IOP Publishing Phys. Med. Biol. 65 (2020) 025009 (12pp) https://doi.org/10.1088/1361-6560/ab5c5b

Physics in Medicine & Biology

IPEM Institute of Physics and Engineering in Medicine

PAPER

Automated air region delineation on MRI for synthetic CT creation

RECEIVED
23 June 2019

REVISED
22 November 2019

ACCEPTED FOR PUBLICATION
27 November 2019

PUBLISHED
16 January 2020

Ranjeeta Thapa, Ergun Ahunbay, Haidy Nasief, Xinfeng Chen and X Allen Li[1]

Radiation Oncology, Medical College of Wisconsin, Milwaukee, WI, United States of America

[1] Author to whom any correspondence should be addressed.

E-mail: ali@mcw.edu

Keywords: MRI, synthetic CT, air region delineation, online adaptive radiation therapy

Supplementary material for this article is available online

Abstract

Automatically and accurately separating air from other low signal regions (especially bone, liver, etc) in an MRI is difficult because these tissues produce similar MR intensities, resulting in errors in synthetic CT generation for MRI-based radiation therapy planning. This work aims to develop a technique to accurately and automatically determine air-regions for MR-guided adaptive radiation therapy.

CT and MRI scans (T2-weighted) of phantoms with fabricated air-cavities and abdominal cancer patients were used to establish an MR intensity threshold for air delineation. From the phantom data, air/tissue boundaries in MRI were identified by CT-MRI registration. A formula relating the MRI intensities of air and surrounding materials was established to auto-threshold air-regions. The air-regions were further refined by using quantitative image texture features. A naive Bayesian classifier was trained using the extracted features with a leave-one-out cross validation technique to differentiate air from non-air voxels. The multi-step air auto-segmentation method was tested against the manually segmented air-regions. The dosimetry impacts of the air-segmentation methods were studied.

Air-regions in the abdomen can be segmented on MRI within 1 mm accuracy using a multi-step auto-segmentation method as compared to manually delineated contours. The air delineation based on the MR threshold formula was improved using the MRI texture differences between air and tissues, as judged by the area under the receiver operating characteristic curve of 81% when two texture features (autocorrelation and contrast) were used. The performance increased to 82% with using three features (autocorrelation, sum-variance, and contrast). Dosimetric analysis showed no significant difference between the auto-segmentation and manual MR air delineation on commonly used dose volume parameters.

The proposed techniques consisting of intensity-based auto-thresholding and image texture-based voxel classification can automatically and accurately segment air-regions on MRI, allowing synthetic CT to be generated quickly and precisely for MR-guided adaptive radiation therapy.

1. Introduction

Magnetic resonance image (MRI) guided radiation therapy, like the integrated MRI and linear accelerator (MR-Linac), is anticipated to improve radiation therapy by enhancing tumor targeting (Lagendijk *et al* 2014). The high soft tissue contrast MRIs acquired before and during each treatment delivery allow for online and real time plan adaptation. The lack of electron density information in MRI requires the generation of a synthetic CT based on the daily MRI for dose calculation (Johansson *et al* 2011, Dowling *et al* 2012, Hsu *et al* 2013, Edmund *et al* 2014, Korhonen *et al* 2014, Gudur *et al* 2014, Delso *et al* 2014, Burgos *et al* 2015, Sjolund *et al* 2015, Siversson *et al* 2015, Kim *et al* 2015, Dowling *et al* 2015, Andreasen *et al* 2015, Arabi *et al* 2016, Andreasen *et al* 2016, Edmund and Nyholm 2017, Kraus *et al* 2017, Guerreiro *et al* 2017, Han 2017, Ahunbay *et al* 2019). Widely used approaches for synthetic CT generation include voxel-, atlas-, and hybrid-based methods. In a voxel-based method (Johansson

et al 2011, Edmund et al 2014, Korhonen et al 2014), the electron density data is directly obtained for each voxel based on its MR intensity in which the electron density conversion is independent of the location of the voxel. The atlas-based method (Dowling et al 2012, Burgos et al 2015, Sjolund et al 2015, Arabi et al 2016, Edmund and Nyholm 2017, Ahunbay et al 2019) uses synthetic CT intensity of a voxel determined from its location in anatomy, typically by registration with a reference image that already has the electron density information, while a hybrid-based method uses both the voxel intensity and the spatial information (Gudur et al 2014, Siversson et al 2015). These methods, however, suffer from several drawbacks. For example, Johansson et al (Johansson et al 2011) showed that the largest errors in the synthetic CT were found at air/tissue and bone/tissue interfaces and can have dosimetric consequences. The voxel-based methods suffer from limitations due to the very short T2 relaxation time of bone, making it hard to discriminate between air and bone (Robson et al 2003, Johansson et al 2011, Shah et al 2011, Edmund et al 2014, Korhonen et al 2014). Although additional MRI sequences have been designed to specifically image air and/or bone regions, these additional scans would extend imaging time and are not desirable for online adaptation.

Furthermore, with the MR-Linac, patients are irradiated in the presence of a magnetic field that can significantly affect the dose distribution due to the Lorentz force perturbing the secondary electrons specifically at air/tissue boundary. This phenomenon is called the electron return effect (Raaymakers et al 2004, Raaijmakers et al 2005). The dose change due to the electron return effect at tissue interfaces is dependent on the magnetic field strength and surface orientation, which can cause an under/over dose at interfaces, e.g. the air cavity walls, and the lung/tissue interfaces (Prior et al 201, Raaymakers et al 2004, Raaijmakers et al 2005, Raaijmakers et al 2007a, Oborn et al 2009). The change in the dose due to the electron return effect has been studied by several groups (Raaymakers et al 2004, Raaijmakers et al 2005, Raaijmakers et al 2007a, Oborn et al 2009, Bol et al 2012, van Heijst et al 2013, Chen et al 2016). For instance, Raaijmakers et al (Raaijmakers et al 2007a) investigated whether intensity modulated radiation therapy (IMRT) could be used to compensate for the electron return effect on prostate, larynx, and oropharyngeal cases, and reported that the presence of a 1.5 T magnetic field does not compromise the ability to achieve desired dose distributions with IMRT. Chen et al (Chen et al 2016) investigated the impact of a 1.5 T magnetic field on realistic IMRT/volumetric modulated arc therapy plans at various tumor sites where air/tissue or lung-tissue interfaces are commonly present in head and neck, breast, pancreas, and lungs and found that the dose effects on air/tissue or tissue/lung interfaces can vary significantly depending on the site, size, and geometry of the air cavities if the magnetic field was not included in the plan optimization. The authors also found that the dose effects from the magnetic field can be substantially reduced, or even eliminated, by including the magnetic field in the plan optimization, resulting in no deterioration in overall plan quality.

Overall, these studies concluded that the presence of air regions can cause significant dosimetric consequences, and it can be even more complicated for MR-guided radiation therapy because of the presence of the magnetic field. Since these regions happen randomly in the abdomen and pelvis, their shapes in daily MRI will be different from the reference image and cannot be precisely segmented with the current deformable image registration methods. Further, a simple image intensity threshold would pick up many other regions, since MR intensities in these regions may be similar to the air intensity as shown in figure 1. An ultimate solution of this issue has not been reported. We have previously developed a general atlas-based method to create synthetic CT by transferring electron density from CT to MRI with special considerations for bone and air regions (Ahunbay et al 2019). This method can be used to generate relative electron density map for extra-cranial sites using the patient's own CT instead of patient's average atlases. However, there was no robust method for air cavity delineation included in the previous study. Thus, in this study, we extend our previous work by introducing a technique to automatically and accurately delineate air on MRI for MRI-guided adaptive radiation therapy of abdominal tumors. We have developed an equation based on phantom and patient data, which will be used to generate synthetic CT from daily MR-guided adaptive radiation therapy.

2. Materials and methods

The proposed air cavity auto-segmentation method uses multiple steps including delineating an air cavity based on an MRI intensity threshold derived from phantom and patient data, and refining the air boundary based on quantitative image texture feature classification.

2.1. General explanation of the method

The proposed method for air volume generation has the following steps. Step 1 is performed offline, while the rest of the steps are online with all online processes are being performed automatically. Details of each step are explained further along in the materials and methods section.

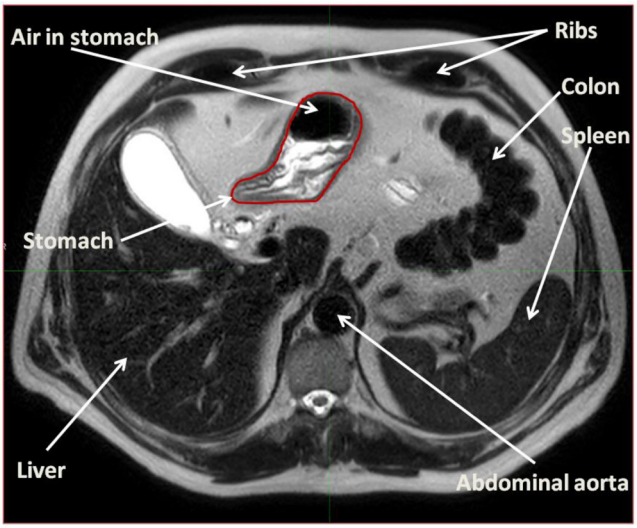

Figure 1. A typical MR image showing many low MR signal regions such as bone, vessel, colon, and liver etc that can be misidentified as air.

1. Generate a gross air volume on the reference CT image. This is the union of already delineated organs of GI track, i.e. stomach, duodenum, bowel, colon and rectum, whichever is present on the CT scan. This structure is expanded by 1 cm. This process is performed only once offline.
2. Transfer the gross air volume to daily MR via deformable image registration.
3. Generate the threshold value automatically via a linear formula. This process includes the generation of the background volume as explained below.
4. Apply the threshold value only to the gross air volume, therefore any regions that would be picked up by the threshold outside the gross air volume will be discarded.
5. Further refine the air volume by texture-based discrimination.

Figure 2 illustrated the step-by-step implementation of the method for MRI-guided online adaptive replanning.

We developed and tested this method with 54 CT and MR images from 51 pancreatic cancer patients and 3 in-house constructed air cavity phantoms of MRI and CT were used. Details of the image acquisitions are provided below.

2.1.1. CT acquisition

All CT data were acquired using a CT scanner (Sensation, Siemens) with standard abdominal protocols and the following settings: 120 kVp, 252 mA, 512 × 512 in-plane image dimensions, 1.28 × 1.28 mm^2 in-plane spatial resolution, and 3 mm slice thickness.

2.1.2. MR acquisition

To ensure the developed technique is applicable for commonly used MRI sequences for MR-guided radiation therapy, we used MRI data acquired using a 3 T MRI (Trio, Siemens) and a 1.5 T scanner in a MR-LINAC (Unity, Elekta) with five different T1 and T2 weighted sequences: (1) Axial T1 weighted 3D breath hold (fat-suppressed) spoiled gradient echo TR/TE of 3.92/1.39 ms, 9° flip angle, 401 hertz/pixel pixel bandwidth, 3 mm slice thickness, and 380 mm in-plane field-of-view (FOV); (2) T1 weighted 3D spoiled gradient echo TR/TE of 7.16/4.5 ms, 30° flip angle, 434 hertz/pixel pixel bandwidth, 4 mm slice thickness, and 71 mm FOV; (3) T1 weighted 3D axial spoiled gradient echo TR/TE of 6.8/3.3 ms, 25° flip angle, 434 hertz/pixel pixel bandwidth, 5 mm slice thickness, and 400 mm FOV; (4) T2 weighted 2D turbo spin echo TR/ TE of 3800/96 ms, 90° flip angle, 373 hertz/pixel pixel bandwidth, 5 mm slice thickness, and 380 mm FOV; and (5) T2 weighted respitory triggered axial 2D single-shot-fast spin echo TR/TE of 2000/93 ms, 90° flip angle, 468 hertz/pixel pixel bandwidth, 5 mm slice thickness, and 210 mm FOV. The first two sequences (mentioned above) were from the Trio scanner and the last three sequences were from the Unity machine.

2.1.3. Standardization of MR images

MRI and CT data from 51 pancreatic cancer patients were used to develop and validate the air auto-segmentation method. Since the MRI data came from 2 different scanners with 5 different types of MR sequences, large intensity

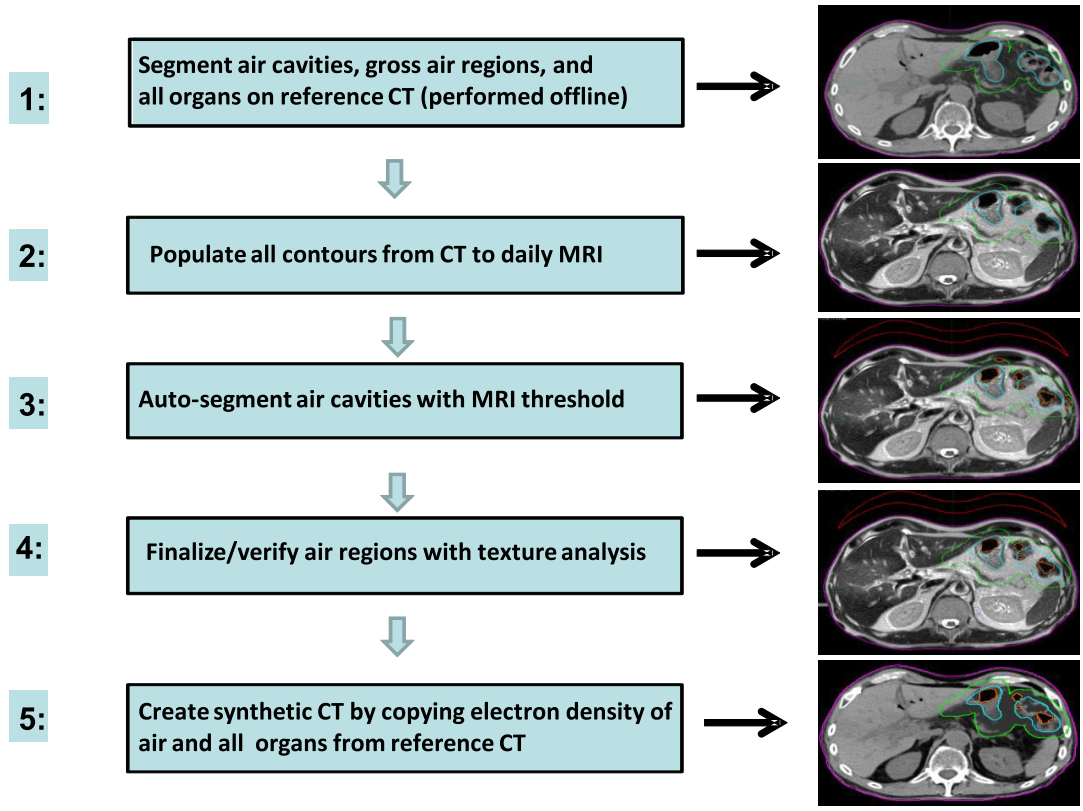

Figure 2. A five-step implementation of the method to automatically generate synthetic CT MRI-guided online adaptation.

variation was expected (supplemental table 1). These variations in intensity could affect thresholding of the image. To account for such variation, the Gaussian approach for intensity normalization (Ellingson *et al* 2012) was used. This approach rescaled the intensity by: $Inew = I/SD$, where I is the intensity and SD is the standard deviation of patient's whole-body contour (Ellingson *et al* 2012). Unlike the situation in the phantoms, the air cavities in the CT and MRI are generally different for a patient. The MRI-based manual air contours (MACs) were created by warping the CT-based MACs to the MRIs via deformable image registration (Admire, Elekta) and followed by manual editing by carefully reviewing the CT and the adjacent MRI slices using the contouring tools in a software (MIM software, Cleveland OH).

2.1.4. Generating the Air Thresholding Formula by Regression (Step 3 mentioned above)

The main component of this method is to develop a formula to automatically generate the threshold value for the daily MR using only the image intensities extracted from automatically generated region of interests (ROIs) on the MR image. We used phantom and patient MR and CT images to derive a linear formula, and empirically derived the parameters for it by linear regression.

The rationale behind this formula is to have an automatically generated air threshold value that is practically the same as a human would visually determine. Since air region varies greatly even for the same patient in daily fractions, manually determining a threshold suitable for the daily image may require multiple iterations that can take a long time and a lot of human intervention.

The formula has the form of:

$$MR_{\text{thres-pat}} = a * MR_{\text{background}} + b * MR_{\text{mean}} + c * MR_{\text{std}} + d * MR_{\text{thres-phan}}. \quad (1)$$

Where we estimated the 4 parameters (a, b, c, and d) via linear regression by fitting to manually determined threshold values from 33 patients for training dataset (training set in supplemental table 1). An open access linear matrix solver software (http://www.bluebit.gr/matrix-calculator/solve.aspx) was used to calculate these coefficients. The values used in the formula are as follows:

$MR_{\text{background}}$: the background MR intensity value from the air signal at 2 cm outside of the patient's external contour (see figure 9). This is acquired automatically by generating expansion and Boolean operations available on MIM software. ROI changes according to the size of the patient. The mean ROI volume for all patients used in our study was 991.1 ± 70.0 cc. The high standard deviation correlates with the large variation in patients' size since the patients' external contour would vary based on patient size.

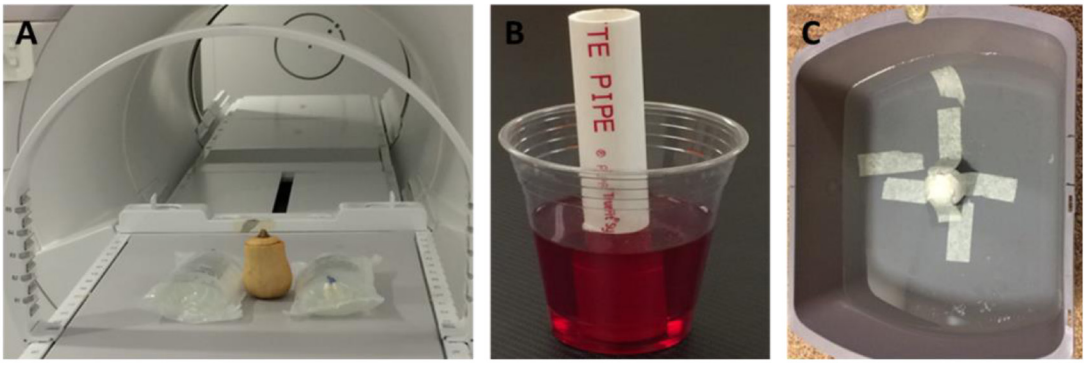

Figure 3. Photographs of three air cavity phantoms: (A) a squash with air cavity, (B) jello and the pipe used to extract the jello to create an air cavity, and (C) a 4 cm diameter ping-pong ball in water.

$MR_{\text{thres-pat}}$: This value is the mean value of the air intensity threshold range (MR_{range}) provided in supplemental table 1 (stacks.iop.org/PMB/65/025009/mmedia). This MR_{range} is established by trying different MR intensity values to contour the air region and judging with the operator's visual observations whether it picks air or not. It is observed that any value within this range can represent the air region equally accurate as far as the operator can tell. So, the mean value of this range is used as $MR_{\text{thres-pat}}$ in the equation.
MR_{mean} and MR_{std}: the mean and standard deviation of the MR intensities in the gross air region (this volume is explained below).

The gross air region is the region where air/gas can be expected. This is used for restricting the volume where air volume will be generated by thresholding. This gross volume is the union of all gas containing structures e.g. stomach, duodenum, bowel, colon, and rectum for abdominal and pelvis patients. The gross volume is generated offline and then transferred to the daily MRI by deformable image registration (as explained in step 2 above). The gross volume on the daily MR is further expanded by 1 cm in all directions to generate the extended gross volume. If the deformable image registration process had no inaccuracies, all air/gas on the daily MR would be expected to be inside the gross volume, and by applying a 1 cm expansion to account for the deformable image registration inaccuracies, we ascertain that all air on the daily MR is within this extended gross volume. All low intensity voxels generated by thresholding (below) that are outside the extended gross volume will not be included in the derived air volume.

2.1.5. *Phantom data to determine $MR_{thres\text{-}phan}$*

Three phantoms (figure 3) were generated with imbedded air cavities, a squash with an air cavity created by drilling and smoothing the inside, a cup of jello with a pipe used to create an air cavity by extracting the jello out of the pipe, and a 4 cm diameter ping-pong ball imbedded in water. CTs and MRIs of these phantoms were acquired in the same setup with the same imaging protocols as those used in our clinic. T2 MRIs with acquisition parameters including 2D turbo spin echo TR/TE of 3800/96 ms were analyzed. The air cavities were delineated in both CTs and MRIs manually, and hereafter will be called manual air contours (MACs). In addition, the CT-based MACs were transferred to the MRIs by rigid image registration and were compared with the MRI-based MACs. Dice coefficient and Housdorff distance (Zou *et al* 2004, Taha and Hanbury 2014, Taha and Hanbury 2015) were used to measure the differences between the CT- and MRI-based MACs, such that for two 3D regions, A and B, the Dice measures the degree of overlap and is defined as

$$Dice = \frac{2 * |A \cap B|}{(|A| + |B|)}. \tag{2}$$

Where $|A|$ is the volume of region A, and $|A \cap B|$ is the volume of the intersection of regions A and B. If A and B perfectly overlap, dice similarity coefficient $= 1$.

The Housdorff distance measures the maximum distance from each point on A to the nearest point on B. More formally, denoting the distance between a point a on the contour of A and point b on the contour of B as $d(a, b)$ with perfectly overlapping contours giving Housdorff distance $= 0$, the Housdorff distance is defined as:

$$\text{Housdorff distance} = \max a \in A \ \{\min b \in B \{d(a, b)\}\}. \tag{3}$$

The MR intensity threshold from phantoms ($MR_{\text{thres-phan}}$) was determined based on the delineation of the MRI contours that leads to minimum difference with the CT-based contours.

2.1.6. Air boundary refinement based on image textures (Step 5 above)

In the abdomen, the air contours determined from the threshold method may contain regions of low intensity that are not actually air, e.g. certain vessels, and/or stool in the stomach or bowels. To check and to refine the air boundary on the contours from the threshold method, image texture features in air and non-air regions were extracted on a voxel-by-voxel basis. Features selected came from the gray level co-occurrence matrix (GLCM) (Haralick *et al* 1973, Amadasun and King 1989, Castellano *et al* 2004, Holub and Ferreira 2006). This feature matrix was selected because it represents the second order statistics that quantify the frequency at which gray-level intensity appears adjacent to other gray-level intensity in a particular direction, thus representing the spatial information of the image. GLCM features were calculated in different 2D directions (e.g. 0, 45, 90, and 135) for each axial slice. To include data from all connected directions, to avoid directional dependence, and to reduce the number of features that can be used for our analysis, GLCMs for a distance $d = 1$ and a particular direction are summed over the set of axial slices. These direction-specific matrices were then summed and averaged to create the final GLCM for the 3D region of interest. A MATLAB script was written as an extension to the MIM imaging software, and was used to extract a total of 23 texture features including contrast, correlation, energy, homogeneity, autocorrelation, entropy, dissimilarity, cluster prominence, cluster tendency, maximum probability, difference entropy, homogeneity2, information measure of correlation 1, information measure2, inverse difference normalized, inverse difference moment normalized, inverse variance, sum average, sum entropy, sum variance, variance, and cluster shade from the air and non-air regions on a voxel by voxel basis. A student *T*-test was performed to identify features that differ significantly between air versus non-air regions. The combinations of different features were tested with a naive Bayesian classifier and a leave-one-out cross validation technique (where one data set is left out for testing and the rest were used for training) to determine the features with highest impact upon distinguishing the air from the non-air regions. The training was done on 33 data sets with a leave one out cross-validation technique. Testing was done on independent datasets from 18 different patients. The data extracted from the MRIs of selected pancreatic cancer cases were divided into a training set consisting of 5000 air and 5000 non-air voxels, and a testing set consisting of 1000 voxels each for air/non-air from different patients' data sets. Classification was performed using the minimum distance to the centroid of the training class. The classifier performance was judged using the area under the receiver operator characteristics (ROC) curve to identify features with highest impact on classification.

2.2. Validation of the air cavity auto-segmentation

The newly developed multi-step air cavity auto-segmentation method based on the MR threshold formula was validated with the CT and MRI data for 18 pancreatic cancer cases (testing set in supplemental table 1). The MRI MACs were compared with the automatic air contours (AACs) created by using the auto-segmentation method. The differences between the MACs and AACs were quantified using the dice coefficient, Housdorff distance and mean distance to agreement measurement. The dosimetric impact from the difference of MACs and ACCs was assessed by creating an IMRT plan based on AACs and recalculating the plan based on MACs using a research planning system (Monaco, Elekta) employing a Monte Carlo dose engine. Statistical uncertainty of 1% and a calculation grid size of 3 mm^3 were used in the dose calculation. All calculations were performed in the presence of a 1.5 T magnetic field orthogonal to the irradiation fields. Relative electron density in MACs and AACs was assigned to be 0.01 based on the recommendation from International Commission of Radiation Units Report 46 (ICRU-46) (White *et al* 1992). Dosimetric differences between MACs and AACs were measured by using commonly used dose volume parameters, e.g. *V*100%: the percentage volume covered by 100% prescription dose (50.40 Gy); D_{max}: the maximum dose; $D_{95\%}$: the dose covering 95% in a volume; $D_{1\,cc}$: the maximum dose covering 1 cm^3 volume; *V*45 Gy: volume covered by 45 Gy.

3. Results

3.1. MR intensity air threshold

The Dice coefficient between CT and MR air contours was maximized by selecting an appropriate image intensity for the contouring threshold. Figure 4 compares air contours using different intensity thresholds in MRI (A) and CT (B) and presents the Dice coefficient as a function of MR and CT intensity (C) in the squash phantom. It is seen that the Dice coefficient showed a maximum at the MR T2 intensity of 50. Therefore $MR_{thres\text{-}phan} = 50$ was used in determining the threshold in patient MR data. Compared to other phantoms, the squash phantom includes pure air and adjacent region that can be confused with air. This scenario is similar to the situation in patient bowel where bowel contents can coexist with air.

Figure 5 compares the CT-based MACs with the MRI-based air contours created by the determined MR threshold in the three phantoms. It is seen that the spacing difference (e.g. Housdorff distance) between the CT and MRI air contours is within 1.0 mm. The average Housdorff distance values were 0.95 ± 0.07, 0.83 ± 0.09, and 0.74 ± 0.01 mm for the squash, jello, and ping-pong ball phantoms, respectively.

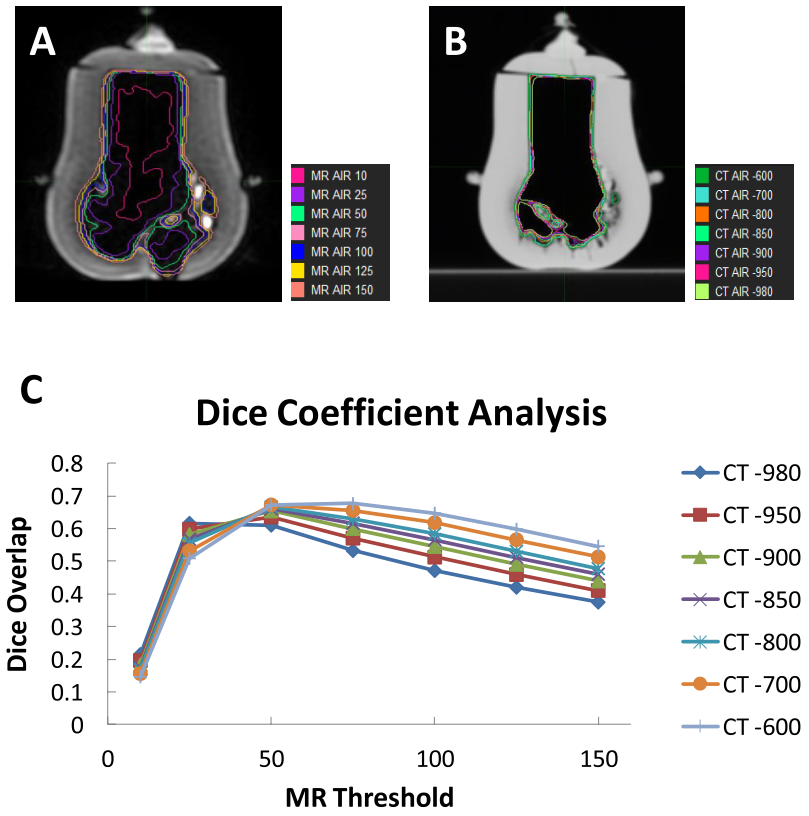

Figure 4. The air contours segmented with the different MR and CT (Housdorff distance) thresholds in MRI (A) and CT (B) and the Dice coefficient as a function of different thresholds (C). The maximum similarity between CT and MR contour pairs was identified at the MR T2 intensity of 50.

Based on the data of the 33 patient cases in training set, the coefficients of a, b, c and d in the threshold equation (1) were estimated to be 0.0403, 0.1564, 0.2961, and −0.1937, respectively. These coefficients were used to calculate the MR threshold for the air cavities on the 18 testing cases in supplemental table 1. The R square value of our model is 0.90. The normalized MR threshold of the patient (MR_{thres_pat}) is plotted in figure 6. The MR thresholds for the formula derived air contours ($MR_{formula}$) fall in the estimated range (MR_{range}) for all tested cases. The MR_{trial_error} value that comes from average of the MR_{range} was also close to the $MR_{formula}$.

3.2. Air boundary refinement

Figure 7 contains examples of histogram plots of three MRI texture features, auto correlation, sum variance, and contrast showing clear distinctions between air and non-air regions.

The best performing feature combinations that distinguish air and non-air regions are tabulated in table 1 along with the area under the ROC curve and the confidence intervals. Figure 8 shows the ROC curve of the combination of features with the highest impact on the classification. The best performing two texture feature combination (autocorrelation and contrast) showed an area under the ROC curve of 81%. The performance increased to 82% with using three features (autocorrelation, sum variance, and contrast).

Figure 9 compares (A) the air contours obtained based on the multi-step auto-segmentation method with (B) the MR-based MACs for a sample case. It is seen that the two contour sets generally agree with each other. The low signal regions, e.g. aorta, were not identified as air since they are outside of the gross air region.

3.3. Validation of the method

The average Dice coefficient, Housdroff distance and mean distance to agreement, values between MRI-based MACs and AACs created by the multi-step air auto-segmentation method was analyzed for all tested cases. The average value of Housdroff distance and mean distance to agreement were 0.84 ± 0.08, and 0.64 ± 0.32, respectively. The MACs and AACs overlap, with an average Dice coefficient of 0.92 ± 0.05. As an example, figure 10 presents comparisons of dose volume histograms obtained based on MACs and AACs. This example was taken from the image set that has the lowest value of Dice coefficient (0.81). There is no substantial difference between the two types of dose volume histograms. The differences between the MACs and AACs values of PTV

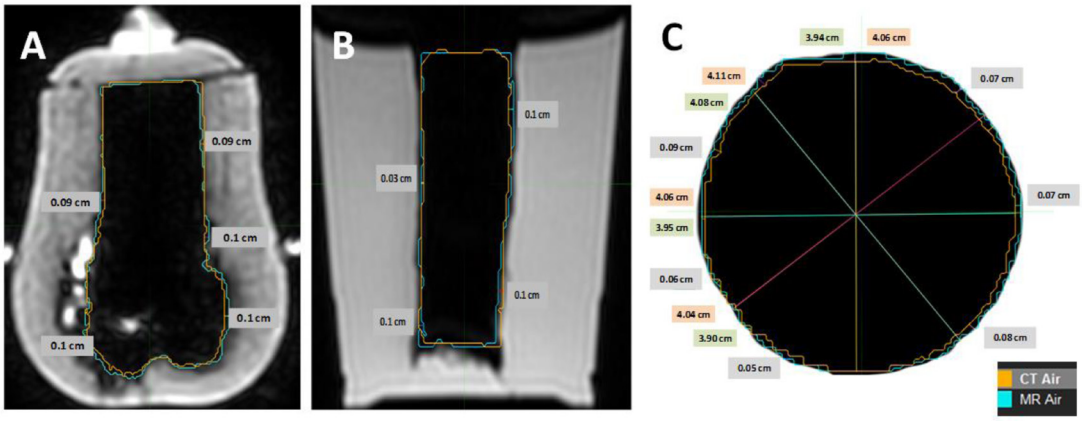

Figure 5. Air contours based on CT (orange) and MRI (cyan) are compared in three phantoms: (A) squash, (B) jello, and C) ping-pong ball. The numbers shown in (A) and (B) are the Housdorff distance between the CT and MRI contours, while the numbers in (C) are the diameters of sphere contours based on CT (yellow) and MRI (green) as well as their Housdorff distance (grey).

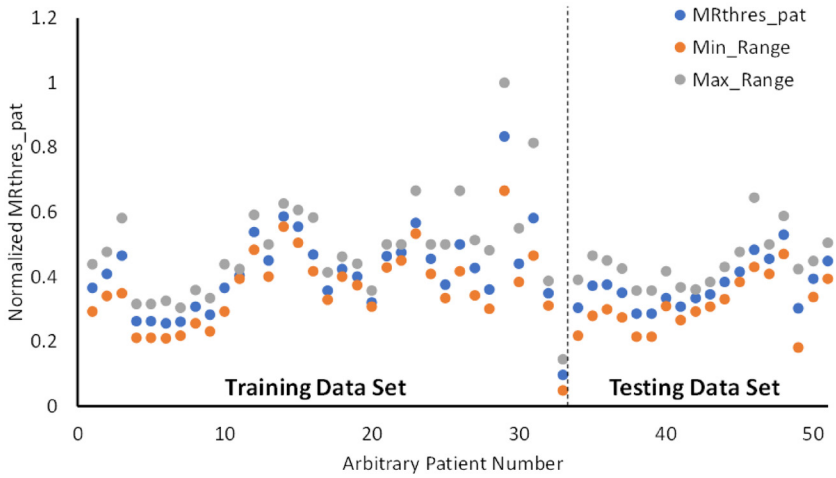

Figure 6. Normalized MR threshold of the patient (MR_{thres_pat}) versus arbitrary patient number. The whole data set is divided into training and testing data (represented by blue dots) separated by a vertical dashed line. The orange and gray dots are established minimum and maximum MR thresholds, respectively.

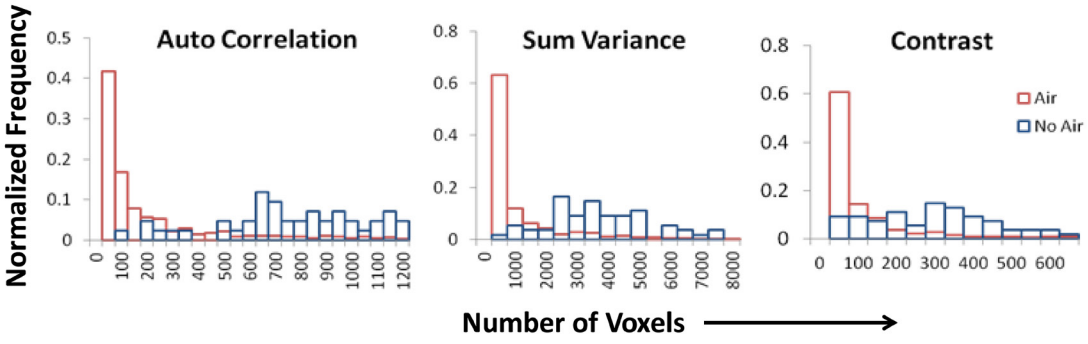

Figure 7. A comparison of histogram plots of three MRI texture features between air and non-air regions. The horizontal axis is the number of voxels with the same probability, and the vertical axis is the normalized frequency. To show the distinction between the histograms for air/no air, different scales are used.

Table 1. The best performing combinations of MRI features that is suitable to distinguish air and non-air regions.

Features	AUC	Confidence interval
Contrast, autocorrelation, sum variance, max probability, sum average	0.80	[0.76, 0.82]
Contrast, autocorrelation, sum variance, max probability	0.77	[0.79, 0.81]
Contrast, autocorrelation, sum variance	0.82	[0.79, 0.84]
Contrast, sum variance	0.80	[0.79, 0.82]
Contrast, autocorrelation	0.81	[0.79, 0.83]

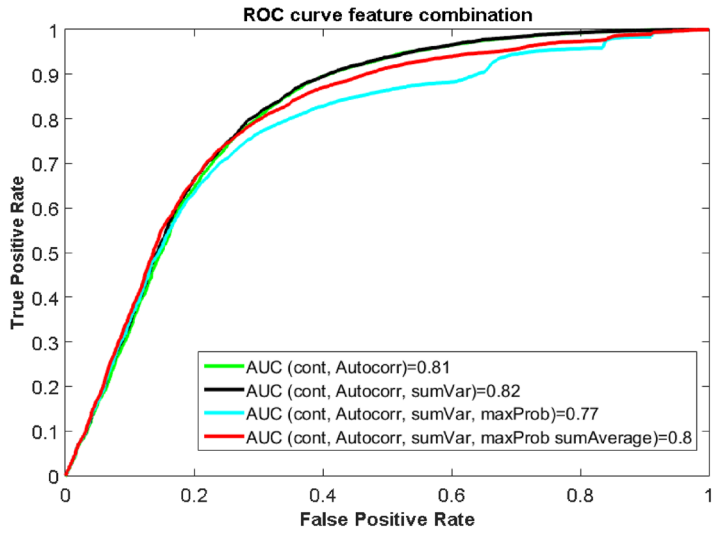

Figure 8. ROC curve of the best performing feature combinations. Green line shows the AUC combining contrast and autocorrelation (AUC = 0.81), black line shows maximum AUC of 0.82 combining contrast, autocorrelation and sum variance, cyan line demonstrates the AUC of 0.77 combining contrast, autocorrelation, sum variance and maximum probability. The red line gives AUC of 0.80 combining contrast, autocorrelation, sum variance, maximum probability, and sum average.

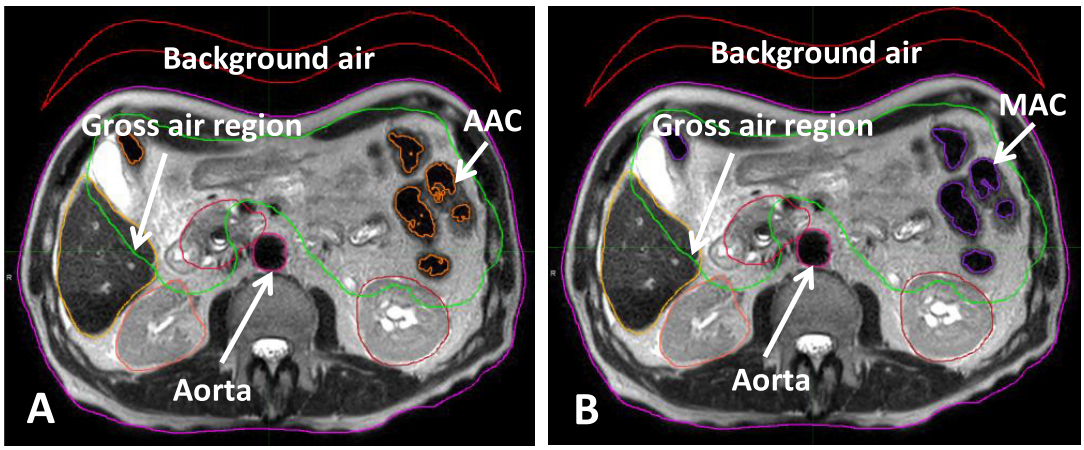

Figure 9. A comparison of (A) AAC (orange) and (B) the MRI-based MAC (purple). Low signal regions, e.g. aorta, are not identified as air because they are outside of the gross air contour (green).

*V*100, *D*max, *D*min, *D*mean, and *D*95, and duodenum *V*45 Gy and *V*18 Gy were 0.29, 0.63, 0.69, 0.78, 0.21, 0.89, and 0.36%, respectively.

4. Discussion

Techniques including voxel-, atlas-, and hybrid- based methods have been developed to generate a synthetic CT from MRIs over the past few years (Johansson *et al* 2011, Dowling *et al* 2012, Hsu *et al* 2013, Edmund *et al* 2014, Korhonen *et al* 2014, Gudur *et al* 2014, Delso *et al* 2014, Burgos *et al* 2015, Sjolund *et al* 2015, Siversson *et al* 2015, Kim *et al* 2015, Dowling *et al* 2015, Andreasen *et al* 2015, Arabi *et al* 2016, Andreasen *et al* 2016, Edmund and

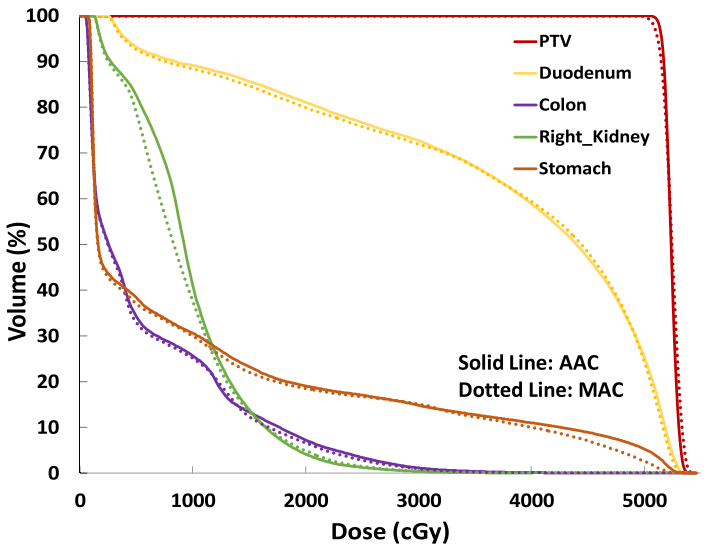

Figure 10. The dose volume histograms comparison between MACs and AACs, showing minimal difference between MACs and AACs for all types of organs at risk.

Nyholm 2017, Kraus *et al* 2017, Guerreiro *et al* 2017, Han 2017, Ahunbay *et al* 2019). However, none of these methods can be used to accurately and automatically assign electron density to air regions.

The accurate and fast delineation of the air for MRI is crucial in MR-guided adaptive radiation therapy. The dosimetric effect due to the magnetic field can be substantial at air/tissue intefaces (Raaymakers *et al* 2004, Raaijmakers *et al* 2005, Raaijmakers *et al* 2007a, Oborn *et al* 2009, Bol *et al* 2012, van Heijst *et al* 2013, Chen *et al* 2016) especially when treating abdominal and pelvic malignancies due to significant inter-and intra-fractional air cavities changes. Inaccurate delineation of air for synthetic CT creation can result in severe dose errors for the target or the critical organs. Thus, the air regions must be appropriately addressed while generating the synthetic CT for MR-guided adaptive radiation therapy (Prior *et al* 201, Raaymakers *et al* 2004, Raaijmakers *et al* 2005, Raaijmakers *et al* 2007b, Oborn *et al* 2009). A solution has been recently proposed by our group (Ahunbay *et al* 2019) in a study investigating the consequences of overriding the whole air containing organs (e.g. bowel, stomach etc) with different electron density against only overriding the air regions with air electron density, concluding that accurately identifying air regions was essential for synthetic CT generation (Ahunbay *et al* 2019). The current study addresses this issue, and focuses on a solution for MR-guided adaptive radiation therapy with MR-Linac.

MRI intensities vary between different sequences, scanners, subjects, and times. This affects algorithm performance, prediction, inference, and even simple things like thresholding an image. Thus, researchers usually enforce intensity normalization for image registration (Wang *et al* 1998, Hellier 2003, Shah *et al* 2011, Ellingson *et al* 2012, Shinohara *et al* 2014, Jean-Philippe *et al* 2016). In this study we used a simple technique; Gaussian approach (Ellingson *et al* 2012), which makes MR intensity comparable regardless of scanners and protocols, thereby improving the accuracy of an automatic segmentation of the air region on MRIs for MRI guided radiation therapy.

The study initially used phantoms with built-in air cavities to determine the MRI threshold. This phantom study was necessary to establish the ground truth of how the true air regions can be detected in an MRI. The air contours created in the phantoms based on the MR threshold formula agree with the ground truth within 1 mm (Housdorff distance). Part of the remaining disagreement may be due to the error in CT and MRI registration (Oh and Kim 2017).

For patient data, although the ground truth can be obtained from CTs, these images are acquired at a different time from the MRI, resulting in different air cavities between the CT and MRI for abdominal sites. Thus, it was necessary in this study to start with the air thresholding in phantoms. We determined the air threshold on patient MR through an iterative process using data obtained from two different MRI scanners with 5 different MR sequences, making the developed method applicable to commonly used clinical protocols. Establishing a thresholding equation allows us to have a more consistent threshold value based on previous visual judgment of operators. Even for the same MR sequences, the users determined a large variety of threshold values which indicates an operator bias. This is nearly eliminated when the formula is based on many instances where the biases in the individual values are averaged out. Using all the MRI data selected, we were able to establish the a, b, c, and d coefficients in the equation and the resulting values were within the range. We determined by visual inspection that the threshold values within the range were equally accurate, and it was not possible to judge any one value to be more accurate than the others. Therefore, any number within that range can be used as an air threshold.

Even if the correct threshold value is used, many non-air regions also would be identified as air as they have intensities less than the threshold value. To distinguish and remove the non-air regions, we used two criteria: (1) a region determined by eliminating all voxels outside 1 cm vicinity of the possible air containing region (Lambin *et al* 2012), (2) quantitative image texture features were extracted on a voxel by voxel basis. The use of texture analysis and a Bayesian classifier on a voxel-by-voxel basis can reduce the misclassification errors.

Although the model has multiple steps such as AACs with gross air region contouring, CT/MR registration and contour transfer with MACs with threshold, it is efficient since the online component of this process is fully automatic. Therefore, it is fast and can be performed within roughly one minute. The longest component is the deformable image registration operation, which takes less than 30 s via ADMIRE (Elekta). The improvement in efficiency is huge, since the manual contouring is rather time consuming especially for the abdominal regions, where one must go through each slice and delineate each of the air regions, which could be several minutes of online time, which is too slow for an online ART workflow. The reduction of time would be more for the abdominal patients, due to the large number of slices and organs to go through. It is hard to put a specific number since the manual process is hard to quantify and dependent on the operator. In this way, we determine the correct segmentation of air cavities in MRI, and thus make accurate electron density assignments for air regions in MRI.

The established formula was proven to provide a reproducible result in the phantom studies where the air regions can be delineated with a large confidence compared to the patient data. The proposed method seems to be effective in differentiating air from non-air regions. However, it needs to be implemented with caution as statistical uncertainties can affect the accuracy of the delineated regions. For instance, uncertainties can come from registration errors, ground truth inter-observer variability, the acquisition time gap between CT and MRI, etc.

5. Conclusion

We have developed a multi-step process that includes MR intensity thresholding and image texture-based voxel classification to automatically and accurately segment air regions on MRI, allowing synthetic CT to be generated quickly and precisely from daily MRI for MRI guided online adaptive radiation therapy.

Acknowledgments

This work is partially supported by MCW Cancer Center Froedtert Foundation.

Conflicts of interest

This work is included in a patent application.

ORCID iDs

Haidy Nasief https://orcid.org/0000-0002-5591-5094
X Allen Li https://orcid.org/0000-0003-2260-0245

References

Ahunbay E E, Thapa R, Chen X, Paulson E and Li X A 2019 A technique to rapidly generate synthetic CT for MRI-guided online adaptive re-planning: an exploratory study *Int. J. Radiat. Oncol. Biol. Phys.* **103** 1261–70

Amadasun M and King R 1989 Textural features corresponding to textural properties *IEEE Trans. Syst. Man Cybern.* **19** 1264–74

Andreasen D, Van Leemput K and Edmund J M 2016 A patch-based pseudo-CT approach for MRI-only radiotherapy in the pelvis *Med. Phys.* **43** 4742–52

Andreasen D, Van Leemput K, Hansen R H, Andersen J L and Edmund J M 2015 Patch-based generation of a pseudo CT from conventional MRI sequences for MRI-only radiotherapy of the brain *Med. Phys.* **42** 1596–605

Arabi H, Koutsouvelis N, Rouzaud M, Miralbell R and Zaidi H 2016 Atlas-guided generation of pseudo-CT images for MRI-only and hybrid PET–MRI-guided radiotherapy treatment planning *Phys. Med. Biol.* **61** 6531–52

Bol G H, Hissoiny S, Lagendijk J J W and Raaymakers B W 2012 Fast online Monte Carlo-based IMRT planning for the MRI linear accelerator *Phys. Med. Biol.* **57** 1375–85

Burgos N *et al* 2015 Robust CT synthesis for radiotherapy planning: application to the head and neck region *MICCAI 2015. Part II, LNCS* vol 9350 pp 476–84

Castellano G, Bonilha L, Li L M and Cendes F 2004 Texture analysis of medical images *Clin. Radiol.* **59** 1061–9

Chen X, Prior P, Chen G, Schultz C J and Li X A 2016 Technical Note: Dose effects of 1.5 T transverse magnetic field on tissue interfaces in MRI-guided radiotherapy *Med. Phys.* **43** 4797

Delso G *et al* 2014 Anatomic evaluation of 3-dimensional ultrashort-echo-time bone maps for PET/MR attenuation correction *J. Nucl. Med.* **55** 780–5

Dowling J A *et al* 2015 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.* **93** 1144–53

Dowling J A, Lambert J, Parker J, Salvado O, Fripp J, Capp A, Wratten C, Denham J W and Greer P B 2012 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.* **83** E5–E11
Edmund J M and Nyholm T 2017 A review of substitute CT generation for MRI-only radiation therapy *Radiat. Oncol.* **12** 28
Edmund J M, Kjer H M, Van Leemput K, Hansen R H, Andersen J A and Andreasen D 2014 A voxel-based investigation for MRI-only radiotherapy of the brain using ultra short echo times *Phys. Med. Biol.* **59** 7501–19
Ellingson B M, Zaw T, Cloughesy T F, Naeini K M, Lalezari S, Mong S, Lai A, Nghiemphu P L and Pope W B 2012 Comparison between intensity normalization techniques for dynamic susceptibility contrast (DSC)-MRI estimates of cerebral blood volume (CBV) in human gliomas *J. Magn. Reson. Imaging* **35** 1472–7
Gudur M S R, Hara W, Le Q T, Wang L, Xing L and Li R J 2014 A unifying probabilistic Bayesian approach to derive electron density from MRI for radiation therapy treatment planning *Phys. Med. Biol.* **59** 6595–606
Guerreiro F *et al* 2017 Evaluation of a multi-atlas CT synthesis approach for MRI-only radiotherapy treatment planning *Phys. Med.* **35** 7–17
Han X 2017 MR-based synthetic CT generation using a deep convolutional neural network method *Med. Phys.* **44** 1408–19
Haralick R M, Shanmugam K and Dinstein I 1973 Textural features for image classification *IEEE Trans. Syst. Man Cybern.* **3** 610–21
Hellier P 2003 Consistent intensity correction of MR images *Image Processing, 2003 ICIP 2003* vol **1**
Holub O and Ferreira S T 2006 Quantitative histogram analysis of images *Comput. Phys. Commun.* **175** 620–3
Hsu S H, Cao Y, Huang K, Feng M and Balter J M 2013 Investigation of a method for generating synthetic CT models from MRI scans of the head and neck for radiation therapy *Phys. Med. Biol.* **58** 8419–35
Jean-Philippe F, Sweeney E M, Muschelli J, Crainiceanu C M and Shinohara R T 2016 Removing inter-subject technical variability in magnetic resonance imaging studies *Neuroimage.* **132** 198–212
Johansson A, Karlsson M and Nyholm T 2011 CT substitute derived from MRI sequences with ultrashort echo time *Med. Phys.* **38** 2708–14
Kim J, Glide-Hurst C, Doemer A, Wen N, Movsas B and Chetty I J 2015 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.* **91** 39–47
Korhonen J, Kapanen M, Keyrilainen J, Seppala T and Tenhunen M 2014 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.* **41** 011704
Kraus K M, Jäkel O, Niebuhr N I and Pfaffenberger A 2017 Generation of synthetic CT data using patient specific daily MR image data and image registration *Phys. Med. Biol.* **62** 1358–77
Lagendijk J J, Raaymakers B W and van Vulpen M 2014 The magnetic resonance imaging-linac system *Semin. Radiat. Oncol.* **24** 207–9
Lambin P *et al* 2012 Radiomics: extracting more information from medical images using advanced feature analysis *Eur. J. Cancer* **48** 441–6
Oborn B M, Metcalfe P E, Butson M J and Rosenfeld A B 2009 High resolution entry and exit Monte Carlo dose calculations from a linear accelerator 6 MV beam under the influence of transverse magnetic fields *Med. Phys.* **36** 3549–59
Oh S and Kim S 2017 Deformable image registration in radiation therapy *Radiat. Oncol. J.* **35** 101–11
Prior P *et al* 2017 Technical Note: Is bulk electron density assignment appropriate for MRI-only based treatment planning for lung cancer? *Med. Phys.* **44** 3437–43
Raaijmakers A J E, Hårdemark B, Raaymakers B W, Raaijmakers C P J and Lagendijk J J W 2007a Dose optimization for the MRI-accelerator: IMRT in the presence of a magnetic field *Phys. Med. Biol.* **52** 7045–54
Raaijmakers A J E, Raaymakers B W and Lagendijk J J W 2005 Integrating 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.* **50** 1363–76
Raaijmakers A J E, Raaymakers B W, Meer S V D and Lagendijk J J W 2007b Integrating a MRI scanner with a 6 MV radiotherapy accelerator: Impact of the surface orientation on the entrance and exit dose due to the transverse magnetic field *Phys. Med. Biol.* **52** 929–39
Raaymakers B W, Raaijmakers A J E, Kotte A N T J, Jette D and Lagendijk J J W 2004 Integrating a MRI scanner with a 6 MV radiotherapy accelerator: Dose deposition in a transverse magnetic field *Phys. Med. Biol.* **49** 4109–18
Robson M D *et al* 2003 Magnetic resonance: an introduction to ultrashort TE (UTE) imaging *J. Comput. Assist. Tomogr.* **27** 825–46
Shah M, Xiao Y and Subbanna N 2011 Evaluating intensity normalization on MRIs of human brain with multiple sclerosis *Med. Image Anal.* **15** 267–82
Shinohara R T *et al* 2014 *Statistical Normalization Techniques for Magnetic Resonance Imaging* vol 6 (Amsterdam: Elsevier) pp 9–19
Siversson C, Nordstrom F, Nilsson T, Nyholm T, Jonsson J, Gunnlaugsson A and Olsson L E 2015 Technical Note: MRI only prostate radiotherapy planning using the statistical decomposition algorithm *Med. Phys.* **42** 6090–7
Sjolund J, Forsberg D, Andersson M and Knutsson H 2015 Generating patient specific pseudo-CT of the head from MR using atlas-based regression *Phys. Med. Biol.* **60** 825–39
Taha A A and Hanbury A 2014 An efficient algorithm for calculating the exact Hausdorff distance *IEEE Trans. Pattern Anal. Mach. Intell.* **37** 2153–63
Taha A A and Hanbury A 2015 Matrics for evaluating 3D medical image segmentation: analysis, selection, and tool *BMC Med. Imaging* **15** 29
van Heijst T C, den Hartogh M D, Lagendijk J J W, van denBongard H J G D and van Asselen B 2013 MR-guided breast radiotherapy: Feasibility and magnetic-field impact on skin dose *Phys. Med. Biol.* **58** 5917–30
Wang L, Lai H M, Barker G J, Miller D H and Tofts P S 1998 Correction for variations in MRI scanner sensitivity in brain studies with histogram matching *Magn. Reson. Med.* **39** 322–7
White D R, Griffith R V and Wilson I J 1992 *Report 46, Journal of the International Commission on Radiation Units and Measurements* vol os24 p 1
Zou K H *et al* 2004 Statistical validation of image segmentation quality based on a spatial overlap index *Acad. Radiol.* **11** 178–89