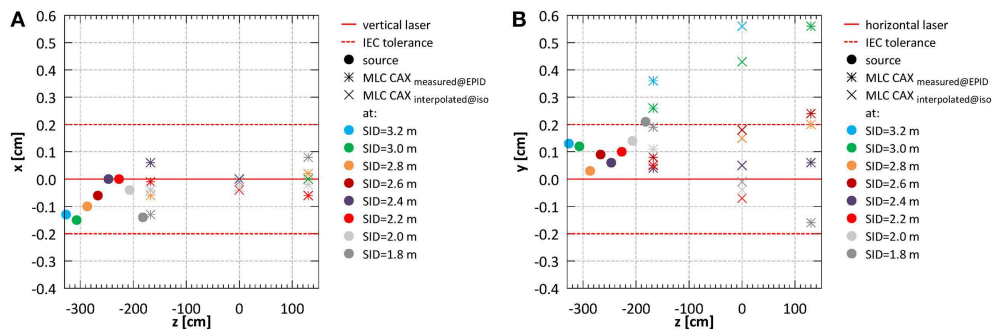
FIGURE 5 | System alignment for all SIDs **(A)** in the horizontal (x) and **(B)** in the vertical (y) direction.
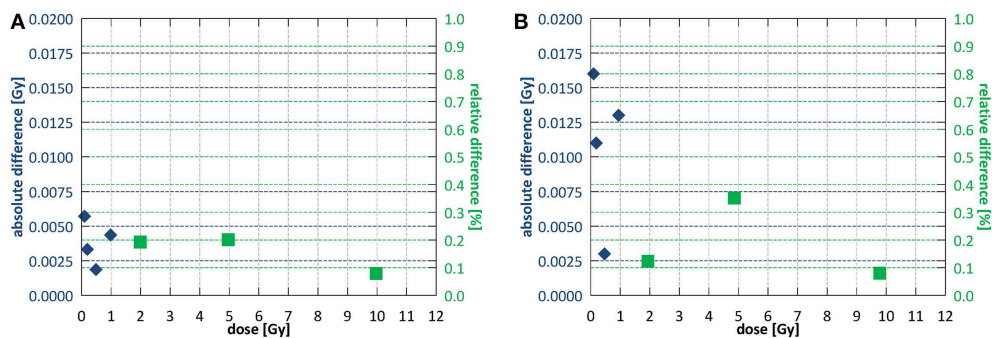

FIGURE 6 | Dose output linearity **(A)** at SID of 2.4 m and **(B)** at SID of 1.8 m. Different symbols indicate regions of applicability of the absolute and the relative deviation criterion according to IEC60976 (39).

at SID of 2.4 m and 9.7 × 9.6 cm² at SID of 1.8 m were 10.6 × 10.1 and 9.7 × 9.6 cm² respectively.

Functional Performance Characteristics
Dose Monitoring System
For reproducibility, proportionality and stability measurements fields of 10.6 × 10.5 and 9.7 × 9.6 cm² were used for SID of 2.4 and 1.8 m, respectively. Short term reproducibility of the monitoring chamber calculated as a coefficient of variation (40) was 0.29% at SID of 2.4 m and 0.49% at SID of 1.8 m. Output after high absorbed dose showed a decrease in of 1.2 ± 0.4% at SID of 2.4 m and 1.1 ± 0.4% at SID of 1.8 m. Stability throughout the week was 1.0 ± 0.6% at SID of 2.4 m and 2.6 ± 0.6% at SID of 1.8 m. Stability after a full day of intensive commissioning measurements yielded 1.7 ± 0.4 and 2.8±1.0% output decrease at SID of 2.4 and 1.8 m, respectively, however such intensive clinical use of the system is not foreseen outside of commissioning or annual quality assurance. **Figure 6** shows the dose output linearity, calculated as per IEC60977 (41), for both commissioned SIDs. At SID of 2.4 m, linearity was better than 0.2% above 1 Gy and better than 0.006 Gy below 1 Gy. At SID of 1.8 m, linearity was better than 0.4% above 1 Gy and better than 0.016 Gy below 1 Gy.

For the measurements of the dependence on the field shape, fields of 5.3 × 20.9 and 21.1 × 5.2 and 5.8 × 19.2 and 19.4 × 5.8 cm² were used for SID of 2.4 and 1.8 m, respectively. Variation

with the field size shape was 1.3 ± 0.4% at SID of 2.4 m and 0.0 ± 0.5% at SID of 1.8 m.

The fringe field magnitude at the location of the monitoring chamber was ∼15 and 45 mT for SID of 2.4 and 1.8 m, respectively.

Depth Dose Characteristics
The dose distributions acquired using EBT3 films in the region of ∼±10 cm around the beam CAX, normalized at depth of 10 cm, at SID of 2.4 m for field sizes of 10.6 × 10.5 and 34.3 × 34.0 cm² are shown in **Figures 7A,B** and at SID of 1.8 m for field sizes of 9.7 × 9.6 and 21.4 × 21.2 cm² in **Figures 7C,D**. The higher dose around the central axis at small depths caused by electron focusing is visible.

The surface doses (normalized to 10 cm depth) measured using EBT3 films were: 430% for 10.6 × 10.5 cm² field and 1,024% for 34.3 × 34.0 cm² at SID of 2.4 m and 576% for 9.7 × 9.6 cm² field and 1,068% for 21.4 × 21.2 cm² at SID of 1.8 m. Measured ±3.5 cm off-axis, at SID of 2.4 m these values were reduced to 79% for 10.6 × 10.5 cm² field and 143% for 34.3 × 34.0 cm² and at SID of 1.8 m to 72% for 9.7 × 9.6 cm² field and 103% for 21.4 × 21.2 cm². The radius at which the surface dose becomes lower than the dose at 10 cm depth was between 2.6 cm for 10.6 × 10.5 cm² field and more than 6 cm for 34.3 × 34.0 cm² field at SID of 2.4 m and between 2.6 cm for 9.7 × 9.6 cm² field and 4 cm for 21.4 × 21.2 cm² at SID of 1.8 m.
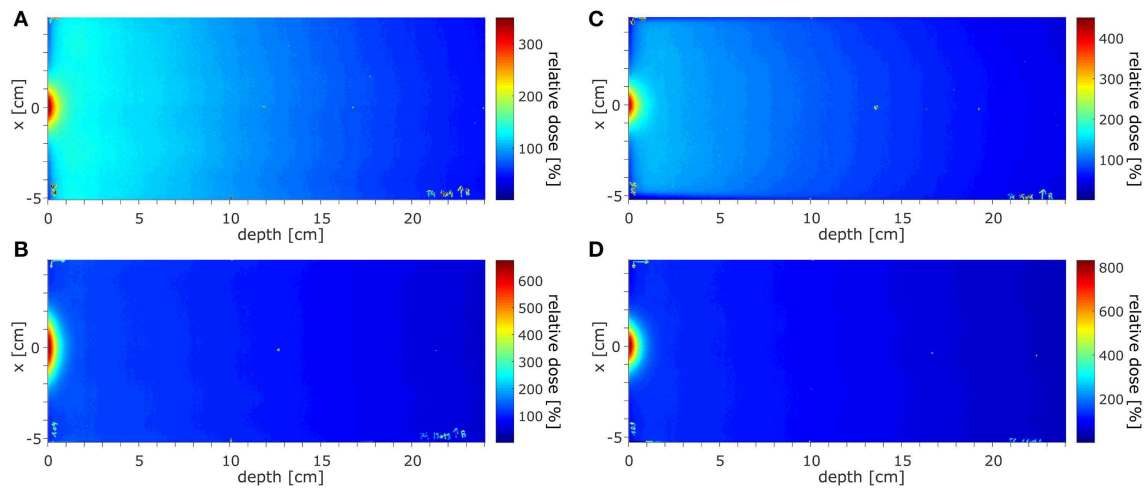
FIGURE 7 | The dose distributions acquired in the region of ±10 cm around the beam CAX, normalized at depth of 10 cm, **(A,B)** at SID of 2.4 m for field sizes of 10.6 × 10.5 and 34.3 × 34.0 cm^2 and **(C,D)** at SID of 1.8 m for field sizes 9.7 × 9.6 and 21.4 × 21.2 cm^2.
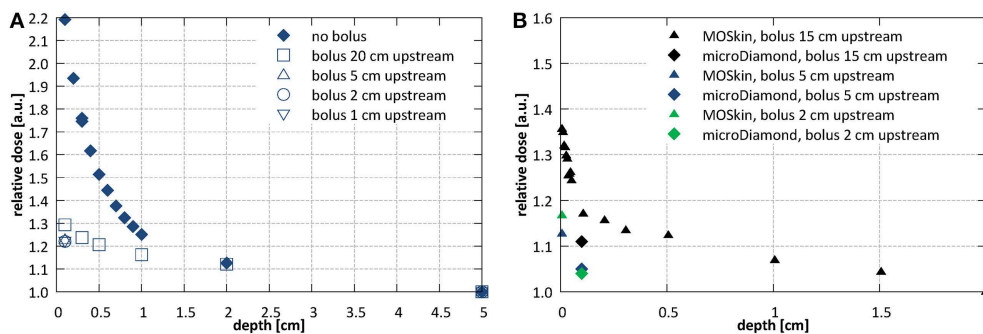

FIGURE 8 | Investigations of the entrance dose at SID of 2.4 m with varying bolus placement **(A)** with microDiamond for a 10.6 × 10.5 cm^2 field at SID of 2.4 m (normalized at depth of 5 cm) and **(B)** with MOSkin^{TM} and microDiamond detector for a 7.9 × 7.8 cm^2 field (normalized at depth of 2 cm).

High dose deposited in the initial section of the PDD by the contaminant electrons focused around the beam central axis hinders the determination of the depth of d$_{max}$ and of the penetrative quality of the beam in the radiation field according to the IEC 60976/977 standard (40, 41). As estimates, these values were extracted ±3.5 cm off-axis yielding: 1.47 cm for field 10.6 × 10.5 cm^2 and 1.54 cm for field 34.3 × 34.0 cm^2 at SID of 2.4 m and 1.45 cm for field 9.7 × 9.6 cm^2 and 1.49 cm for field 21.4 × 21.2 cm^2 at SID of 1.8 m. The penetrative quality was 7.19 cm for the 10.6 × 10.5 cm^2 at SID of 2.4 m and 7.02 cm for the 9.7 × 9.6 cm^2 at SID of 1.8 m.

Use of a bolus placed upstream of the entrance surface to mitigate the presence of the contaminant electrons was investigated. Bolus thickness of 2 cm was selected based on the observed penetration depth of the electrons. **Figure 8A** shows the initial section of the PDD acquired using the microDiamond detector for a 10.6 × 10.5 cm^2 field at SID of 2.4 m with bolus placed 20, 10, 5, 2, and 1 cm upstream from the surface of the phantom. Measurements were normalized to the dose at 5 cm depth due to the contaminant electrons affecting the dose

maximum position. The presence of the bolus lead to a significant reduction of the electron hotspot: from more than 220% at 1 mm depth to about 120–130%, depending on the distance at which the bolus was placed. Placing the bolus close to the surface, i.e., reducing the length of the air column where new electrons can be generated, resulted with lower surface dose, however only down to a distance of ∼5 cm upstream from the phantom surface. **Figure 8B** shows higher resolution data (normalized at depth of 2 cm) acquired with the MOSkin^{TM} detector, which reveal presence of a further dose enhancement and steep dose fall-off within the initial 1 mm of the PDD, i.e., at depths smaller than the EPOM of the microDiamond.

For TPR$_{20/10}$ measurements fields of 10.6 × 10.5 and 9.7 × 9.6 cm^2 were used for SID of 2.4 and 1.8 m, respectively. The measured TPR$_{20/10}$ values were 0.633 ± 0.001 at SID of 2.4 m and 0.634 ± 0.004 at SID of 1.8 m.

Beam Uniformity

Results of the beam symmetry and flatness measurements performed with the Starcheck^{maxiMR} array are summarized in

TABLE 1 | Beam symmetry, flatness and penumbra at depth of 10 cm at SID of 2.4 m.

Field size (cm^2)	Symmetry (%)		Flatness		Penumbra (cm)	
	x	y	x	y	x	y
5.3 × 5.2	100.4	100.3	1.04	1.03	1.06[+] 0.90[*]	0.66[+] 0.73[*]
10.6 × 10.5	100.4	100.5	1.08	1.08	1.35[+] 1.10[*]	0.86[+] 0.84[*]
29.0 × 28.7	101.8	100.6	1.17	1.15	n.a.[+] 1.48[*]	n.a.[+] 1.29[*]
34.3 × 34.0	102.2	100.4	1.22	1.22	n.a.[+] 1.49[*]	n.a.[+] 1.23[*]

[*] Measured with Starcheck $^{maxi\ MR}$.
[+] Measured with film.

TABLE 2 | Beam symmetry, flatness and penumbra at depth of 10 cm at SID of 1.8 m.

Field size (cm^2)	Symmetry (%)		Flatness		Penumbra (cm)	
	x	y	x	y	x	y
5.8 × 5.8	101.4	100.8	1.05	1.04	0.89[+] 0.86[*]	0.56[+] 0.67[*]
9.7 × 9.6	102.0	101.6	1.10	1.09	1.02[+] 0.94[*]	0.74[+] 0.76[*]
21.4 × 21.2	101.5	102.9	1.18	1.18	n.a.[+] 1.17[*]	n.a.[+] 1.01[*]

[*] Measured with Starcheck $^{maxi\ MR}$.
[+] Measured with film.

Table 1 for SID = 2.4 m and in **Table 2** for SID of 1.8 m. At the larger SID, the IEC criteria for symmetry were fulfilled for fields up to 34.3 × 34.0 cm^2 while at shorter SID for fields up to 21.4 × 21.2 cm^2.

Beam penumbra measured with films (for field sizes not exceeding the size of the solid water blocks) and with the Starcheck array (for all investigated field sizes) at 10 cm depth is summarized in **Table 1** for SID of 2.4 m and in **Table 2** for SID of 1.8 m. The results obtained with films and with the array agree on average within 1 mm. In the direction perpendicular to the leaf motion (y), where the penumbra is steeper, compared to the direction perpendicular to the leaf motion (y), the penumbra values were 3 mm lower when measured with films and 4 mm lower when measured with the array.

Isocentre
Offset of the radiation focal spot from the lasers defining the position of the imaging isocentre was 0.0 ± 2.1 mm in the horizontal (x) and 0.6 ± 2.1 mm in the vertical (y) direction at SID of 2.4 m and −1.4 ± 1.6 mm in the horizontal and 2.1 ± 1.6 mm in the vertical direction at SID of 1.8 m.

Geometry of the Beam Limiting System
Offset of the projection of the MLC center from the lasers defining the position of the imaging isocentre was 0.0 ± 2.1 mm in the horizontal (x) and 0.8 ± 2.1 mm in the vertical (y) direction

at SID of 2.4 m and −0.1 ± 1.6 mm in the horizontal and −0.2 ± 1.6 mm in the vertical direction at SID of 1.8 m. The MLC bank tilt was 0.28° measured at SID of 2.4 m and 0.13° measured at SID of 1.8 m.

Beam Base Data Acquisition
Based on the surface dose measurements performed as part of the functional characteristic tests, the base data relevant for beam modeling (PDDs, beam profiles and absolute dose calibration) was acquired for both cases: without and with the electron absorbing bolus placed 5 cm upstream from the surface of the phantom, as currently its use is foreseen for first patient treatments.

Depth Dose Characteristics and Surface Dose
Example PDD curves measured at SID of 2.4 m for field sizes 2.6 × 2.6, 10.6 × 10.5, 18.5 × 18.3, and 34.3 × 34.0 cm^2 normalized to the 10.6 × 10.5 cm^2 field at a depth of 10 cm are shown in **Figure 9A** without and **Figure 9B** with the bolus. Example PDD curves measured at SID of 1.8 m for field sizes 1.9 × 1.9, 9.7 × 9.6, 17.5 × 17.3, and 21.4 × 21.4 cm^2 normalized to the 9.7 × 9.6 cm^2 field at depth of 10 cm are shown in **Figure 9C** without and **Figure 9D** with the bolus.

Beam Profiles
Example profiles acquired at depths of 1, 5, 10, and 20 cm at SID of 2.4 m for field sizes 2.6 × 2.6, 10.6 × 10.5, and 18.5 × 18.3 cm^2 normalized to the CAX value of the profile of the 10.6 × 10.5 cm^2 field at 10 cm depth are shown in **Figures 10A–C**. Example profiles acquired at SID of 1.8 m for field sizes 1.9 × 1.9, 9.7 × 9.6, and 17.5 × 17.3 cm^2 normalized to the CAX value of the profile of the 9.7 × 9.6 cm^2 field at 10 cm depth are shown in **Figures 10D–F**. Horizontal (x) half-profiles are shown on the negative and corresponding vertical (y) half-profiles on the positive x axis. Only data measured without the bolus is shown.

Tissue Phantom Ratios
Results of the beam quality measurements required by IEC 60976/977 standard (40, 41) are presented as part of functional performance characterization in section Depth Dose Characteristics.

Output Factors
Field size output factors were measured for fields between 2.6 × 2.6 and 26.4 × 26.1 cm^2 at SID of 2.4 m and normalized to the field of 10.6 × 10.5 cm^2. At SID of 1.8 m, the field sizes between 1.9 × 1.9 and 25.3 × 25.0 cm^2 were used and the results were normalized to the field of 9.7 × 9.6 cm^2. The total (S_{CP}) and collimator (S_C) scatter factors measured under isocentric conditions are shown in **Figure 11**.

At SID of 2.4 m the total scatter factors measured with the ion chamber and with the microDiamond detector showed a very good agreement for field sizes down to ~5 × 5 cm^2, with an average deviation of 0.3%. Below, at field size of 2.6 × 2.6 cm^2, the ion chamber underestimated the output by 5.1%. At SID of 1.8 m the agreement between ion chamber and microDiamond was good down to field size of ~4 × 4 cm^2, with an average
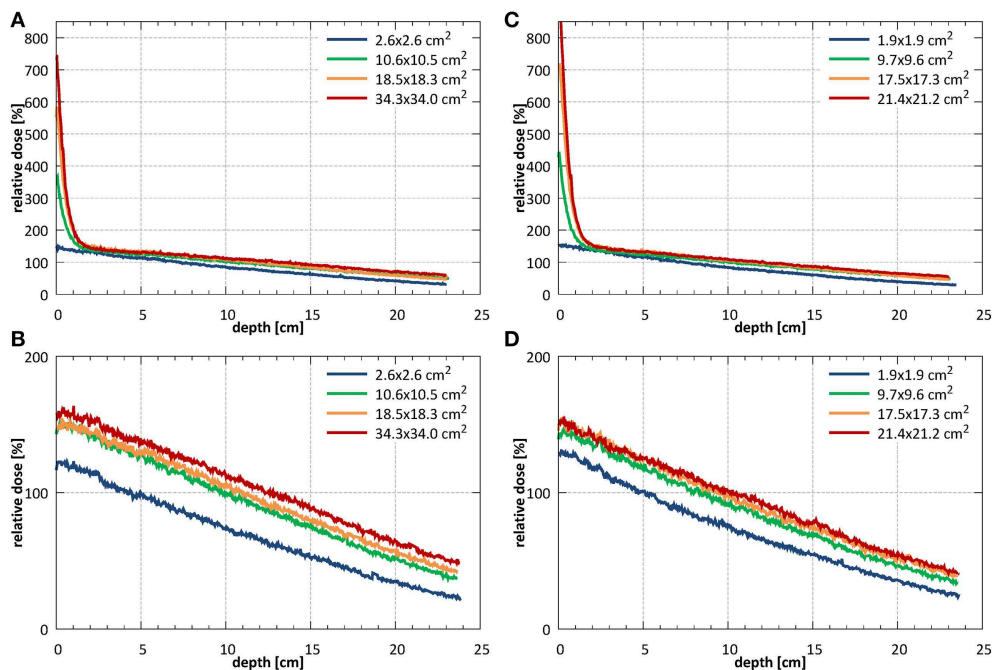
FIGURE 9 | PDDs measured **(A)** without **(B)** with the bolus at SID of 2.4 m for field sizes 2.6 × 2.6, 10.6 × 10.5, 18.5 × 18.3, and 34.3 × 34.0 cm^2 (normalized to the 10.6 × 10.5 cm^2 field at a depth of 10 cm) and **(C)** without and **(D)** with the bolus at SID of 1.8 m for field sizes 1.9 × 1.9, 9.7 × 9.6, 17.5 × 17.3, and 21.4 × 21.2 cm^2 (normalized to the 9.7 × 9.6 cm^2 field at a depth of 10 cm).

deviation of 0.2%. Below, at field size of 1.9 × 1.9 cm^2, the ion chamber underestimated the output by 17.9%.

Beam Output Calibration

Beam output calibration was performed at a field size of 10.6 × 10.5 and 9.7 × 9.6 cm^2 for SID of 2.4 and 1.8 m, respectively. The measured beam output was 0.1376 ± 0.0002 Gy/MU at SID of 2.4 m and 0.0915 ± 0.0002 Gy/MU at SID of 1.8 m. The respective factors measured with the electron absorbing bolus in place were 0.1247 ± 0.0001 Gy/MU at SID of 2.4 m and 0.0830 ± 0.0001 Gy/MU i.e., 9.42 and 9.37% lower.

MLC Characterization

Positional accuracy of the MLC (symmetry around the isocentre and tilt) is presented in section Geometry of the Beam Limiting System and the measured beam penumbra values in section Beam Uniformity.

Average transmission through the leaves, relative to an open field dose, measured with the EBT3 film (averaged within a radius of 2.5 cm around the center of the field) at SID of 2.4 m was 1.06% and at SID of 1.8 m 1.17%. The corresponding ion chamber measurements were slightly higher and yielded 1.23 and 1.52%, respectively. Interleaf transmission peak-to-peak amplitude was ∼0.4% at SID of 2.4 m and 0.3% at SID of 1.8 m (**Figure 12**).

DISCUSSION

This work is the first report on the dosimetric characterization of an inline MRI-linac system.

The Australian MRI-linac features a rail system which enables variation of the source-to-isocentre distance. The largest distance (3.2 m) corresponds to the decoupling of the MRI and linac components (B ≈ 0 T). On the other hand, the fringe field at the linac location for the shortest distance (1.8 m) influences the electron transport, and therefore radiation beam generation in the linac, leading to output loss and profile distortion. In order to utilize the shortest SIDs, physical alignment of the radiation beam to the magnetic field was fine-tuned and magnetic shielding of the radiation head was optimized. Aligning and shielding allowed the system commissioning at two SIDs: at 2.4 m where full range of field sizes fulfilled the symmetry criteria and at 1.8 m where fields up to ∼21 × 21 cm^2 fulfilled the symmetry criteria, but the clinical treatments will benefit from smaller leaf width projection and sharper penumbra.

Alignment of the radiation field and the imaging isocentre below the 2 mm tolerance specified in IEC standard was obtained: the deviation of the center of MLC shaped fields from the isocentre was 0.0 ± 2.1 mm in the horizontal (x) and 0.8 ± 2.1 mm in the vertical (y) direction at SID of 2.4 m and −0.1 ± 1.6 mm in the horizontal (x) and −0.2 ± 1.6 mm in the vertical (y) direction at SID of 1.8 m. These offsets stem from the limitations of physical linac and MLC assembly alignment and vary with the SID due to factors such as: residual angle between the rail system, the linac and the magnetic field axis as well as the influence of the fringe field on the electron beam in the linac and hence on its point of incidence on the target. While in the horizontal (x) direction this could be
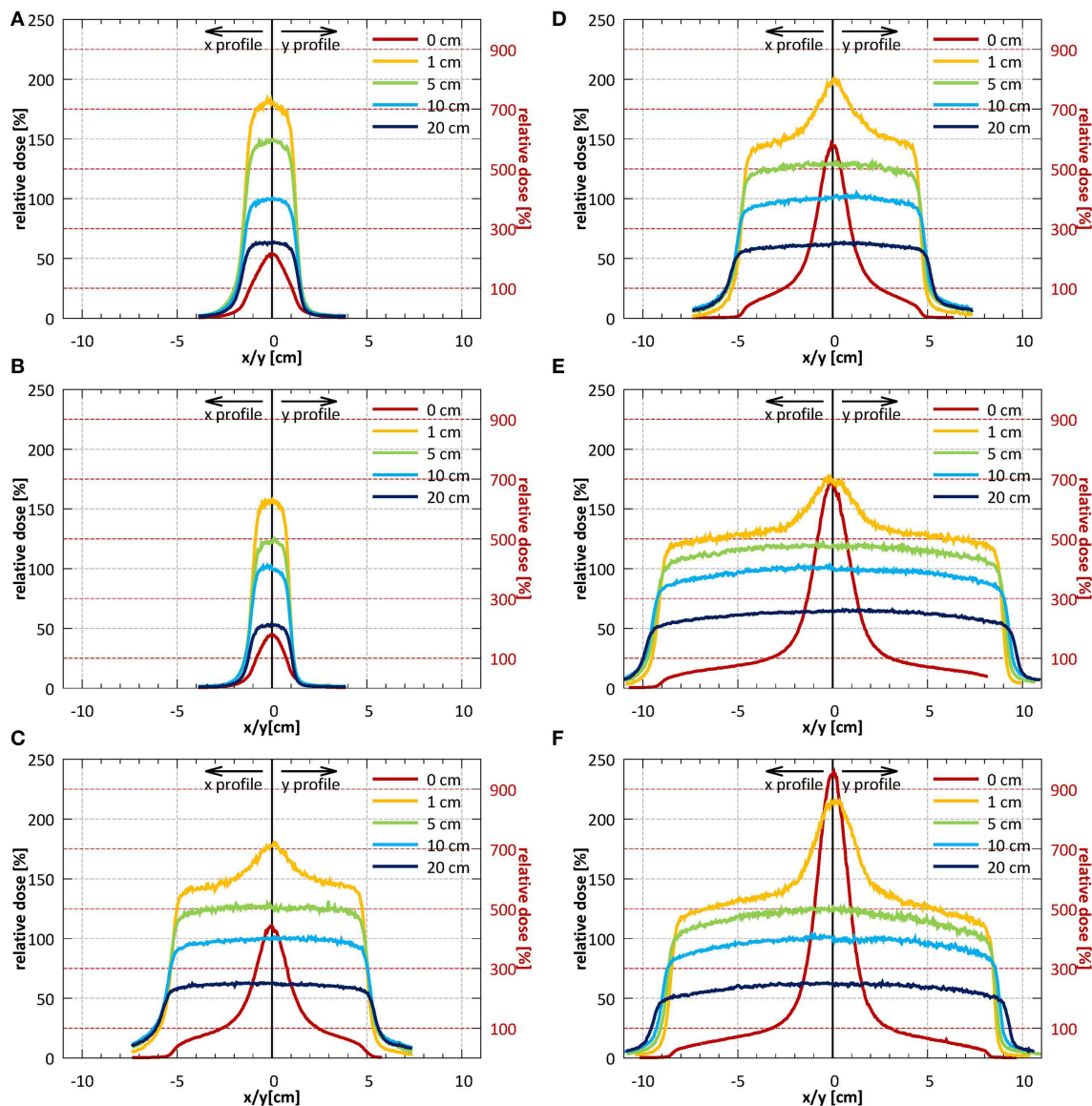
FIGURE 10 | Profiles in the horizontal (x) direction (solid lines) and in the vertical (y) direction (dotted lines) measured without the bolus at the surface and at depths of 1, 5, 10, and 20 cm **(A–C)** at SID of 2.4 m for field sizes 2.6 × 2.6, 10.6 × 10.5, and 18.5 × 18.3 cm² (normalized to CAX value of the 10.6 × 10.5 cm² field at a depth of 10 cm) and **(D–F)** at SID of 1.8 m for field sizes 1.9 × 1.9, 9.7 × 9.6, and 17.5 × 17.3 cm² (normalized to CAX value of the 9.7 × 9.6 cm² field at a depth of 10 cm). Note: the secondary y-axis was used for surface profiles and primary y-axis was used for all remaining profiles.

minimized by software-controlled adjustment of the MLC axis, in the vertical direction (y) a compromise in the MLC height placement was made. MLC leakage yielded 1.06% at SID of 2.4 m and 1.17% at SID of 1.8 m measured with film and 1.23% at SID of 2.4 m and 1.52% at SID of 1.8 m when measured with an ionization chamber. In similar measurement, Arnfield et al. (48) reported a value of 1.34 ± 0.03% for the same MLC model.

Profile symmetry was better than 103% for the commissioned field size ranges and the flatness values of 1.03–1.22 were comparable with values reported for a 6 MV FFF beam in

literature (50). Penumbra values could be measured with films, offering high spatial resolution, only for a subset of fields required by IEC standard. For an ~10 × 10 cm² field the penumbra was 1.35 cm in the leaf motion direction (x) and 0.86 cm in the direction perpendicular to the leaf motion (y) at SID of 2.4 m and 1.02 cm in the leaf motion direction and 0.74 cm in the direction perpendicular to the leaf motion at SID of 1.8 m. For comparison, penumbra values measured for the same field size on a commercial transverse 1.5 T system, equipped with an Elekta Agility MLC, with an SID of 1.4 m determined using artificial flattening were between 0.74 and 0.87 cm (28). No penumbra
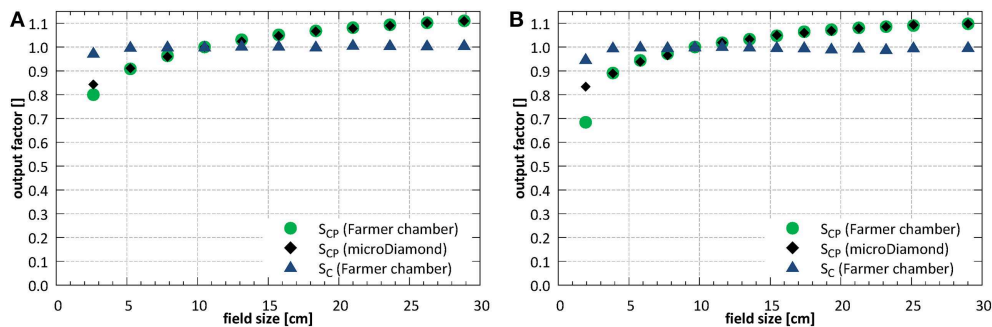
FIGURE 11 | Total (S_{CP}) and collimator (S_C) scatter factors **(A)** at SID of 2.4 m and **(B)** at SID of 1.8 m.
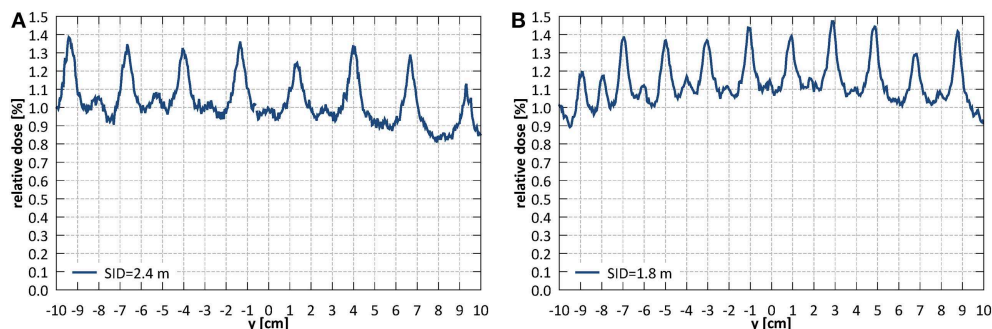

FIGURE 12 | Leakage dose (relative to an open field dose) through a closed MLC **(A)** at SID of 2.4 m and **(B)** at SID of 1.8 m.

asymmetry and profile shift, characteristic for transverse systems (28), has been observed, which simplifies beam modeling in the TPS for inline systems.

The dose monitoring system of the Linatron-MP consists of a single, parallel plate, unsealed monitoring chamber. Reproducibility and proportionality of the chamber met the IEC criteria. However, at shorter SID, the long-term stability over 1 week reached $2.6 \pm 0.6\%$, exceeding the IEC defined tolerance. An independent monitoring chamber is currently being installed to mitigate this for patient treatments as well as to ensure dose monitoring redundancy as per IEC requirements (40, 41). Fringe field effects on beam the beam properties (e.g., presence of the focused electrons, backscatter) and on the dose monitoring system response required a separate beam output calibration at the two commissioned SIDs. Beam quality instead remained the same at both distances ($TPR_{20,10}$ was 0.633 ± 0.001 at SID of 2.4 m and 0.634 ± 0.004 at SID of 1.8 m) within measurement uncertainty. It should be noted that in inline configurations, similar to transverse systems, $TPR_{20,10}$ as opposed to $\%dd(10)_x$ is more applicable as beam quality measure. Although, contrarily to transverse systems, photon build-up remains unaffected by the inline magnetic fields, determination of d_{max} and D_{max} may be confounded by the presence of electron focusing. In this work, as an approximate estimate the value of d_{max} measured off-axis was reported and amounted to 1.47 cm at SID of 2.4 m and 1.45 cm at SID of 1.8 m for $\sim 10 \times 10 \, cm^2$ fields. For comparison, a

reduced d_{max} of 1.3 cm was reported for a 7 MV FFF beam of a commercial 1.5 T transverse system (28).

Electron focusing effect, modeled (42) and observed experimentally (16, 32), was quantified and methods to minimize it have been investigated. Surface dose enhancement around the central axis was observed, which was dependent on the field size and reached 400–600% for $10 \times 10 \, cm^2$ fields and more than 1,000% for largest fields, relative to the dose at 10 cm depth. This could be counteracted by placing of an absorbing bolus upstream of the phantom or by irradiation using off-axis fields. The former resulted in significant reduction of the entrance dose although keeping the maximum dose value on the surface: $\sim$140–150% for $\sim 10 \times 10 \, cm^2$ fields and 160% for small fields, relative to the dose at 10 cm depth. The efficacy of the latter is dependent on the off-axis distance: the distance at which the surface dose becomes lower than the dose in 10 cm depth was between 2.6 cm for $10.6 \times 10.5 \, cm^2$ field and more than 6 cm for $34.3 \times 34.0 \, cm^2$ field at SID of 2.4 m and between 2.6 cm for $9.7 \times 9.6 \, cm^2$ field and 4 cm for $21.4 \times 21.2 \, cm^2$ at SID of 1.8 m.

For absolute dose measurements, the correction factor k_B was applied to the ionization chamber reading. However, it should be emphasized, that the effect of the magnetic field on the dosimeters and the sensitivity to the detector orientation has been shown to be less pronounced in inline as compared to transverse configuration (19, 20). To avoid the effect of air gaps, the chamber holder has been filled out with water whenever applicable in this

work. Nevertheless, it should also be noted that these effects have been shown to be smaller for the inline relative to the transverse configuration: 0.4% (30) compared to 0.7–1.2% (26).

The beam base data acquired with the bolus has been inserted into Pinnacle (Philips Healthcare, The Netherlands) TPS system for beam modeling, as currently its use is foreseen for first patient treatments. This data as well as data acquired without the bolus will be instead used to improve and validate the MC model of the system.

Last but not least, while this work focuses of strictly dosimetric aspects of the Australian MRI-linac, its imaging performance, including potential interactions between imaging and beam delivery, had been described previously (29). It should also be emphasized that the integration of the whole system has been tested and that the first live animal treatments have been conducted recently (51), as a further step prior to clinical treatments.

CONCLUSION

Owing to the fundamentally different design, the inline systems display a different set of dosimetric issues as compared to transverse designs, most notably: no field shift and penumbra asymmetry, no build-up depth reduction, no electron return effect, the presence of electron focusing, weaker effects on detector response and less pronounced air gap effects. In this work, the methods were developed and employed to experimentally investigate and demonstrate these properties for the first time on an inline MRI-linac system. The collected measurements were used to fine-tune and commission the radiation related aspects of the Australian MRI-linac,

constituting a key step toward the application of inline MRI-linacs for patient treatments.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

AUTHOR CONTRIBUTIONS

UJ, PK, and GL contributed to the design of this work. UJ, BD, JB, NR, and BW contributed to the acquisition of data. UJ, JB, NR, and BW contributed to the analysis and interpretation of data. UJ drafted the manuscript. All authors revised the manuscript and approved the content for publication.

FUNDING

ACKNOWLEDGMENTS

REFERENCES

1. Lagendijk JJ, Raaymakers BW, Raaijmakers AJ, Overweg J, Brown KJ, Kerkhof EM, et al. MRI/linac integration. *Radiother Oncol.* (2008) 86:25–9. doi: 10.1016/j.radonc.2007.10.034

2. Liney GP, Whelan B, Oborn B, Barton M, Keall P. MRI-linear accelerator radiotherapy systems. *Clin Oncol.* (2018) 30:686–91. doi: 10.1016/j.clon.2018.08.003

3. Raaymakers BW, Jürgenliemk-Schulz IM, Bol GH, Glitzner M, Kotte ANTJ, van Asselen B, et al. First patients treated with a 1.5 T MRI-Linac: clinical proof of concept of a high-precision, high-field MRI guided radiotherapy treatment. *Phys Med Biol.* (2017) 62:L41–50. doi: 10.1088/1361-6560/aa9517

4. Mutic S, Dempsey JF. The ViewRay system: magnetic resonance-guided and controlled radiotherapy. *Semin Radiat Oncol.* (2014) 24:196–9. doi: 10.1016/j.semradonc.2014.02.008

5. Fallone BG, Murray B, Rathee S, Stanescu T, Steciw S, Vidakovic S, et al. First MR images obtained during megavoltage photon irradiation from a prototype integrated linac-MR system. *Med Phys.* (2009) 36:2084–8. doi: 10.1118/1.3125662

6. Keall PJ, Barton M, Crozier S. The Australian magnetic resonance imaging-linac program. *Semin Radiat Oncol.* (2014) 24:203–6. doi: 10.1016/j.semradonc.2014.02.015

7. Tijssen RHN, Philippens MEP, Paulson ES, Glitzner M, Chugh B, Wetscherek A, et al. MRI commissioning of 1.5T MR-linac systems - a multi-institutional study. *Radiother Oncol.* (2019) 132:114–20. doi: 10.1016/j.radonc.2018.12.011

8. St Aubin J, Santos DM, Steciw S, Fallone BG. Effect of longitudinal magnetic fields on a simulated in-line 6 MV linac. *Med Phys.* (2010) 37:4916–23. doi: 10.1118/1.3481513

9. Whelan B, Holloway L, Constantin D, Oborn B, Bazalova-Carter M, Fahrig R, et al. Performance of a clinical gridded electron gun in magnetic fields: implications for MRI-linac therapy. *Med Phys.* (2016) 43:5903–14. doi: 10.1118/1.4963216

10. Raaymakers BW, Raaijmakers AJ, Kotte AN, Jette D, Lagendijk JJ. Integrating a MRI scanner with a 6 MV radiotherapy accelerator: dose deposition in a transverse magnetic field. *Phys Med Biol.* (2004) 49:4109–18. doi: 10.1088/0031-9155/49/17/019

11. Bielajew AF. The effect of strong longitudinal magnetic fields on dose deposition from electron and photon beams. *Med Phys.* (1993) 20:1171–9. doi: 10.1118/1.597149

12. Keyvanloo A, Burke B, Warkentin B, Tadic T, Rathee S, Kirkby C, et al. Skin dose in longitudinal and transverse linac-MRIs using Monte Carlo and realistic 3D MRI field models. *Med Phys.* (2012) 39:6509–21. doi: 10.1118/1.4754657

13. Hackett SL, van Asselen B, Wolthaus JWH, Bluemink JJ, Ishakoglu K, Kok J, et al. Spiraling contaminant electrons increase doses to surfaces outside the photon beam of an MRI-linac with a perpendicular magnetic field. *Phys Med Biol.* (2018) 63:095001. doi: 10.1088/1361-6560/aaba8f

14. Raaijmakers AJ, Raaymakers BW, Lagendijk JJ. Integrating a MRI scanner with a 6 MV radiotherapy accelerator: dose increase at tissue-air interfaces in a lateral magnetic field due to returning electrons. *Phys Med Biol.* (2005) 50:1363–76. doi: 10.1088/0031-9155/50/7/002

15. Oborn BM, Ge Y, Hardcastle N, Metcalfe PE, Keall PJ. Dose enhancement in radiotherapy of small lung tumors using inline magnetic fields: a Monte Carlo based planning study. *Med Phys.* (2016) 43:368–77. doi: 10.1118/1.4938580

16. Liney GP, Dong B, Begg J, Vial P, Zhang K, Lee F, et al. Experimental results from a prototype high-field inline MRI-linac. *Med Phys.* (2016) 43:5188–94. doi: 10.1118/1.4961395

17. Oborn BM, Metcalfe PE, Butson MJ, Rosenfeld AB, Keall PJ. Electron contamination modeling and skin dose in 6 MV longitudinal field MRIgRT: Impact of the MRI and MRI fringe field. *Med Phys.* (2012) 39:874–90. doi: 10.1118/1.3676181

18. Meijsing I, Raaymakers BW, Raaijmakers AJ, Kok JG, Hogeweg L, Liu B, et al. Dosimetry for the MRI accelerator: the impact of a magnetic field on the response of a Farmer NE2571 ionization chamber. *Phys Med Biol.* (2009) 54:2993–3002. doi: 10.1088/0031-9155/54/10/002

19. Reynolds M, Fallone BG, Rathee S. Dose response of selected ion chambers in applied homogeneous transverse and longitudinal magnetic fields. *Med Phys.* (2013) 40:042102. doi: 10.1118/1.4794496

20. Reynolds M, Fallone BG, Rathee S. Technical Note: response measurement for select radiation detectors in magnetic fields. *Med Phys.* (2015) 42:2837–40. doi: 10.1118/1.4919681

21. O'Brien DJ, Roberts DA, Ibbott GS, Sawakuchi GO. Reference dosimetry in magnetic fields: formalism and ionization chamber correction factors. *Med Phys.* (2016) 43:4915–27. doi: 10.1118/1.4959785

22. Reynolds M, Fallone BG, Rathee S. Dose response of selected solid state detectors in applied homogeneous transverse and longitudinal magnetic fields. *Med Phys.* (2014) 41:092103. doi: 10.1118/1.4893276

23. van Asselen B, Woodings SJ, Hackett SL, van Soest TL, Kok JGM, Raaymakers BW, et al. A formalism for reference dosimetry in photon beams in the presence of a magnetic field. *Phys Med Biol.* (2018) 63:125008. doi: 10.1088/1361-6560/aac70e

24. Malkov VN, Rogers DWO. Monte Carlo study of ionization chamber magnetic field correction factors as a function of angle and beam quality. *Med Phys.* (2018) 45:908–25. doi: 10.1002/mp.12716

25. O'Brien DJ, Dolan J, Pencea S, Schupp N, Sawakuchi GO. Relative dosimetry with an MR-linac: Response of ion chambers, diamond, and diode detectors for off-axis, depth dose, and output factor measurements. *Med Phys.* (2018) 45:884–97. doi: 10.1002/mp.12699

26. Hackett SL, van Asselen B, Wolthaus JW, Kok JG, Woodings SJ, Lagendijk JJ, et al. Consequences of air around an ionization chamber: are existing solid phantoms suitable for reference dosimetry on an MR-linac? *Med Phys.* (2016) 43:3961–8. doi: 10.1118/1.4952727

27. Agnew J, O'Grady F, Young R, Duane S, Budgell GJ. Quantification of static magnetic field effects on radiotherapy ionization chambers. *Phys Med Biol.* (2017) 62:1731–43. doi: 10.1088/1361-6560/aa5876

28. Woodings SJ, Bluemink JJ, de Vries JHW, Niatsetski Y, van Veelen B, Schillings J, et al. Beam characterisation of the 1.5 T MRI-linac. *Phys Med Biol.* (2018) 63:085015. doi: 10.1088/1361-6560/aab566

29. Liney GP, Dong B, Weber E, Rai R, Destruel A, Garcia-Alvarez R, et al. Imaging performance of a dedicated radiation transparent RF coil on a 1.0 Tesla inline MRI-linac. *Phys Med Biol.* (2018) 63:135005. doi: 10.1088/1361-6560/aac813

30. Billas M. Homer M, Duane S. *Report on Dose Measurements on the MRI-linac at Liverpool Hospital (Australia) Performed by NPL.* London: National Physical Laboratory (2018). NPL Report IR 46.

31. Woodings SJ, Wolthaus JWH, van Asselen B, de Vries JHW, Kok JGM, Lagendijk JJW, et al. Performance of a PTW 60019 microDiamond detector in a 1.5 T MRI-linac. *Phys Med Biol.* (2018) 63:05NT04. doi: 10.1088/1361-6560/aaa1c6

32. Roberts NF, Patterson E, Jelen U, Causer T, Holloway L, Liney G, et al. Experimental characterization of magnetically focused electron contamination at the surface of a high-field inline MRI-linac. *Med Phys.* (2019) 46:5780–9. doi: 10.1002/mp.13847

33. IEC. *Medical Electrical Equipment – Dosimeters With Ionization Chambers as Used in Radiation Therapy (60731).* International Electrotechnical Commission (1997).

34. Smit K, Kok JG, Lagendijk JJ, Raaymakers BW. Performance of a multi-axis ionization chamber array in a 1.5 T magnetic field. *Phys Med Biol.* (2014) 59:1845–55. doi: 10.1088/0031-9155/59/7/1845

35. Raaijmakers AJ, Raaymakers BW, Lagendijk JJ. Experimental verification of magnetic field dose effects for the MRI-accelerator. *Phys Med Biol.* (2007) 52:4283–91. doi: 10.1088/0031-9155/52/14/017

36. Delfs B, Schonfeld AA, Poppinga D, Kapsch RP, Jiang P, Harder D, et al. Magnetic fields are causing small, but significant changes of the radiochromic EBT3 film response to 6 MV photons. *Phys Med Biol.* (2018) 63:035028. doi: 10.1088/1361-6560/aa9bd5

37. Niroomand-Rad A, Blackwell CR, Coursey BM, Gall KP, Galvin JM, McLaughlin WL, et al. Radiochromic film dosimetry: recommendations of AAPM Radiation Therapy Committee Task Group 55. *Med Phys.* (1998) 25:2093–115. doi: 10.1118/1.598407

38. Micke A, Lewis DF, Yu X. Multichannel film dosimetry with nonuniformity correction. *Med Phys.* (2011) 38:2523–34. doi: 10.1118/1.3576105

39. Whelan B, Kolling S, Oborn BM, Keall P. Passive magnetic shielding in MRI-Linac systems. *Phys Med Biol.* (2018) 63:075008. doi: 10.1088/1361-6560/aab138

40. IEC. *Medical Electrical Equipment – Medical Electron Accelerators – Functional Performance Characteristics (60976).* International Electrotechnical Commission (1989).

41. IEC. *Medical Electrical Equipment – Medical Electron Accelerators in the Range 1 MeV to 50 MeV – Guidelines for Functional Performance Characteristics (60977).* International Electrotechnical Commission (1989).

42. Oborn BM, Kolling S, Metcalfe PE, Crozier S, Litzenberg DW, Keall PJ. Electron contamination modeling and reduction in a 1 T open bore inline MRI-linac system. *Med Phys.* (2014) 41:0517908. doi: 10.1118/1.4871618

43. Budgell G, Brown K, Cashmore J, Duane S, Frame J, Hardy M, et al. IPEM topical report 1: guidance on implementing flattening filter free (FFF) radiotherapy. *Phys Med Biol.* (2016) 61:8360–94. doi: 10.1088/0031-9155/61/23/8360

44. Fogliata A, Garcia R, Knoos T, Nicolini G, Clivio A, Vanetti E, et al. Definition of parameters for quality assurance of flattening filterfree (FFF) photon beams in radiation therapy. *Med Phys.* (2012) 39:6455–64. doi: 10.1118/1.4754799

45. Das IJ, Cheng CW, Watts RJ, Ahnesjö A, Gibbons J, Li XA, et al. Accelerator beam data commissioning equipment and procedures: report of the TG-106 of the Therapy Physics Committee of the AAPM. *Med Phys.* (2008) 35:4186–215. doi: 10.1118/1.2969070

46. Zhu XR, Kang Y, Gillin MT. Measurements of in-air output ratios for a linear accelerator with and without the flattening filter. *Med Phys.* (2006) 33:3723–33. doi: 10.1118/1.2349695

47. Andreo P, Burns D, Hohlfeld K, Huq M, Kanai T, Laitano F, et al. *IAEA TRS-398: Absorbed Dose Determination in External Beam Radiotherapy: An International Code of Practice for Dosimetry based on Standards of Absorbed Dose to Water.* Vienna: International Atomic Energy Agency (2000).

48. Arnfield MR, Siebers JV, Kim JO, Wu Q, Keall PJ, Mohan R. A method for determining multileaf collimator transmission and scatter for dynamic intensity modulated radiotherapy. *Med Phys.* (2000) 27:2231–41. doi: 10.1118/1.1312190

49. Patel I, Glendinning AG, Kirby MC. Dosimetric characteristics of the Elekta Beam Modulator. *Phys Med Biol.* (2005) 50:5479–92. doi: 10.1088/0031-9155/50/23/004

50. Clivio A, Belosi MF, Cozzi L, Nicolini G, Vanetti E, Bolard G, et al. On the determination of reference levels for quality assurance of flattening filter free photon beams in radiation therapy. *Med Phys.* (2014) 41:021713. doi: 10.1118/1.4861817

51. Liney GP, Jelen U, Byrne H, Dong B, Roberts TL, Kuncic Z, et al. Technical Note: the first live treatment on a 1.0 Tesla inline MRI-linac. *Med Phys.* (2019) 46:3254–8. doi: 10.1002/mp.13556

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: info@frontiersin.org | +41 21 510 17 00

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