

# Journal of Medical Imaging

MedicalImaging.SPIEDigitalLibrary.org

## **MRI-based pseudo CT synthesis using anatomical signature and alternating random forest with iterative refinement model**

Yang Lei  
Jiwoong Jason Jeong  
Tonghe Wang  
Hui-Kuo Shu  
Pretesh Patel  
Sibo Tian  
Tian Liu  
Hyunsuk Shim  
Hui Mao  
Ashesh B. Jani  
Walter J. Curran  
Xiaofeng Yang

**SPIE.**

Yang Lei, Jiwoong Jason Jeong, Tonghe Wang, Hui-Kuo Shu, Pretesh Patel, Sibon Tian, Tian Liu, Hyunsuk Shim, Hui Mao, Ashesh B. Jani, Walter J. Curran, Xiaofeng Yang, "MRI-based pseudo CT synthesis using anatomical signature and alternating random forest with iterative refinement model," *J. Med. Imag.* 5(4), 043504 (2018), doi: 10.1117/1.JMI.5.4.043504.

# MRI-based pseudo CT synthesis using anatomical signature and alternating random forest with iterative refinement model

Yang Lei,<sup>a</sup> Jiwoong Jason Jeong,<sup>a</sup> Tonghe Wang,<sup>a</sup> Hui-Kuo Shu,<sup>a</sup> Pretesh Patel,<sup>a</sup> Sibor Tian,<sup>a</sup> Tian Liu,<sup>a</sup> Hyunsuk Shim,<sup>a,b</sup> Hui Mao,<sup>b</sup> Ashesh B. Jani,<sup>a</sup> Walter J. Curran,<sup>a</sup> and Xiaofeng Yang<sup>a,\*</sup>

<sup>a</sup>Emory University, Winship Cancer Institute, Department of Radiation Oncology, Atlanta, Georgia, United States

<sup>b</sup>Emory University, Winship Cancer Institute, Department of Radiology and Imaging Sciences, Atlanta, Georgia, United States

**Abstract.** We develop a learning-based method to generate patient-specific pseudo computed tomography (CT) from routinely acquired magnetic resonance imaging (MRI) for potential MRI-based radiotherapy treatment planning. The proposed pseudo CT (PCT) synthesis method consists of a training stage and a synthesizing stage. During the training stage, patch-based features are extracted from MRIs. Using a feature selection, the most informative features are identified as an anatomical signature to train a sequence of alternating random forests based on an iterative refinement model. During the synthesizing stage, we feed the anatomical signatures extracted from an MRI into the sequence of well-trained forests for a PCT synthesis. Our PCT was compared with original CT (ground truth) to quantitatively assess the synthesis accuracy. The mean absolute error, peak signal-to-noise ratio, and normalized cross-correlation indices were  $60.87 \pm 15.10$  HU,  $24.63 \pm 1.73$  dB, and  $0.954 \pm 0.013$  for 14 patients' brain data and  $29.86 \pm 10.4$  HU,  $34.18 \pm 3.31$  dB, and  $0.980 \pm 0.025$  for 12 patients' pelvic data, respectively. We have investigated a learning-based approach to synthesize CTs from routine MRIs and demonstrated its feasibility and reliability. The proposed PCT synthesis technique can be a useful tool for MRI-based radiation treatment planning. © 2018 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: 10.1117/1.JMI.5.4.043504]

Keywords: MRI-based treatment planning; pseudo computed tomography; random forest; feature selection.

Paper 18197R received Sep. 4, 2018; accepted for publication Nov. 12, 2018; published online Dec. 5, 2018.

## 1 Introduction

Magnetic resonance imaging (MRI) has several important advantages over computed tomography (CT) for radiation treatment planning. The principle advantage is a more accurate and reliable organ-at-risk and target delineation with superior soft tissue contrast offered by MRI over CT.<sup>1</sup> A potential treatment planning process with MRI as a sole imaging modality<sup>2,3</sup> could eliminate systematic MRI-CT coregistration errors (for example, an inherent uncertainty of  $\sim 2$  mm in cranial MRI-CT registration was reported by Ulin et al.<sup>4</sup>), reduce medical cost, minimize patient radiation exposure, and simplify clinical workflow. Despite these advantages, MRI contains neither unique nor quantitative information on the attenuation properties which are required for accurate dose calculations and the generation of reference images for patient setup based on two-dimensional planar images.<sup>2</sup> Synthesizing PCT images with electron density information from MRIs has been proposed.<sup>5,6</sup> Conventional PCT synthesis techniques can be broadly categorized as atlas-, segmentation-, sequence-based methods, and hybrid methods. Atlas-based methods map a single or multiple CT atlases to a newly acquired MRI to generate the corresponding PCT of the new patient based on deformable registration algorithms.<sup>7,8</sup> However, these methods' efficiency is highly limited by the accuracy of intersubject registration algorithms. Segmentation-based methods first manually or automatically segment MR images into different classes, such as bone, air, fat, and soft tissue, and then assign a uniform electron density

value to each class.<sup>9–11</sup> However, segmentation-based methods cannot reliably distinguish between bone and air regions due to their similar MRI intensities. Sequence-based methods use specialized MR sequences, such as the ultrashort echo time (UTE), or a mixture of standard and specialized sequences to acquire different types of MRIs and perform some postprocessing on these MRIs to generate the final PCT.<sup>12–14</sup> The current image quality of UTE sequences is inadequate for accurate delineation of blood vessels from bone, as they both appear dark.<sup>15,16</sup> Moreover, the use of nonstandard MR sequences may introduce additional scanning time to the existing MR scanning workflow and may increase the probability of discomfort and motion degradation. Hybrid methods combine two or more of the above methods into a mixed framework to generate a PCT.<sup>10,11,17</sup>

Recently, as machine learning has become more and more popular in the medical imaging field, using machine-learning-based methods to generate PCTs from MRIs has been developed and can be classified into the following three categories:

1. Dictionary-learning (DL)-based methods: These methods first rank the corresponding CT patches based on the similarity of MR image patches and then linearly or nonlinearly combine these top-ranked patches together to generate a PCT patch and finally reconstruct a whole PCT through combining all synthetic patches together.<sup>18–22</sup> However, these methods are sensitive to MR intensities which can vary as a function of scanning parameters for a given tissue.

\*Address all correspondence to: Xiaofeng Yang, E-mail: [xyang43@emory.edu](mailto:xyang43@emory.edu)

2. Random-forest (RF)-based methods: RFs constitute an intuitive model that offers a flexible probabilistic framework for solving the learning tasks that can be used for learning MRI-CT mapping.<sup>23-25</sup> However, these methods are imperfect due to uninformative and redundant components of extracted feature vectors, which may introduce ambiguous binary splitting in the decision trees of an RF to result in an erroneous synthesis.
3. Convolutional-neural-network (CNN)-based methods: A CNN provides a complex nonlinear mapping from MRI to CT through a multilayer and fully trainable model.<sup>26-28</sup> However, CNN-based methods suffer from a long training time and their performance can be affected by many parameters.

In this work, we incorporate anatomical signature, alternating random forest (ARF), and iterative refinement (IR) model into a learning-based framework to derive PCTs from standard MRIs for MRI-based treatment planning. The contributions of the paper are as follows: (1) a feature selection (FS) is introduced to identify the most informative components of features, which are served as an anatomical signature. (2) To reduce the uncertainty of both MRI anatomical signature and CT target, a joint information gain (JIG) is incorporated into the decision trees of an RF. (3) To cope with a lack of global optimization in RF-based methods, an ARF scheme is used to train an RF-based model. (4) An IR model is introduced to enhance the accuracy of PCT synthesis by combining the contextual information of a PCT image and MRI anatomical signature.

This paper is organized as follows: we first provide an overview of our proposed MR-based PCT synthesis framework, followed by the details on the FS, the construction of the ARF with a JIG, and the IR model for synthesis improvement. We then compare our methods with DL- and RF-based PCT methods and display our contribution within the experimental results.

## 2 Methods

### 2.1 System Overview

For a given pair of MR and CT training images, the CT image was used as the regression target of the MR image. Prior to the training stage, we performed image preprocessing to improve the training quality by removing noise and uninformative regions by a nonlocal means method.<sup>29</sup> An intrasubject registration was performed to align each pair of MR and CT images, and all pairs were then aligned into a common space. This registration is performed by a commercial software Velocity AI 3.2.1 (Varian Medical Systems, Palo Alto, California) using rigid registration. In the training stage, for each MRI, we first extracted multilevel features, i.e., discrete cosine transform (DCT), local binary pattern (LBP), and pair-wise voxel difference features<sup>23</sup> from multiscale images, which consisted of an original and three derived images with a sequence of downsampling factors (0.75, 0.5, and 0.25). Then, these extracted features were concatenated as a patch-based feature vector. (The patch size of MRI is  $15 \times 15 \times 15$  voxel size.) Next, we performed an FS process through least absolute shrinkage and selection operator to identify the anatomical signature from these extracted features. The anatomical signature of each voxel,

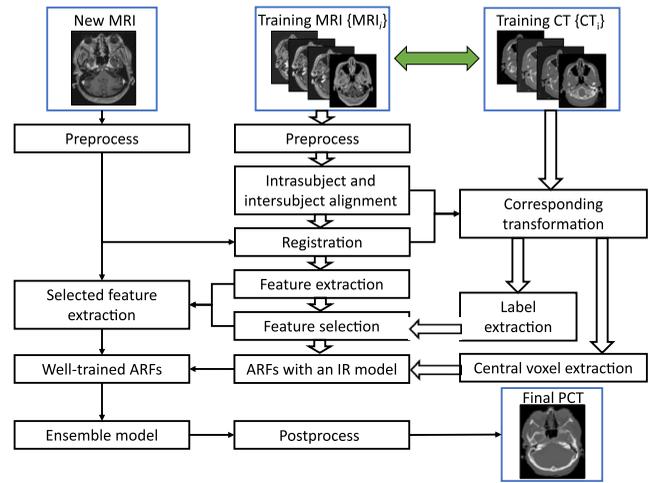


Fig. 1 The flowchart of the proposed method.

together with the corresponding CT target, was used to train a sequence of ARFs under an IR model. In the synthesizing stage, the selected features (anatomical signature) were extracted from a new MR image and then were fed into the sequence of well-trained ARFs for the PCT synthesis. Figure 1 outlines the workflow schematic of our synthesis method.

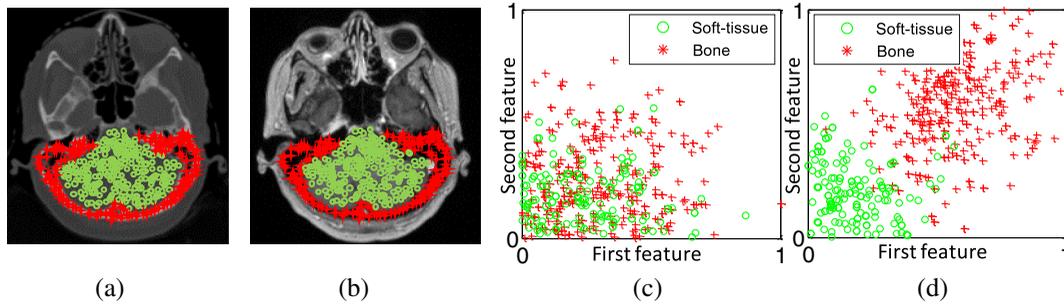
### 2.2 Feature Selection

The patch-based features may contain noisy and redundant features that could overfit a learning model, affecting the prediction accuracy and degenerate the final PCT synthesis performance. Therefore, an FS should be performed to identify the most informative features, which are called as anatomical signature to differentiate with original feature vector. As recommended in Ref. 22, minimizing a logistic sparsity regression is used for our FS, as follows:

$$w = \min_w \left\{ \sum_{j=1}^n \log(1 + \exp(-l(y_j)(w^T f(x_j) + b))) + \mu \sum_{i=1}^m |w_i| / \beta_i \right\}. \quad (1)$$

For each MRI patch centered at voxel  $x_j$ ,  $f(x_j)$  denotes the extracted feature vector which concatenates DCT, LBP, and pair-wise voxel difference features at multiple scales,  $y_j$  denotes the CT voxel at the same position of  $x_j$ ,  $n$  denotes the number of voxel samples  $x_j$ ,  $\beta_i$  denotes the optimization scalar, which is computed by discriminative power, i.e., Fisher's score,<sup>30</sup>  $\mu$  is a regularization parameter and we set  $\mu = 0.5$ ,  $b$  is an intercept scalar,  $l(\cdot)$  denotes an anatomical label, and  $w = \{w_i\}$  is a sparse coefficient vector. Each component  $w_i$  denotes the  $i$ 'th component's importance of  $f(x_j)$  and  $m$  denotes the length of  $w$ .

We first used label  $l(y_j) = 1$  if  $y_j$  belongs to nonair regions and  $l(y_j) = -1$  if  $y_j$  belongs to air regions. After first FS, we used  $l(y_j) = 1$  if  $y_j$  belongs to bone regions and  $l(y_j) = -1$  if  $y_j$  belongs to soft tissue regions. The corresponding label was obtained by setting the CT intensity range of  $[-\infty, -400)$  as air,  $[-400, 300)$  as soft tissue, and  $[300, +\infty)$  as bone. As studied in our previous work,<sup>22</sup> the  $w$  is obtained by Eq. (1). The anatomical signature, denoted as  $fs(x_j)$ , is composed by a subset of feature components with corresponding nonzero entries in  $w$ . The total dimensionality of the original feature space is 2400.



**Fig. 2** An example of discriminating the bone from soft tissue regions with and without FS. (a) and (b) Axial MRI and CT images, where the sample voxels belonging to soft tissue are highlighted by green circles, and sample voxels belonging to bone are highlighted by red asterisks. (c) Scatter plots of random 2 features of original extracted features generated from MRI patches that centered on corresponding samples. (d) Scatter plots of 2 top-ranked features with a high Fisher's separation score.

After FS, 63 selected informative features are adequate. These selected components have superior power to distinguish among bone, air, and soft tissue. An example is given in Fig. 2, (a) and (b) show the sample voxels drawn from a pair of MR/CT images. A sample belonging to the bone is highlighted by a red asterisk, and a sample belonging to the soft tissue is represented by a green circle. Figure 2(c) shows a scatter plot of two random features of the corresponding samples without any FS. Figure 2(d) shows a scatter plot of the two top-ranked features with a high Fisher's separation score.<sup>30</sup> It can be observed that the selected features can better separate bone and soft tissue voxels in an MRI.

## 2.3 Alternating Regression Forest

### 2.3.1 Joint information gain

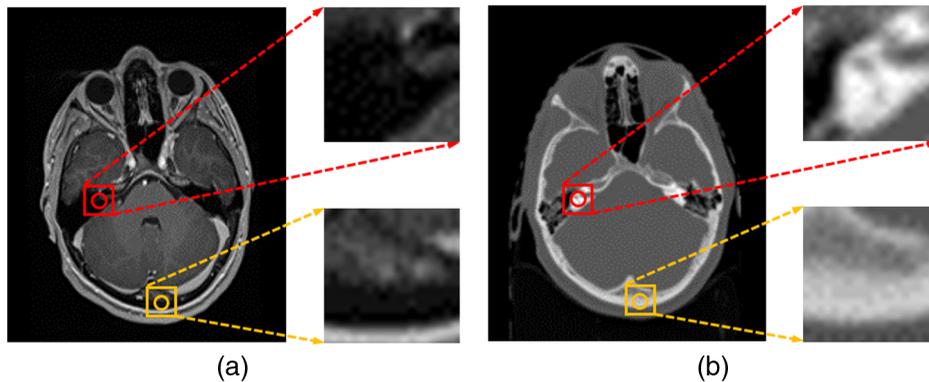
An RF includes a series of decision trees, each of which is fed with a random subset of training data and is trained independently from the others. Each decision tree recursively splits a training sample  $S = \{f_s(x_j), y_j\}$  from a parent node into two disjoint child nodes, such that the uncertainty of the CT target variables in the resulting subsets (child node) is minimized. However, if the splitting of decision trees is only based on CT target variables, two CT patches with a similar CT intensity in their central voxels may come from two totally different locations with significantly different MRI anatomical signature and vice versa. In Figs. 3(a) and 3(b), the red and yellow circles

denote the central voxel of the patches of MRI and CT, respectively. In Fig. 3(a), these two MRI patches from two different locations have totally different extracted features and anatomical signatures, while in Fig. 3(b) the central voxel values of the two corresponding CT patches have the similar intensity values. Based on similar CT intensities, these two samples with different MRI anatomical signatures will be distributed in a uniform partition (child node), no matter which feature in anatomical signature is chosen and which splitting function is applied for a separation. Thus, the splitting procedure pertaining to the uncertainty reduction of CT targets may induce ambiguity for training a decision tree.

To cope with this issue, we proposed a JIG strategy to reduce the uncertainty between an MRI anatomical signature and a CT target:

$$H_C(S, \theta) = \sum_i |y_i - \bar{y}|^2 / |S| + \lambda \sum_i |f_s(x_i) - m(f_s(x_i))|^2 / |S|, \quad (2)$$

where  $\lambda$  is a smoothing parameter,  $m(\cdot)$  denotes the mean of anatomical signatures,  $\bar{y}$  denotes the mean of CT targets, and  $|S|$  is the number of samples in  $S$ . Here, we empirically set  $\lambda = 0.05$ . The  $i$  in Eq. (2) means the  $i$ 'th position of central location in both CT and MR image. In first term,  $y_i$  denotes the CT patch whose central location is  $i$ . In second term,  $f_s(x_i)$  denotes



**Fig. 3** An example of (a) a CT and (b) its corresponding MR image with highlighted patches and center voxels.

the anatomic feature vector of MR patch  $x_i$  whose central location is  $i$ . By minimizing Eq. (2), the binary splitting procedure not only requires forwarding the training anatomical signatures with similar CT targets into the same child node but also satisfying the restraint that the anatomical signatures themselves are similar, which potentially eases the inference model in a leaf. The splitting procedure continues until a stopping criterion is met and then final leaf nodes are created.

### 2.3.2 Alternating random forest

Since how data are further split is only decided at a local node level for an RF-based method, the training procedure is not globally supervised by an appropriate measurement from its whole regressor's performance. To cope with this issue, we introduced an alternating RF concept to provide a global optimization.<sup>31</sup>

The global loss function is computed as a breadth-first missing error that is measured by all weak inference learners (regressor) at each depth.<sup>31</sup>

$$\sum_{\{x_i, f_s(x_i)\}} L(y_i, R_{d-1}(x_i, \theta) + r_d(x_i, \theta_d)), \quad (3)$$

where  $L(\cdot)$  is a differentiable loss function and  $R_{d-1}(x_i, \theta) = \sum_{j=1}^{d-1} r_j(x_i, \theta_j)$  denotes a regressor, which is trained from a root node to the child nodes at the  $(d-1)$ 'th depth,  $\theta$  is a collection of the thresholds,  $r_d(x_i, \theta_d)$  denotes a regressor for the  $d$ 'th depth, and  $\theta_d$  is a set of splitting thresholds to be trained in the  $d$ 'th depth by minimizing Eq. (2). Then, the splitting function was optimized by maximizing the JIG and minimizing the global loss at each depth:

$$\arg \max_{\theta_d} (H_C(S, \theta_d) - \gamma \sum_{\{x_i, f_s(x_i)\}} L(y_i, R_{d-1}(x_i, \theta) + r_d(x_i, \theta_d))). \quad (4)$$

The term  $H_C(S, \theta_d)$  is the JIG in Eq. (2), the second term of Eq. (4) is the global loss regularization in Eq. (3), and  $\gamma$  is a parameter balancing the need for a regression. Here, we set  $\gamma = 0.05$ . Figure 4 briefly shows the brief architecture of an ARF.

## 2.4 Iterative Refinement Model

It has been known that the context, i.e., the surrounding information with respect to an object of interest, plays a vital role in interpreting image content.<sup>32</sup> Similarly, the synthesis of a PCT could be enhanced by the information from its surrounding neighbors. An IR model was used to iteratively leverage the surrounding information by utilizing an autocontext method (ACM) to improve the synthesis performance.<sup>32</sup> An IR model is used to characterize the context information from the previous synthesis results and then uses such information as context features,<sup>32</sup> together with the anatomical signature extracted from the input MR image, to recursively train a series of ARFs. The process is then repeated to train a series of ARFs until a synthesis error criterion is met. In the synthesizing stage, a new MR image is fed into these well-trained ARFs with the IR model to iteratively generate and refine the PCT. The architecture of an IR model is shown in Fig. 5.

## 3 Experiments

### 3.1 Datasets

To test the proposed method, we applied it to 14 patients' brain data and 12 patients' pelvic data. For the brain images, T1-weighted MRI was captured using a GE MRI scanner with magnetization-prepared rapid gradient echo (MP-PAGE) sequence and  $1.0 \times 1.0 \times 1.4 \text{ mm}^3$  voxel size (TR/TE: 950/13 ms, flip angle: 90 deg) and CT was captured with a Siemens CT scanner with  $1.0 \times 1.0 \times 1.0 \text{ mm}^3$  voxel size. For the pelvic images, MRI was acquired using a Siemens standard T2-weighted MRI scanner with three-dimensional T2-SPACE sequence and  $1.0 \times 1.0 \times 2.0 \text{ mm}^3$  voxel size (TR/TE: 1200/123 ms, flip angle: 95 deg) and CT was captured with a Siemens CT scanner with  $1.0 \times 1.0 \times 1.2 \text{ mm}^3$  voxel size. We performed the leave-one-out cross-validation method to evaluate the proposed method. Our PCTs were compared with the original planning CT images which served as ground truth. Three widely used metrics, mean absolute error (MAE), peak signal-to-noise ratio (PSNR), and normalized cross correlation (NCC), were used to quantify the absolute difference, relative difference, and image similarity within the body outline. They are defined as

$$\text{MAE} = |I_{\text{CT}} - I_{\text{PCT}}|/C, \quad (5)$$

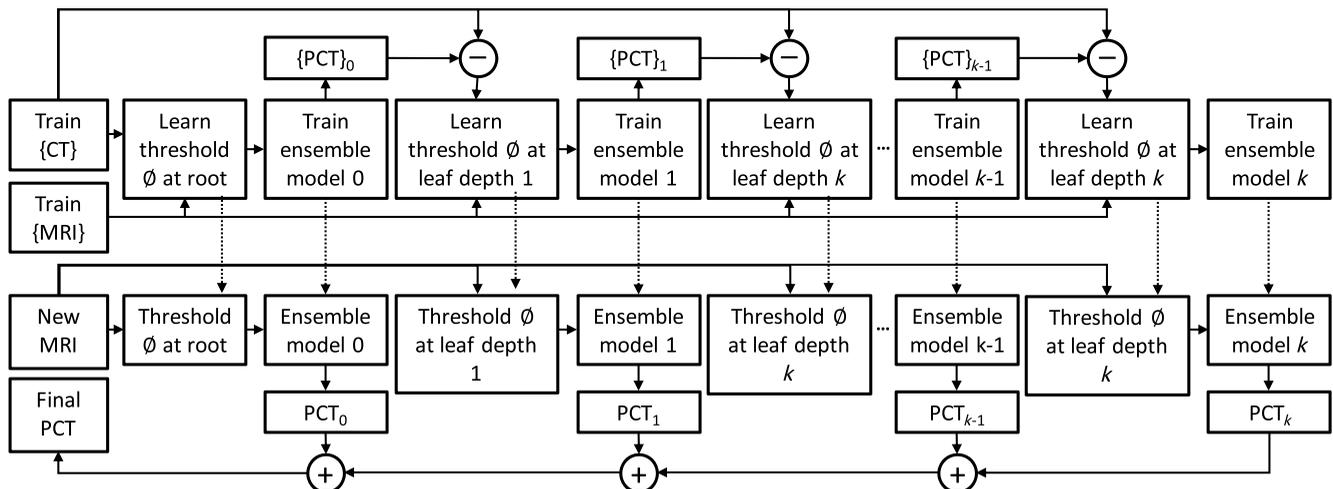


Fig. 4 The architecture of an alternating regression forest.

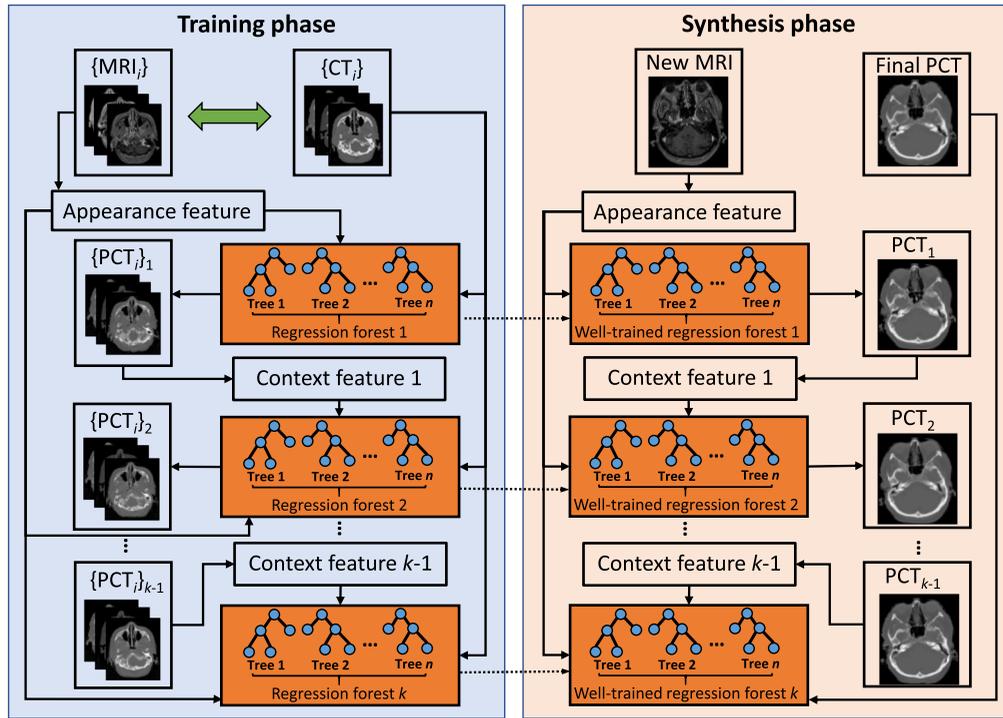


Fig. 5 The architecture of the proposed IR model.

$$PSNR = 10 \log_{10}(C \cdot Q^2 / \|I_{CT} - I_{PCT}\|_2^2), \quad (6)$$

$$NCC = \sum_{x,y,z} C \cdot (I_{CT}(x, y, z) - \mu_{CT})(I_{PCT}(x, y, z) - \mu_{PCT}) / \sigma_{CT} \sigma_{PCT}, \quad (7)$$

where  $I_{CT}$  is the ground truth CT,  $I_{PCT}$  is the corresponding PCT,  $Q$  is the maximum intensity value of  $I_{CT}$  and  $I_{PCT}$ .  $C$  is the number of image voxels within the body outline,  $\mu_{CT}$  and  $\mu_{PCT}$  are the mean of CT and PCT intensities' values, respectively, and  $\sigma_{CT}$  and  $\sigma_{PCT}$  are the standard deviations of CT and PCT intensities' values, respectively. Generally, better synthesis results are associated with lower MAE, higher PSNR, and higher NCC values.

### 3.2 Parameter Setting

Generally, the maximum number of training data in a leaf node is set to 5 for regression tasks, as recommended in the guidance of a regression forest.<sup>23</sup> In theory, the synthesis performance can be improved with more decision trees and a deeper depth for each decision tree. However, we need to balance the trade-off between the computation time and the regression performance. To achieve the optimal balance between the two, we conducted a fourfold cross validation to decide the number of trees and the maximum depth of each tree, as shown in Fig. 6. Figures 6(a1)–6(a3) show the changes of the mean MAE, PSNR, and NCC metrics with varying numbers of trees while fixing the other parameters to its default values as shown in Table 1. Based

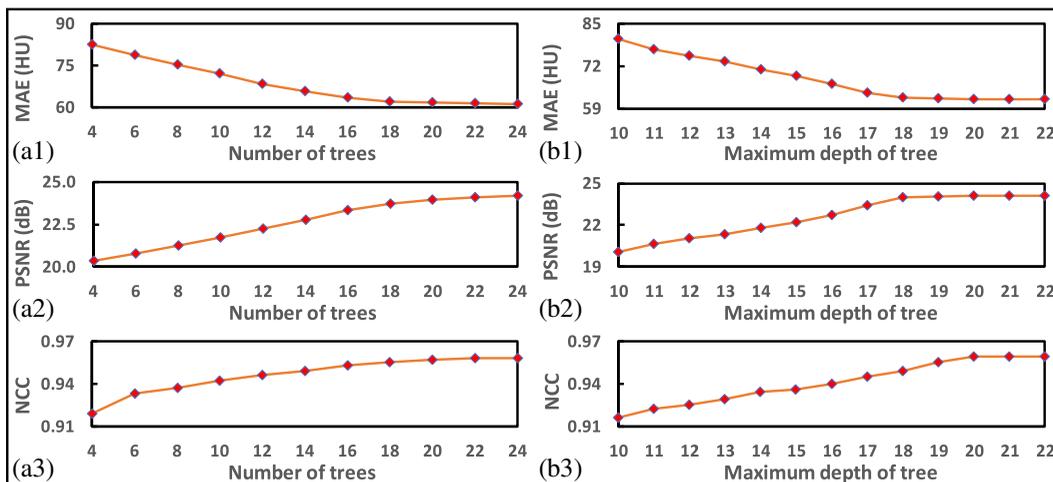


Fig. 6 (a1)–(a3) MAE, PSNR, and NCC performance on brain data with varying number of trees, respectively, and (b1)–(b3) MAE, PSNR, and NCC metrics as a function of the maximum depths of trees, respectively.

on these curves, the tree number of 20 is sufficient for a good PCT synthesis. Figures 6(b1)–6(b3) show the changes of the MAE, PSNR, and NCC metrics with different maximum depth of trees while fixing the other parameters to the default as shown in Table 1. A maximum depth of 20 is adequate for a good PCT synthesis.

### 3.3 Influence of a Feature Selection

Figure 7 provides a detailed visual comparison between the proposed RF method with FS (RF + FS) and without FS (RF) on the brain dataset. To demonstrate the detailed difference with and

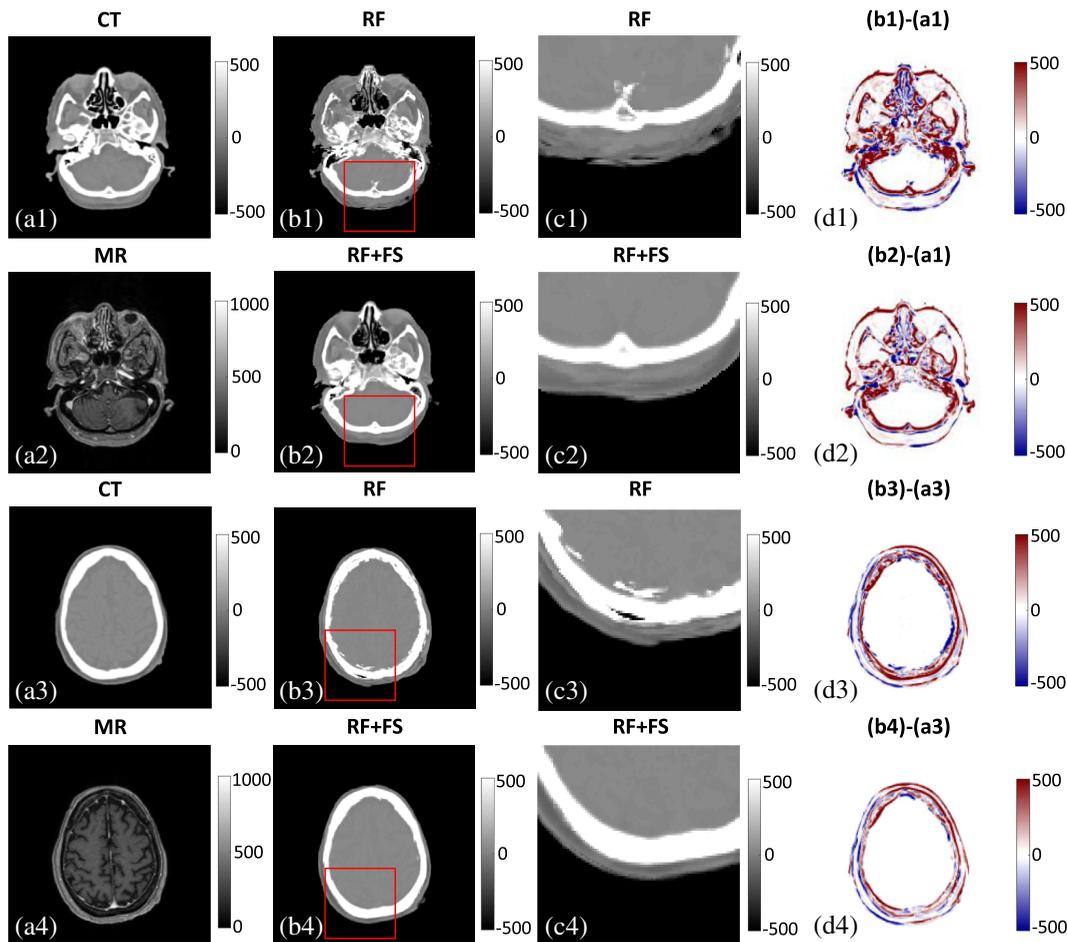
without an FS, two regions of interest (ROIs) inside red boxes were zoomed in Figs. 7(c1)–7(c4). As illustrated in Fig. 7, the RF + FS can better preserve the continuity, coalition, and smoothness in the synthesis results as it preserves informative features in training the model. In Table 3, the comparison between RF and RF + FS quantitatively describes the averaged MAE, PSNR, and NCC results of RF with and without an FS based on leave-one-out experiments. Compared with the RF without an FS, the RF with an FS could significantly reduce the MAE ( $p = 0.005$ ).

### 3.4 Influence of the Proposed Method Based on a JIG and ARF

To evaluate the influence of the proposed JIG and ARF, all the experiments were performed with the same training and test samples images in our brain dataset. Figure 8 provides a comparison among the traditional RF-based method, the RF-based method with a JIG (RF + JIG), and the ARF-based method with a JIG (ARF + JIG). Again, Figs. 8(c1)–8(c6) provide the detailed comparison of two ROIs in Figs. 8(b1)–8(b6) among these three methods. In Table 3, the comparison between RF, RF + JIG, and ARF + JIG quantitatively describes the averaged MAE, PSNR, and NCC results of these three methods based on

**Table 1** Default RF parameter setting.

Parameter	Default value
Number of trees	20
Maximum number of training data in a leaf node	5
Maximum depth of tree	20



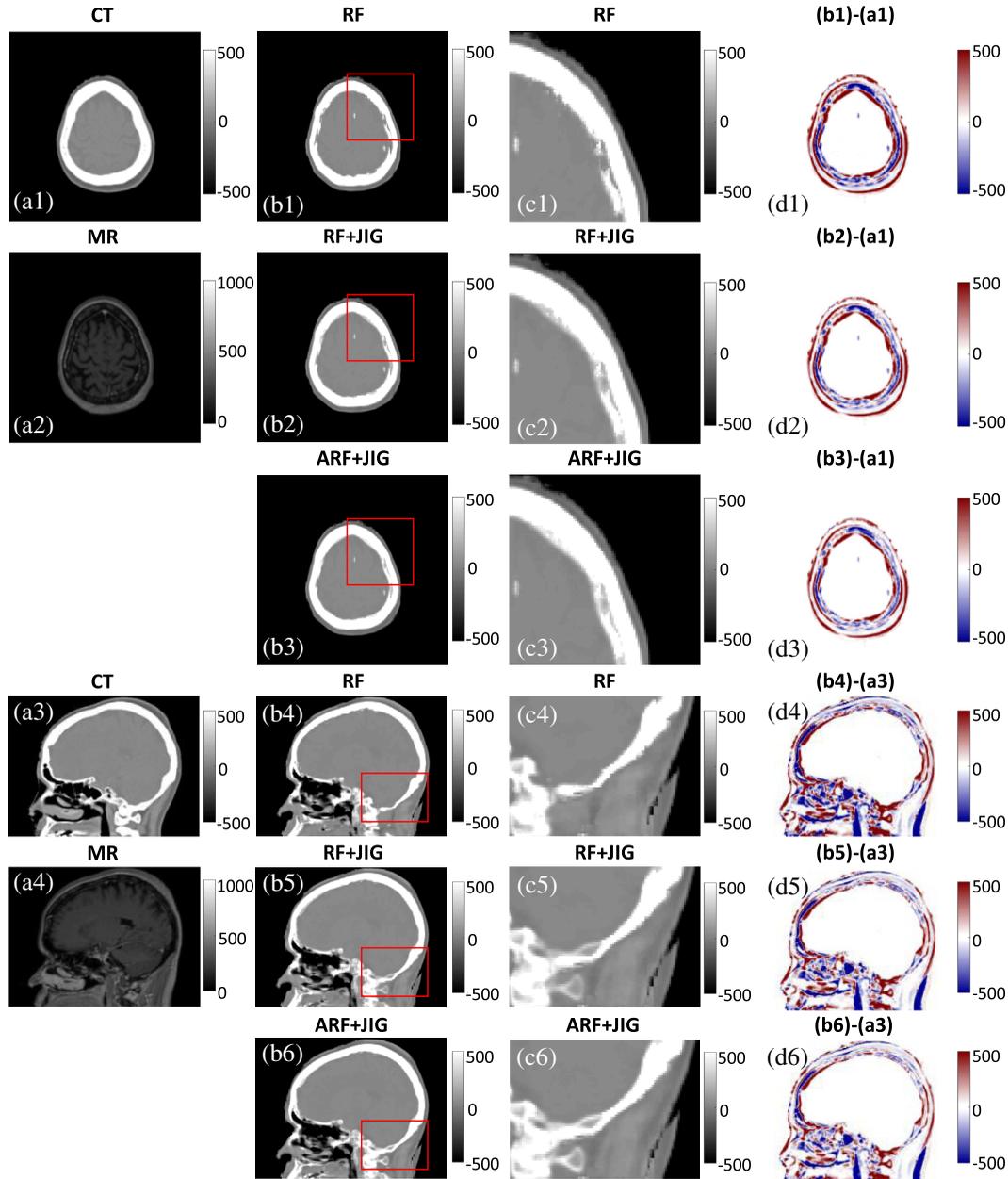
**Fig. 7** Comparison with and without FS: (a1) and (a3) are the axial view of original CT scans, (a2) and (a4) are the corresponding MRI scans, (b1) and (b3) are the PCTs by RF, (b2) and (b4) are the PCTs by RF + FS, (c1)–(c4) are the close-up of the highlighted regions in (b1)–(b4), (d1) and (d3) are the difference images between the original CT and the PCTs by RF, and (d2) and (d4) are the difference images between the original CT and the PCTs by RF + FS.

leave-one-out experiments. A smaller MAE, higher PSNR, and NCC demonstrate the consistent improvement of the ARF + JIG-based method over the RF-based method as well as the RF + JIG-based method. The significant difference ( $p < 0.001$  via  $t$ -test) in the MAE, PSNR, and NCC values between the RF + JIG-based and the RF-based methods demonstrates that the JIG could significantly improve the synthesis performance of an RF-based method. There is also a significant improvement of the NCC between the ARF + JIG-based and RF + JIG-based methods ( $p < 0.001$  via  $t$ -test). Since the joint information from both MRI and CT, as well as a breadth-wise global loss,

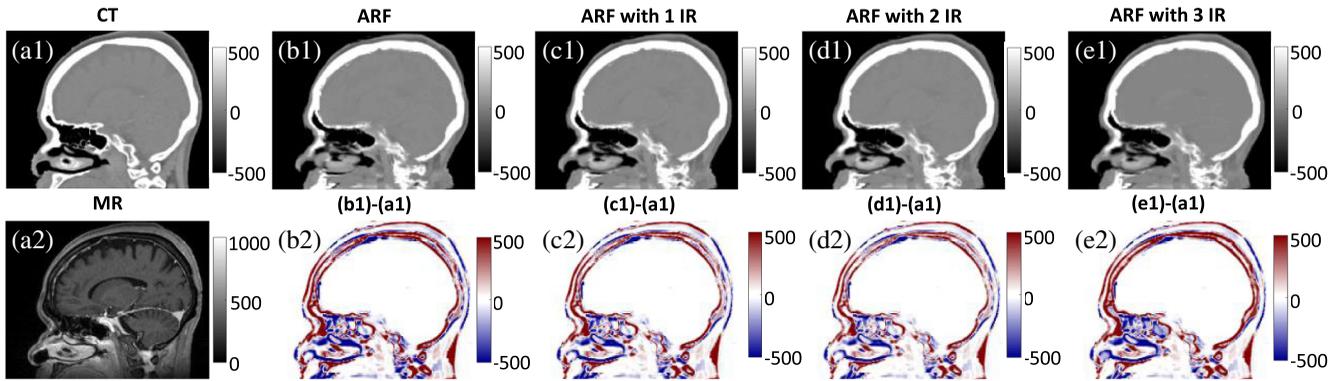
were introduced into our proposed ARF + JIG-based method, the proposed method significantly shows a synthesis performance improvement over the RF-based method.

### 3.5 Influence of an Iterative Refinement Model

The influence of the IR model was demonstrated by comparing the generated PCTs with different refinement iterations. Figure 9 shows the PCT results at different refinement iterations in the sagittal plane. Table 2 quantitatively shows the MAE, PSNR, and NCC at different refinement iterations.



**Fig. 8** Comparison of RF, RF + JIG, and ARF + JIG: (a1) and (a3) show the axial view of CT scan; (a2) and (a4) are the corresponding MRI scans; (b1) and (b4) are the PCTs by RF; (b2) and (b5) are the PCTs by RF + JIG; (b3) and (b6) are the PCTs by ARF + JIG; (c1)–(c6) show the highlighted region in greater detail; (b1)–(b6), (d1), and (d4) are the difference images between the original CT and the PCTs by RF; (d2) and (d5) are the difference images between the original CT and the PCTs by RF + JIG; and (d3) and (d6) are the difference images between the original CT and the PCTs by ARF + JIG.



**Fig. 9** Comparison of the results at different iterations of the IR model: (a1) shows the axial view of CT scan; (a2) is the corresponding MRI scan; (b1)–(e1) are the PCTs of initial ARF, ARF with 1 IR, ARF with 2 IR, and ARF with 3 IR, respectively; and (b2)–(e2) are the corresponding difference images between the original CT and the PCT in (b1)–(e1).

**Table 2** Performance at different iterations of the IR model on the brain dataset.

Iteration	1	2	3	4
MAE	$63.185 \pm 6.734$	$62.415 \pm 6.494$	$61.735 \pm 6.029$	$60.495 \pm 5.910$
PSNR	$24.048 \pm 1.901$	$24.444 \pm 1.852$	$24.802 \pm 1.880$	$24.923 \pm 1.903$
NCC	$0.948 \pm 0.016$	$0.949 \pm 0.017$	$0.951 \pm 0.017$	$0.951 \pm 0.017$

### 3.6 Comparison of State-of-the-Art Methods

To further evaluate our proposed PCT generation method, we compared our ARF + FS + JIG + IR-based methods with a state-of-the-art DL-based approach,<sup>21</sup> an RF-based approach (RF + ACM),<sup>23</sup> and a CNN-based approach, which is called the generative adversarial networks (GAN) method.<sup>28</sup> The parameters of all the algorithms were set according to their best performance. The comparison results of these three methods are shown in Fig. 10. Due to linear averaging of similar patches, the DL-based method often results in a loss in image quality and erroneous synthesis as shown in Figs. 10(b1), 10(b3), and 10(b5). Due to the uninformative patch-based features, the RF + ACM-based method has some errors around the bone boundary as shown in Figs. 10(c1), 10(c3), and 10(c5). Compared with the results from the DL- and the RF + ACM-based methods, the PCTs generated by our ARF + FS + JIG + IR-based method are much closer to the ground truth CTs as shown in Figs. 10(d1), 10(d3), and 10(d5). Table 3 quantitatively shows the averaged MAE, PSNR, and NCC results of these three methods. Our method obtained an average MAE of  $60.87 \pm 15.10$  HU, a significant improvement over an MAE of  $94.64 \pm 23.91$  HU using the DL-based method,  $77.86 \pm 21.53$  HU using the RF + ACM-based method ( $p < 0.001$  via  $t$ -test), and  $64.86 \pm 19.92$  HU using the GAN-based method ( $p = 0.058$  via  $t$ -test). Our method also resulted in an average PSNR of  $24.63 \pm 1.73$  dB versus  $22.38 \pm 1.86$  dB,  $22.91 \pm 1.60$  dB, and  $23.21 \pm 1.65$  dB ( $p < 0.01$  via  $t$ -test), and an average NCC of  $0.954 \pm 0.013$  versus  $0.912 \pm 0.035$ ,  $0.937 \pm 0.021$ , and  $0.938 \pm 0.018$  obtained by the DL-, the RF + ACM-, and the GAN-based methods ( $p < 0.01$  via  $t$ -test), respectively. Since the bone has the

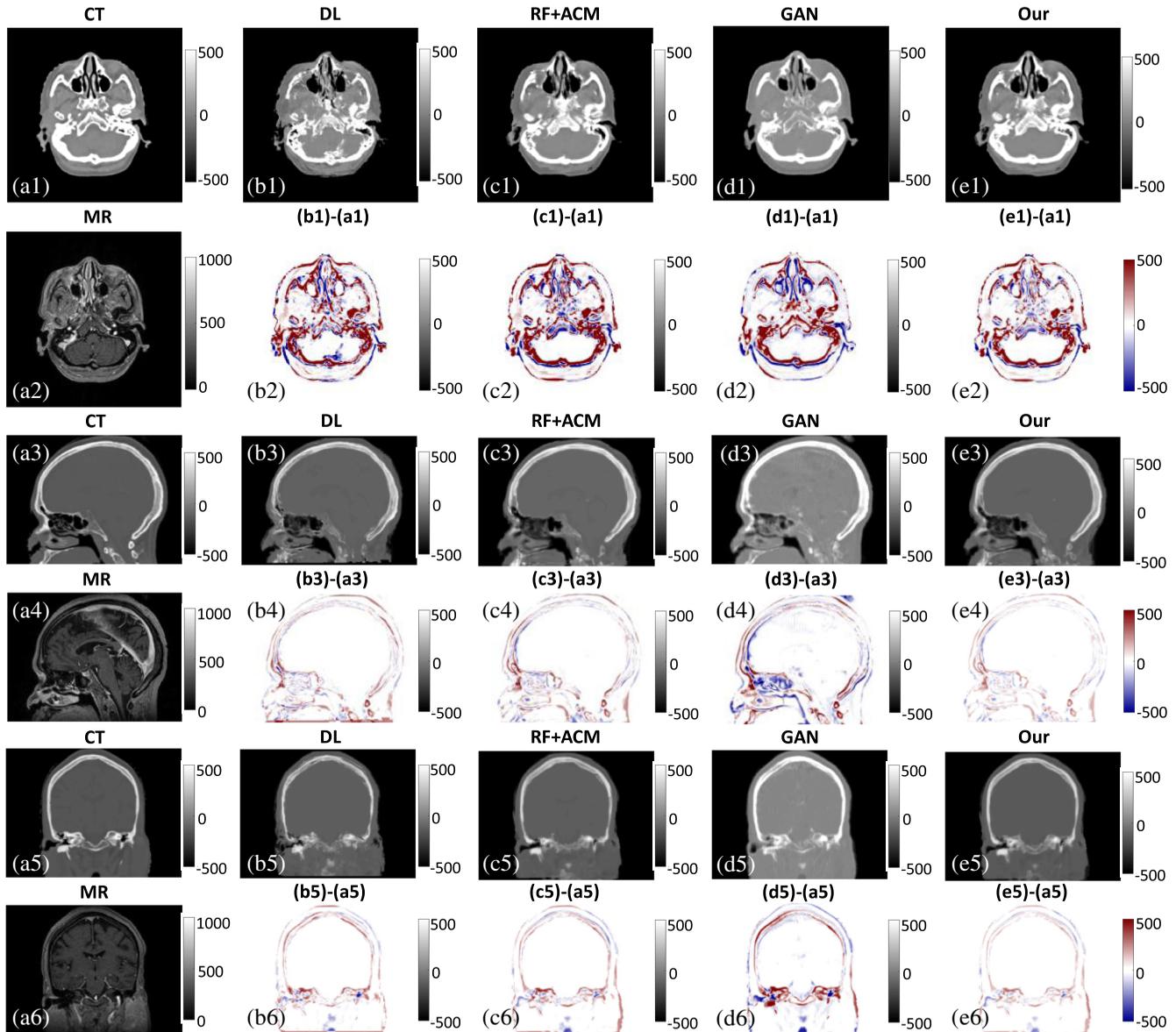
most significant effect on the PCT estimation, we specifically showed the bone ( $\geq 300$  HU) comparison of generated PCTs in Fig. 11. Table 4 shows the averaged MAE in the three different regions (bone, soft tissue, and air). Again, our method based on ARF + FS + JIG + IR significantly reduces the MAE of the bone, soft tissue, and air regions compared with the DL- and RF + ACM-based methods ( $p < 0.01$ ,  $t$ -test).

### 3.7 Results of the Pelvic Data

To further test our proposed method, we applied our proposed method to the pelvic data. Figure 12 shows visual comparison results between PCT and CT. Figures 13(a2)–13(c2) show the quantitative results of the MAC, PSNR, and NCC for each patient's pelvic images. Overall, the average MAE was  $29.86 \pm 10.4$  HU, the average PSNR was  $34.18 \pm 3.31$  dB, and the average NCC was  $0.980 \pm 0.025$  within the body outline.

## 4 Discussion

We presented a machine-learning-based approach to synthesize PCT images from routine MR images. To the best of our knowledge, this is the first work that incorporates the FS and JIG into the RF-based learning framework for this PCT synthesis task. Several mechanisms were also incorporated to improve the synthesis performance, such as the IR strategy and the alternating RF scheme. Our method significantly outperformed the state-of-the-art DL- and RF-based methods for PCT synthesis. Recently, Nie et al.<sup>28</sup> proposed to use the GAN to generate PCT and reported that the MAE was  $92.5 \pm 13.9$  HU for the brain data and  $39.0 \pm 4.6$  HU for the pelvic data. The MAE from our proposed method was  $60.9 \pm 15.1$  HU for the brain data



**Fig. 10** Comparison of the results from different methods: (a1), (a3), and (a5) show the axial, sagittal, and coronal view of CT scan, respectively; (a2), (a4), and (a6) are the corresponding MRI scans; (b1), (b3), and (b5) are the PCTs by DL; (c1), (c3), and (c5) are the PCTs by RF + ACM; (d1), (d3), and (d5) are the PCTs by GAN; (e1), (e3), and (e5) are the PCTs by our final algorithm, i.e., ARF + FS + JIG + IR; (b2), (b4), and (b6) are the difference images between the original CT and the PCTs by DL; (c2), (c4), and (c6) are the difference images between the original CT and the PCTs by RF + ACM; (d2), (d4), and (d6) are the difference images between the original CT and the PCTs by GAN; and (e2), (e4), and (e6) are the difference images between the original CT and the PCTs by our final algorithm.

and  $29.9 \pm 10.4$  HU for the pelvic data. Compared with the deep-learning-based method proposed by Nie et al., our method obtained a much smaller MAE for both sites, which further demonstrates the performance of our PCT synthesis method.

Although the proposed JIG and ARF can enhance the performance, optimizing the proposed binary splitting procedure is more complex than binary splitting in a traditional RF-based method and takes more computation time (about 7 h versus 2 h in the RF-based method by Huynh et al.<sup>23</sup> and 19 h in the GAN-based method by Nie et al.<sup>28</sup>). Although we have tested our algorithm on 14 patients' brain data and 12 patients' pelvic data, in the future, we plan to enroll more patients to further ascertain the robustness of our algorithm. In addition, in this

study, we only used routine T1-weighted (brain) or T2-weighted (pelvis) MRIs to synthesize our PCTs. However, it is possible for our method to generate PCTs using other sequence-based MRIs if our training database includes these images. We plan to combine several types of MRIs based on widely used sequences into our training database to generate PCTs using multisequence MRIs to test whether this could improve our accuracy of PCT.

The accuracy of MRI-CT registration may affect our proposed method from two folds: (1) the mismatches between MRI and CT may affect our model training and (2) the mismatches between MRI and CT will affect our MAE, PSNR, and NCC accuracy. To deal with such small mismatch effects

**Table 3** Numerical results of the different methods on the brain dataset.

Method	MAE (HU)	PSNR (dB)	NCC
DL	94.64 ± 23.91	22.38 ± 1.86	0.912 ± 0.035
RF	82.62 ± 26.06	22.28 ± 1.92	0.927 ± 0.022
RF + FS	75.82 ± 19.69	23.08 ± 1.58	0.936 ± 0.019
RF + JIG	66.43 ± 18.18	24.09 ± 1.62	0.946 ± 0.015
ARF + JIG	65.33 ± 16.59	24.05 ± 1.91	0.948 ± 0.015
RF + ACM	77.86 ± 21.53	22.91 ± 1.60	0.937 ± 0.021
GAN	64.86 ± 19.92	23.21 ± 1.65	0.938 ± 0.018
ARF + FS + JIG + IR	60.87 ± 15.10	24.63 ± 1.73	0.954 ± 0.013

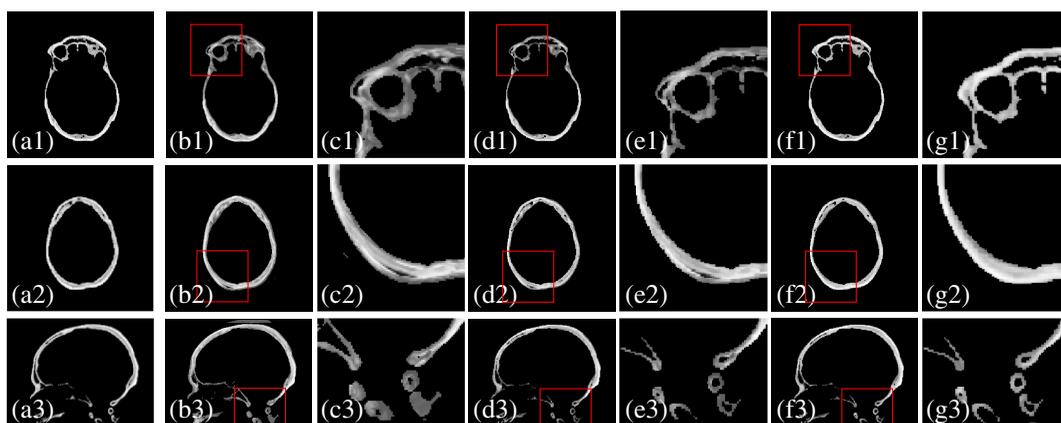
for our model training, we divided the whole image into multiple small patches with dense overlap. Most of the patch pairs between MRI and CT match very well. Moreover, not all mismatched patch pairs are located in the similar anatomical regions for each patient, which do not significantly affect our model training. In the synthesizing stage, for a new MRI feature arriving at each leaf node followed by the previous splitting rules, to infer its correspondence to the CT target, an ensemble model is introduced to generate the final estimation by computing the median of all the nodes within each decision tree, and then computing the median of all the decision trees. Thus, even if there are some positioning differences in pair-wise MRI and CT patches, the median estimation can avoid most of the potential bias. However, our PCT is directly generated from MRI and has same anatomical structures as the MRI, so the mismatches will directly affect our MAE evaluation, which calculates the voxel-based HU difference through a subtract processing between planning CT and PCT.

The MR imaging acquisition parameters as well as magnetic field inhomogeneity and patient-specific distortion may

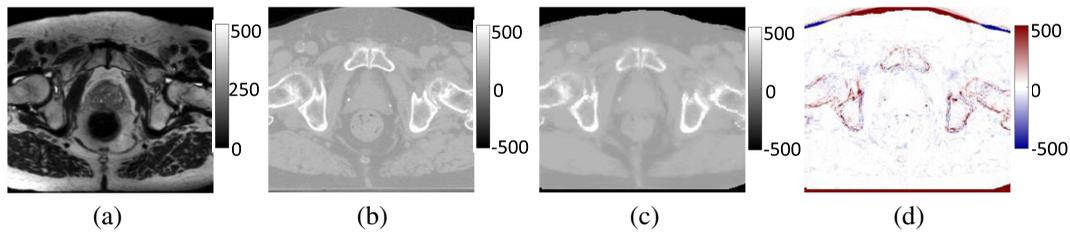
**Table 4** Numerical results of the three methods on the three different regions of brain CT images.

Method	MAE (HU)		
	Bone region	Soft tissue region	Air region
DL	492.51 ± 143.15	52.48 ± 16.17	23.20 ± 7.42
RF + ACM	369.14 ± 87.28	43.33 ± 12.90	22.39 ± 7.46
ARF + FS + JIG + IR	323.83 ± 76.33	38.74 ± 11.28	21.98 ± 7.48

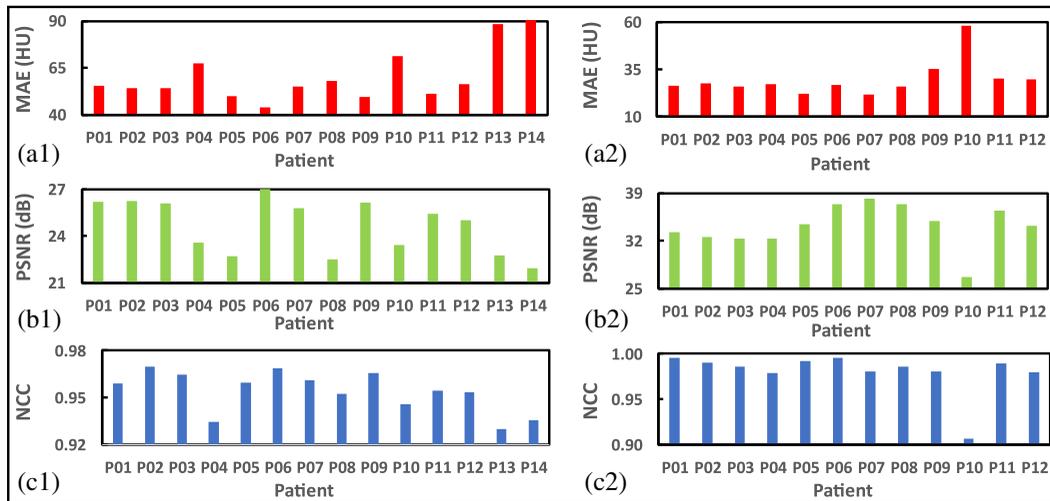
influence the performance of the proposed method, with implications on dosimetry calculations and patient setup. In our study, all MRI images were preprocessed using an N3 algorithm<sup>33</sup> to effectively reduce distortion before training or synthesizing. Other innovative methods, such as a real-time image distortion correction method,<sup>34</sup> have been reported to have excellent performance, and combining these preprocessing methods with our method could increase the accuracy of the PCTs. Due to the high fraction of air, large motion, and distortion, the lung with low resolution and intensity in MRI is very unique and difficult site for our MRI-based radiotherapy, which will be one of our future plans. In this study, we demonstrated the accuracy of PCT in HU numbers. Accurate HU numbers in PCT is of great importance for its usage for dose calculation in MRI-only radiation therapy treatment planning.<sup>35</sup> For example, dose calculation is sensitive around tissue and bone boundaries due to the significant changes in electron density. In MR images, bony tissues pose a significant susceptibility artifact, and its ambiguous boundary with air may introduce synthesizing error of perturbation and shift. Such effect on dose calculation accuracy especially for surrounding tissues needs more related studies. In future, we need to conduct studies to investigate how much those factors on MR imaging will affect the dose calculation and patient setup during our MRI-only-based cancer radiotherapy. These studies would be important to further evaluate the clinical utility of our method.



**Fig. 11** Comparison of the results in the bone region: (a1)–(a3) are the bone from a ground truth CT, (b1)–(b3) are the bone results of the DL-based method, (c1)–(c3) show the close-ups of the regions inside red rectangles in (b1)–(b3), (d1)–(d3) are the bone results of the RF + ACM-based method, (e1)–(e3) show the close-ups of the regions indicated by red rectangles in (d1)–(d3), (f1)–(f3) are the bone results of the proposed ARF + FS + JIG + IR-based method, and (g1)–(g3) show the close-ups of the regions indicated by red rectangles in (f1)–(f3).



**Fig. 12** Comparison between PCT and CT for pelvic data: (a) the MRI, (b) the CT, (c) the PCT obtained from proposed algorithm, and (d) the difference image between PCT and CT.



**Fig. 13** MAE, PSNR, and NCC for (a1), (b1), and (c1) each brain and (a2), (b2), and (c2) pelvic patient.

In addition, the PCT generated by our method has potential to be used in dose calculation for other treatment modality, such as proton radiation therapy. However, proton dose deposition is widely known to be more sensitive and has more uncertainty than photon beams, thus a comprehensive study is necessary to test its feasibility. The PCT generation in this study could also be used to generate an attenuation correction map of PET. This application was not the focus of this study and using our method for PET attenuation correction will be a potential future study.

### 5 Conclusion

We proposed a learning-based approach to synthesize a pseudo-CT (PCT) image from a routine MR image for potential MRI-based treatment planning, which combines alternating RF with patch-based anatomical signature to effectively capture the relationship between the CT and MR images. We demonstrated that the proposed method is capable of reliably generating a CT image from its MRI counterpart on brain and pelvic data. This PCT synthesis technique could be a useful tool for MRI-based radiation treatment planning.

### Disclosures

The authors declare no conflicts of interest.

### Acknowledgments

This research was supported in part by the National Cancer Institute of the National Institutes of Health under Award No. R01CA215718 and the Department of Defense

(DoD) Prostate Cancer Research Program (PCRP) Award No. W81XWH-13-1-0269.

### References

1. M. A. Schmidt and G. S. Payne, "Radiotherapy planning using MRI," *Phys. Med. Biol.* **60**(22), R323–R361 (2015).
2. S. Devic, "MRI simulation for radiotherapy treatment planning," *Med. Phys.* **39**(11), 6701–6711 (2012).
3. V. S. Khoo and D. L. Joon, "New developments in MRI for target volume delineation in radiotherapy," *Br. J. Radiol.* **79**(Special issue 1), S2–15 (2006).
4. K. Ulin, M. M. Urie, and J. M. Cherlow, "Results of a multi-institutional benchmark test for cranial CT/MR image registration," *Int. J. Radiat. Oncol.* **77**(5), 1584–1589 (2010).
5. J. M. Edmund and T. Nyholm, "A review of substitute CT generation for MRI-only radiation therapy," *Radiat. Oncol.* **12**(1), 28 (2017).
6. E. Johnstone et al., "Systematic review of synthetic computed tomography generation methodologies for use in magnetic resonance imaging-only radiation therapy," *Int. J. Radiat. Oncol.* **100**(1), 199–217 (2018).
7. J. Uh et al., "MRI-based treatment planning with pseudo CT generated through atlas registration," *Med. Phys.* **41**(5), 051711 (2014).
8. J. Sjolund et al., "Generating patient specific pseudo-CT of the head from MR using atlas-based regression," *Phys. Med. Biol.* **60**(2), 825–839 (2015).
9. X. Yang and B. Fei, "Multiscale segmentation of the skull in MR images for MRI-based attenuation correction of combined MR/PET," *J. Am. Med. Inform. Assoc.* **20**(6), 1037–1045 (2013).
10. J. Kim et al., "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.* **91**(1), 39–47 (2015).

11. J. Korhonen et al., "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**(1), 011704 (2014).
12. J. M. Edmund et al., "A voxel-based investigation for MRI-only radiotherapy of the brain using ultra short echo times," *Phys. Med. Biol.* **59**(23), 7501–7519 (2014).
13. J. Cabello et al., "MR-based attenuation correction using ultrashort-echo-time pulse sequences in dementia patients," *J. Nucl. Med.* **56**(3), 423–429 (2015).
14. A. Johansson, M. Karlsson, and T. Nyholm, "CT substitute derived from MRI sequences with ultrashort echo time," *Med. Phys.* **38**(5), 2708–2714 (2011).
15. S. H. Hsu et al., "Investigation of a method for generating synthetic CT models from MRI scans of the head and neck for radiation therapy," *Phys. Med. Biol.* **58**(23), 8419–8435 (2013).
16. M. S. Gudur et al., "A unifying probabilistic Bayesian approach to derive electron density from MRI for radiation therapy treatment planning," *Phys. Med. Biol.* **59**(21), 6595–6606 (2014).
17. M. Kapanen and M. Tenhunen, "T1/T2\*-weighted MRI provides clinically relevant pseudo-CT density data for the pelvic bones in MRI-only based radiotherapy treatment planning," *Acta Oncol.* **52**(3), 612–618 (2013).
18. D. Andreassen et al., "Patch-based generation of a pseudo CT from conventional MRI sequences for MRI-only radiotherapy of the brain," *Med. Phys.* **42**(4), 1596–1605 (2015).
19. D. Andreassen, K. Van Leemput, and J. M. Edmund, "A patch-based pseudo-CT approach for MRI-only radiotherapy in the pelvis," *Med. Phys.* **43**(8), 4742–4752 (2016).
20. S. Aouadi et al., "Sparse patch-based method applied to MRI-only radiotherapy planning," *Phys. Med.* **32**(Suppl. 3), 309 (2016).
21. A. Torrado-Carvajal et al., "Fast patch-based pseudo-CT synthesis from T1-weighted MR images for PET/MR attenuation correction in brain studies," *J. Nucl. Med.* **57**(1), 136–143 (2016).
22. Y. Lei et al., "Magnetic resonance imaging-based pseudo computed tomography using anatomic signature and joint dictionary learning," *J. Med. Imaging* **5**(3), 034001 (2018).
23. T. Huynh et al., "Estimating CT image from MRI data using structured random forest and auto-context model," *IEEE Trans. Med. Imaging* **35**(1), 174–183 (2016).
24. X. F. Yang et al., "Pseudo CT estimation from MRI using patch-based random forest," *Proc. SPIE* **10133**, 101332Q (2017).
25. D. Andreassen et al., "Computed tomography synthesis from magnetic resonance images in the pelvis using multiple random forests and auto-context features," *Proc. SPIE* **9784**, 978417 (2016).
26. D. Nie et al., "Estimating CT image from MRI data using 3D fully convolutional networks," in *Deep Learning and Data Labeling for Medical Applications*, G. Carneiro et al., Eds., pp. 170–178, Springer, Cham (2016).
27. X. Han, "MR-based synthetic CT generation using a deep convolutional neural network method," *Med. Phys.* **44**(4), 1408–1419 (2017).
28. D. Nie et al., "Medical image synthesis with deep convolutional adversarial networks," *IEEE Trans. Biomed. Eng.* **65**(12), 2720–2730 (2018).
29. X. Yang et al., "A learning-based approach to derive electron density from anatomical MRI for radiation therapy treatment planning," *Int. J. Radiat. Oncol.* **99**(2), S173–S174 (2017).
30. M. W. Ayeche and D. Ziou, "Automated feature weighting and random pixel sampling in k-means clustering for terahertz image segmentation," in *2015 IEEE Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 35–40 (2015).
31. S. Schulter et al., "Alternating regression forests for object detection and pose estimation," in *Proc of IEEE Int. Conf. on Computer Vision*, pp. 417–424 (2013).
32. Z. Tu and X. Bai, "Auto-context and its application to high-level vision tasks and 3D brain image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.* **32**(10), 1744–1757 (2010).
33. N. J. Tustison et al., "N4ITK: improved N3 bias correction," *IEEE Trans. Med. Imaging* **29**(6), 1310–1320 (2010).
34. S. P. M. Crijns, B. W. Raaymakers, and J. J. W. Lagendijk, "Real-time correction of magnetic field inhomogeneity-induced image distortions for MRI-guided conventional and proton radiotherapy," *Phys. Med. Biol.* **56**(1), 289–297 (2011).
35. T. Wang et al., "MRI-based treatment planning for brain stereotactic radiosurgery: dosimetric validation of a learning-based pseudo-CT generation method," *Med. Dosim.* (2018).

Biographies for the authors are not available.