Computed Tomography Image Estimation by Statistical Learning Methods

Fekadu L. Bayisa* Xijia Liu† Anders Garpebring‡ Jun Yu§

Abstract

There is increasing interest in computed tomography (CT) image estimations from magnetic resonance (MR) images. The estimated CT images can be utilised for attenuation correction, patient positioning, and dose planning in diagnostic and radiotherapy workflows. This study presents a statistical learning method for CT image estimation. We have used predefined tissue-type information in a Gaussian mixture model to explore the estimation. The performance of our method was evaluated using cross-validation on real data. In comparison with the existing model-based CT image estimation methods, the proposed method has improved the estimation, particularly in bone tissues. Evaluation of our method shows that it is a promising method to generate CT image substitutes for the implementation of fully MR-based radiotherapy and PET/MRI applications.

Keywords— Computed tomography; magnetic resonance imaging; CT image estimation; pseudo-CT image; supervised learning; Gaussian mixture model

1 Introduction

Diagnostic imaging technologies such as magnetic resonance imaging (MRI) and computed tomography (CT) are used to obtain detailed anatomical images, although the ionising radiation of CT is associated with an increased risk of radiation-induced cancer (Mathews et al., 2013). MRI is suitable for tumour localisation, which increases the effectiveness of radiation dose planning (Suit and du Bois W.,

*Department of Mathematics and Mathematical Statistics, Umeå University, 901 87, Umeå, Sweden.

†Department of Mathematics and Mathematical Statistics, Umeå University, SE–901 87 Umeå, Sweden.

‡Department of Radiation Sciences, Umeå University, 901 87, Umeå, Sweden.

§Department of Mathematics and Mathematical Statistics, Umeå University, SE–901 87 Umeå, Sweden.

1991; Just et al., 1991; Heesters et al., 1993). The interest in removing MRI-to-CT image fusion uncertainties, the development of MRI linear accelerators, and the tumour-localising ability of MRI have encouraged the development of fully MRI-based radiotherapy planning (Boydev et al., 2017). However, it is challenging for MRI to delineate solid bone structures due to the short signal lifetimes and low proton densities in bone tissue (Wiesinger et al., 2016). Besides, MR images do not provide electron density information for dose calculation. In addition to its geometrical and intensity distortions, MRI also lacks digitally reconstructed radiograph generation for verification of patient positioning (Tanner et al., 2000; Yu et al., 2001; Karger et al., 2003; Boydev et al., 2017). Therefore, coregistered CT and MR images are regarded as a complementary procedure (Chen et al., 2004; Boettger et al., 2008; Karlsson et al., 2009).

Recently, there has been a growing interest in the development and application of integrated positron emission tomography (PET)/MRI scanners, and these scanners acquire PET and MR images simultaneously in brain imaging studies (Herzog et al., 2010; Catana et al., 2012, 2013). The implementation of integrated PET/MRI scanners requires information about electron densities for attenuation correction (Burgos et al., 2014), and this information is obtained from CT images. We can obtain MR images from the integrated scanners; however, MR images do not provide attenuation-related measurement of electron densities (Schmidt and Payne, 2015; Chen et al., 2017). As a result, we need to develop a reliable CT image estimation method for the implementation of fully MRI-based radiotherapy and for the integration of PET and MRI scanners.

The estimated CT images can be used to enable accurate MRI-based radiotherapy, and they are essential for attenuation correction in PET imaging. In this regard, Korhonen et al. (2014) and Korhonen (2015) used a voxel-based dual second-order polynomial model on image data of prostate cancer patients that allowed for the clinical implementation of in-house-developed MRI-only methods. Hsu et al. (2013) utilised probabilistic classification of voxels to obtain CT images that are currently supporting the workflows of radiation oncology treatment planning in the brain. Furthermore, solutions to the problem of CT substitute generation are becoming available commercially (Siversson et al., 2015; Frantzen-Steneker, 2015; Maspero et al., 2017; Kemppainen et al., 2017). Several CT image estimation methods have been reported, and the estimated CT images are acceptable for clinical implementation. However, the lack of established methods for evaluating the methods and for reporting the consistency of the methods has challenged and delayed widespread clinical implementation (Edmund and Nyholm, 2017).

Huynh et al. (2016) used a structured random forest and auto-context model to estimate CT images from MR images, but this approach is subjective and its performance depends on the quality of features that are extracted from the MR images. A three-dimensional fully convolutional neural network model has also been developed for patch-wise estimation of CT images from MR images (Nie et al., 2016). In terms of mean absolute error and peak signal-to-noise ratio, this method has better performance on experimental data than structured random forest and atlas-based

methods. However, its success depends on the geometry of the data. Arabi et al. (2016) proposed a two-step atlas-based algorithm to generate pseudo-CT images from MR image sequences. Their method provides better bone tissue identification accuracy than the conventional segmentation and atlas-based methods. A Gaussian mixture model (GMM) has also been utilised to explore CT image estimation and the associated estimation uncertainty (Johansson et al., 2011, 2012). Taking contextual information into account, Kuljus et al. (2017) have also investigated CT image estimation using a hidden Markov model (HMM) and a Markov random field model (MRF). In terms of mean absolute error, HMM has better performance than MRF and GMM. In comparison with HMM and GMM, MRF has superior estimation quality in bone tissues. However, MRF is computationally expensive. Tissue-dependent performance of the statistical models is the main challenge in CT image estimation, and the estimation quality is poor for air and for bone tissues (Johansson et al., 2012; Kuljus et al., 2017). This challenge led Bayisa et al. to examine CT image estimation by partitioning the training data into non-bone and bone tissues, and they used GMM and a skewed-Gaussian mixture model (SGMM) to estimate CT image intensities on each partition (Bayisa and Yu, 2017). Compared to HMM and MRF, the partitioning approach has better dense bone tissue estimation. The predictive quality of the approach depends on the performance of GMM and SGMM on the full data, however, and it is evident that GMM has poor CT image estimation quality on the full data (Kuljus et al., 2017). This issue has motivated the present work to further investigate and improve the quality of CT image estimation by combining a supervised statistical learning method and GMM.

The main purpose of this article is to further explore CT image estimation and to improve the quality of CT image prediction using statistical learning methods. To achieve this aim, we propose a classification method, the RUSBoost algorithm, at a supervised learning stage of the CT image estimation process. The RUSBoost algorithm uses random undersampling in an adaptive boosting algorithm to overcome the class imbalance problem in the classification of real datasets (Galar et al., 2012; Seiffert et al., 2010). Decision trees are exploited in the RUSBoost algorithm as weak classifiers[a] because they are usually taken to be the best out-of-the-box learners (Praveena and Jaiganesh, 2017). The trained RUSBoost method is used to predict the tissue types of the new MR images for which we seek the corresponding CT image. The predicted tissue types and GMM are utilised to obtain the desired estimated CT image. Hereafter we call our method RGMM. We have also compared the CT image estimation performance of RGMM, HMM, GMM, MRF, and GMM* (when GMM is used at the supervised learning stage, see Bayisa and Yu (2017)).

This article is organised as follows. Section 2 describes the details of our approach, and the results of the study are presented in section 3. We discuss the implication of the results in section 4 and present the conclusions of the study in

[a] Let $\epsilon > 0$ and $\delta > 0$. A weak classifier is the algorithm that can achieve at least slightly better error rate ϵ than random guessing $\left(\epsilon \geq \frac{1}{2} - \lambda, \text{where } \lambda > 0\right)$ with a probability $(1 - \delta)$.

section 5.

2 Materials and Methods

In this section, we describe the data acquisition, CT image estimation, and assessment methods.

2.1 Data acquisition

Three-dimensional CT and MR images were acquired from the heads of nine patients. Four MR images were obtained from each patient using two dual echo UTE sequences with flip angles of 10 and 30 degrees. The UTE sequences were sampled from a first echo (free induction decay) and a second echo (gradient echo) from the same excitation with echo times of 0.07 and 3.76 milliseconds, respectively. Each MR image was reconstructed to a 192 $\times$ 192 $\times$ 192 dimensional matrix. An entry in the matrix represents a signal intensity corresponding to a voxel with size 1.33 $\times$ 1.33 $\times$ 1.33 mm^3. For each patient, one CT image was acquired using a GE Lightspeed with a 2.5 mm slice thickness, and the image was reconstructed with an in-plane resolution of 0.78 $\times$ 0.78 mm^2. A binary mask was also developed for each patient to demarcate the area of interest. The binary mask, the CT image, and the four MR images of each patient were co-registered and resampled to the same resolution using linear interpolation. For further technical details, we refer to Johansson et al. (2011). We arranged the voxel values of the CT image, the binary mask, and the four MR images into six columns to obtain data for each patient. We arranged the voxel values of the CT image, the binary mask, and the four MR images into six columns to obtain data for each patient. The organised data of the patients were column stacked, and data corresponding to the region of interest were selected for model fitting. Figure 1 shows a sagittal image slice for a representative patient.

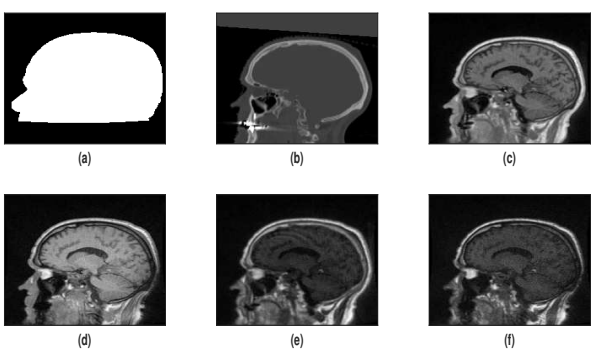

Figure 1. Binary mask (a), CT image (b) and MR images (c-f).

2.2 Gaussian mixture model

Let Y_i and $\mathbf{X}_i = (X_{i1}, X_{i2}, \cdots, X_{id})$ represent the CT image intensity and its corresponding d-dimensional intensities of the MR images at voxel i. In our real data, we have four MR images, $d = 4$. Let T_i be its corresponding known tissue type, which represents some knowledge about tissue types. The class probability density of the random vector $(Y_i, \mathbf{X}_i)$ is assumed to be Gaussian:

$$(Y_i, \mathbf{X}_i) \mid T_i = k \sim \mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k),$$

where $i = 1, 2, \cdots, n$, $k = 1, 2, \cdots, K$, n is the number of voxels, and K is the known number of classes. In this work, $K = 2$. For each known tissue type, we assume that there is a latent variable Z_i that represents an unobserved subtissue type at voxel i. To take this into account, we have assumed a more flexible model given by

$$(Y_i, \mathbf{X}_i) \mid T_i = k, Z_i = j \sim \mathcal{N}(\boldsymbol{\mu}_{kj}, \boldsymbol{\Sigma}_{kj}),$$

where $j = 1, 2, \cdots, J_k$, and J_k is the number of subtissue types within the known tissue type k. Consequently, we have a Gaussian mixture model (GMM) for each known tissue type, and this is given by

$$(Y_i, \mathbf{X}_i) \mid T_i = k \sim \sum_{j=1}^{J_k} \pi_{kj} \mathcal{N}(\boldsymbol{\mu}_{kj}, \boldsymbol{\Sigma}_{kj}),$$

where

$$\pi_{kj} \geq 0, \ \sum_{j=1}^{J_k} \pi_{kj} = 1.$$

The GMM parameters can be estimated by an iterative maximising procedure using the concept of incomplete data via the EM-algorithm (Dempster et al., 1977). We use the mean squared error to select the optimal estimates.

Two stages are used during the parameter estimation process. First, we estimate the model parameters given the number of unknown tissue types. The mean squared error is utilised to select the optimal parameter estimates for the given number of unknown tissue types. Second, we select the number of unknown tissue types by using the estimated models on the validation dataset. The final optimal parameter estimates are the ones that have smaller mean squared errors or mean absolute errors for the validation dataset. The stopping criterion used by Bayisa and Yu (2017) is used to control the convergence of the estimation process.

2.3 Estimation of CT images

We are interested in estimating the CT image intensity Y_i given new MR images $\mathbf{X}_i$ and its known tissue type T_i at voxel i. A point estimator of CT image intensity

can be given by

$$E\left[Y_i \mid \mathbf{X}_i, T_i = k, \boldsymbol{\Theta}\right] = \sum_{j=1}^{J_k} \beta_j E\left[Y_i \mid \mathbf{X}_i, T_i = k, Z_i = j, \boldsymbol{\Theta}\right], \tag{1}$$

where

$$\boldsymbol{\Theta} = \left\{\pi_{kj}, \boldsymbol{\mu}_{kj}, \boldsymbol{\Sigma}_{kj} \mid j = 1, 2, \cdots, J_k; k = 1, 2, \cdots, K\right\},$$

is the set of all parameters. Bayes' theorem can be used to compute the weight $\beta_j = P\left(Z_i = j \mid \mathbf{X}_i, T_i = k, \boldsymbol{\Theta}\right)$. The expected value $E\left[Y_i \mid \mathbf{X}_i, T_i = k, Z_i = j, \boldsymbol{\Theta}\right]$ is the conditional expectation in the multivariate normal distribution.

To predict the CT image from the new MR images, the first-layer tissue type T_i for the MR images at voxel i needs to be estimated. For this purpose, we train a classifier at a supervised learning stage. The predicted tissue types and GMM parameters are exploited to obtain the desired CT image from the new MR images.

2.4 Supervised learning method

Using the observed CT image intensity y_i that corresponds to the intensity $\mathbf{x}_i$ of the MR images, we define a class label or tissue type of the MR images at voxel i as follows.

$$t_i = \begin{cases} 0, & \text{if } \ y_i \leq 100\,\text{HU}, \\ 1, & \text{otherwise}, \end{cases}$$

where 100 HU is used as CT image intensity threshold value, see Bayisa and Yu (2017) and HU is a Hounsfield unit.

2.4.1 Extraction of spatial features

Important information in datasets can be captured by well-conceived new features, and it is reasonable to consider spatial features during the CT image estimation process. CT and MR images are three-dimensional images, and a given voxel in the images that is not on the boundary of the images has six closest and a maximum of twenty-six neighbouring voxels. Let $\mathbf{x}_i^s$ be a vector of intensities of MR images at the six or the twenty-six closest voxels to voxel i.

2.4.2 RUSBoost algorithm

Let $\mathbf{x}_i^c = (\mathbf{x}_i, \mathbf{x}_i^s)$ be a vector of the original and the extracted spatial features at voxel i, and let t_i be its corresponding known class label or tissue type. We use the RUSBoost algorithm to train a supervised learning method using the feature vector $\mathbf{x}_i^c$ and the class label t_i. The RUSBoost algorithm is shown in Algorithm 1. We have used decision trees as weak learners in the RUSBoost algorithm, and we have explored all possible binary splits on every predictor. The optimisation of the binary split is subjected to the maximum number of splits, the purity, and the

Algorithm 1 RUSBoost Algorithm

a Let $\mathbf{D}$ be a training set given by $\mathbf{D} = \{(\mathbf{x}_1^c, t_1), (\mathbf{x}_2^c, t_2), \cdots, (\mathbf{x}_n^c, t_n)\}$, where $\mathbf{x}_i^c \in \mathbf{X}^c$, $t_i \in \mathbf{T}$, $i = 1, 2, \cdots, n$, $\mathbf{X}^c$ is a feature space and $\mathbf{T}$ is a label set;

b Random undersampling (RUS) based on a minority class label in $\mathbf{T}$, maximum number of learners in the ensemble M and weak learning algorithm **WeakLearn**; Let $\mathbf{S} = \{(i,t) \mid i = 1, 2, \cdots, n, t \neq t_i\}$ be the set of possible mislabels. In this notation, i and t represent the index of $\mathbf{x}_i^c$ and one of the incorrect labels of $\mathbf{x}_i^c$;

c Using $\mathbf{S}$, initialise the mislabel distribution $\mathbf{W}_1$, which is given by

$$\mathbf{W}_1 = \left\{ W_1(i,t) = \frac{1}{|\mathbf{S}|} : (i,t) \in \mathbf{S} \right\},$$

where $|\mathbf{S}|$ denotes the cardinality of the set $\mathbf{S}$;

d For $j = 1, 2, \cdots, M$, do

1 Using RUS, obtain the modified training dataset $\mathbf{D}'_j$ and its distribution $\mathbf{W}'_j$; Call **WeakLearn** and provide it with the training dataset $\mathbf{D}'_j$ and $\mathbf{W}'_j$; Obtain a weak classifier h_j: $\mathbf{X}^c \times \mathbf{T} \to [0,1]$;

2 Compute the pseudo-loss of the classifier h_j and Compute the weight update parameter α_j

$$\epsilon_j = \sum_{(i,t) \in \mathbf{S}} W_j(i,t) \left[1 - h_j(\mathbf{x}_i^c, t_i) + h_j(\mathbf{x}_i^c, t)\right]; \quad \alpha_j = \frac{\epsilon_j}{1 - \epsilon_j};$$

3 Update and normalise the weight $W_j(i,t)$

$$\begin{aligned} W_{j+1}(i,t) &= W_j(i,t)\, \alpha_j^{\frac{1}{2}\left[1 + h_j(\mathbf{x}_i^c, t_i) - h_j(\mathbf{x}_i^c, t)\right]}, \\ W_{j+1}(i,t) &= \frac{W_{j+1}(i,t)}{\sum_{(i,t) \in \mathbf{S}} W_{j+1}(i,t)}; \end{aligned}$$

e Obtain the final strong learner $H(\mathbf{x}^c) = \underset{t \in \mathbf{T}}{\operatorname{argmax}} \sum_{j=1}^{M} h_j(\mathbf{x}^c, t) \log \frac{1}{\alpha_j}$;

minimum leaf size of a node. The split predictor among all possible splits of all predictors can be obtained using the GINI index given by

$$GINI(v) = 1 - \sum_{t=1}^{K} p_{tv}^2,$$

where K is the number of class labels, v is a node, and p_{tv} is the proportion of the class label t at node v. GINI index is a measure of node impurity and minimizing it can lead to a purity of a node.

We use the trained RUSBoost method to predict the first-layer tissue types of the new MR images. The predicted first-layer tissue types and the estimated GMM parameters can then be used to obtain the desired CT image estimate. We have summarised the method utilised in this work in Figure 2.

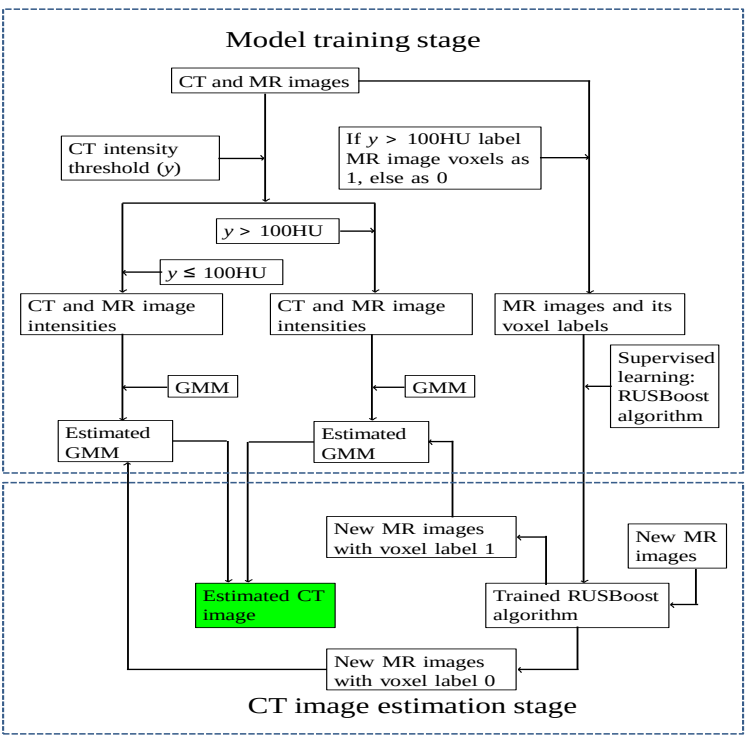

Figure 2. Conceptual model of RGMM that involves the training stage of the models and the CT image estimation stage.

2.5 Model validation method

The conceptual model involves two stages. The first stage is the model training stage that consists of training GMM and the RUSBoost algorithm. The second stage is the CT image estimation stage. To evaluate the quality of the RUSBoost classification algorithm, we use a 10-fold cross-validation method. Using the model trained on out-of-fold observations, we can compute the classification error of the method for the in-fold observations. This is calculated by

$$err = \frac{1}{n}\sum_{i=1}^{n} 1_{\{\hat{y}_i \neq y_i\}},$$

where

$$1_{\{\hat{y}_i \neq y_i\}} = \begin{cases} 1, & \text{if } \hat{y}_i \neq y_i, \\ 0, & \text{otherwise,} \end{cases}$$

and $\hat{y}_i$ is the predicted class label. The number of weak learners can be plotted against the cumulative of the averages of the classification errors in order to validate the predictive quality of the method.

An overall accuracy measure can be exploited to select the best-trained classification model. However, it is not a proper measure in imbalanced class distribution. The reason is that less prevalent classes have very little impact on the accuracy measure compared to the prevalent classes (Joshi et al., 2001). Buckland and Gey (1994) and Powers (2011) utilised precision (Pr), recall (Re), and F-score (Fs) to investigate the performance of a classification method for skewed class distribution. The measures are obtained by

$$\begin{aligned} Pr &= \frac{TP}{TP+FP}, \; Re = \frac{TP}{TP+FN}, \\ Fs &= \frac{\left(1+\beta^2\right) \times Re \times Pr}{\beta^2 \times Re + Pr}, \end{aligned}$$

where TP, FP and FN represent True Positive, False Positive, and False Negative. The positive and negative classes represent the minority and the majority classes in the data. The parameter β denotes the relative importance of precision versus recall and is usually set to 1. Standard classification methods usually produce classifiers that do not accurately predict the minority class. Therefore, we need to obtain a classification method that can improve the recall without affecting the precision. However, the roles of recall and precision are usually conflicting. As a result, we need to use the F-score to assess the goodness of fit of the classifier for the minority class.

We evaluate the CT image estimation performance of RGMM using a leave-one-out cross-validation method. One dataset from a patient is kept for validating

the model, and the remaining datasets from the other patients are used for training the model. This procedure is repeated for each patient. For a given validation dataset, let Y_i; and $\hat{Y}_i$ be the intensity of the CT image, and its estimated intensity, respectively, at voxel i. We use the mean absolute error (MAE) for measuring estimation accuracy. This is a robust measure that is fairly insensitive to outliers and is given by

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{Y}_i - Y_i|.$$

Better models have lower average MAEs. In addition, we use a smoothed residual plot to assess the predictive quality of the model through the tissues of the head. A moving average over non-overlapping windows in CT image intensities is utilised to obtain the smoothed residual plot, and we can exploit the average of residuals and absolute residuals as a summary measure over the windows.

In summary, the MAE, smoothed residual plot, and smoothed absolute residual plot are used to compare the CT image estimation performance of RGMM with the existing model-based methods. We used the same nine datasets as were used by Kuljus et al. (2017). Kuljus et al. (2017) experienced numerical difficulties in estimating the existing model-based methods such as MRF and GMM for the nine datasets. Following that, the authors used nine and five (excluding four datasets) datasets to evaluate the estimation performance of HMM, MRF, and GMM. In line with Kuljus et al., we needed to explore the robustness of our method based on the nine and five datasets. When comparing RGMM and HMM, all nine patients were used. When including MRF, GMM, and GMM* in the comparison, only five of the patients could be included due to instability in MRF and GMM.

3 Results

Histogram plots of the data from the nine patients showed that four of the datasets have high intensities in the MR images acquired with a 10 degree uniform flip angle. These high intensities are observed in some regions of bone tissues.

3.1 Result: Supervised learning method

We utilised CT image intensities to label each voxel of its corresponding MR images as 0 or 1. For the nine datasets, the data show that 18.49% of instances belong to the minority class (with label 1). We began the training of the RUSBoost method with 500 decision trees and explored the training of the method using several splits and leaf sizes for the decision trees. The investigation suggested that a maximum of 400 splits and a minimum of 5 leaf sizes are enough to obtain the desired number of decision trees. Using 10-fold cross-validation, the evaluation of the predictive quality of the algorithm showed that the method achieves a classification error of

under 9.05% by employing 150 or more decision trees. The classification error of our method is shown in Figure 3. We used the F-score to select the best model. The predictive performance of the method depends on the number of extracted spatial features, and when we consider the six extracted spatial features, the best model has an F-score of 74.07% with a corresponding overall accuracy of 90.05%. On the other hand, the best model has an F-score of 76.48% when we take the twenty-six extracted spatial features into account. In this case, the overall accuracy of the best model is 91%. Although the change in F-score is small, the effect of using the twenty-six spatial features on CT image estimation is non-negligible.

For the five datasets, the data indicate that 17.45% of the instances belong to the minority class. of the instances belong to the minority class. Under the same setting as for nine datasets, the method achieves a classification error of under 5.95% by employing 150 or more trees. Figure 3 demonstrates the classification error of the algorithm. The best model has an F-score of 81.61% when the six extracted spatial features are taken into account, and it has a 93.55% overall accuracy. When we consider the twenty-six extracted spatial features, the best model has an F-score of 83.27% and an overall accuracy of 94.12%.

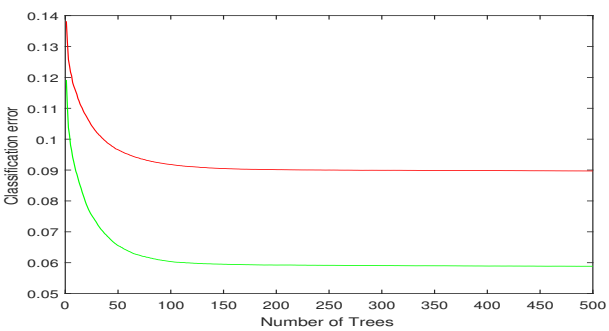

Figure 3. Classification error of the RUSBoost algorithm for the nine patients (red) and the five patients (green).

The F-score and overall accuracy appear to depend on the number of extracted spatial features. Hereinafter, we use the trained RUSBoost algorithm consisting of 150 trees and take the twenty-six extracted spatial features into account.

3.2 Result: RGMM

In this section, we present results obtained by RGMM and compare its performance with the existing model-based methods.

For the case of nine datasets, the optimal estimates of the RGMM parameters are obtained for $J_0 = 5$ and $J_1 = 6$, which are the optimal underlying numbers of classes. According to Kuljus et al. (2017), the optimal estimates of HMM pa-

rameters are obtained for 8 underlying numbers of classes, and they showed that HMM outperforms GMM, which is numerically less stable on the nine datasets. Kuljus et al. (2017) also encountered numerical difficulty in estimating the MRF for the nine datasets. Therefore, we needed to compare the CT image estimation performance of RGMM and HMM for the nine datasets. Using the leave-one-out cross-validation method, we have summarised the mean absolute errors of the CT image estimation in Table 1.

Table 1. Summary of the mean absolute errors of the estimation for the nine patients (or the nine datasets).

Patient	RGMM	HMM
1	145.80	146.31
2	139.51	146.15
3	291.76	297.35
4	143.21	157.00
5	179.50	259.67
6	152.02	199.34
7	302.08	351.73
8	140.45	153.21
9	155.49	153.87
mean	**183.31**	**207.18**

Table 1 shows that RGMM outperforms HMM by approximately 23 HU on average. This implies that the supervised learning method played a major role in improving CT image estimation. One of the main advantages of our method is that it is more robust across the patients than HMM. For instance, HMM has poor performance for patients 3, 5, 6, and 7, especially on bone tissues, see Table 2. We noticed that the MR images of the four heads acquired at 10 degree uniform flip angles had peculiarly high intensities in certain regions of bone tissues. This was not observed in the same regions of the remaining heads. Table 2 shows the robustness of our method on bone tissue-dominated regions, and our approach improved the estimation by approximately 107 HU on average.

Table 2. Summary of the mean absolute errors for the nine patients in bone tissue-dominated regions.

Patient	RGMM	HMM
1	284.93	331.11
2	295.42	359.93
3	564.65	615.97
4	269.80	342.15
5	358.15	724.20
6	303.35	498.62
7	557.98	616.41
8	238.37	296.02
9	297.64	355.82
mean	**352.25**	**460.03**

We merged the CT image intensities of the nine patients and denoted this by mCT. Similarly, we merged the estimated CT image intensities of the nine patients and represented this by sCT. Over non-overlapping windows on mCT with a window size of 20 HU, we computed the average over the windows for both mCT and $|mCT–sCT|$. Plotting the averages, we obtained the smoothed absolute residual plot shown in Figure 4. The smoothed plot shows that RGMM has better CT image estimation quality than HMM. In particular, RGMM has superior bone tissue estimation quality.

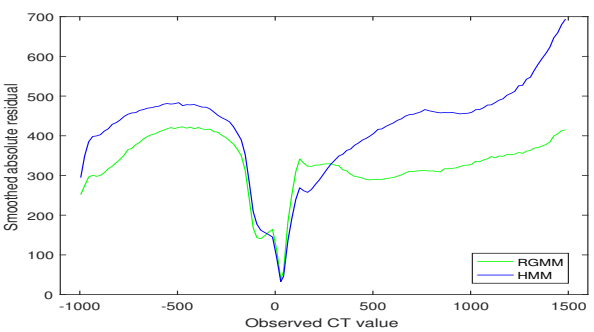

Figure 4. Smoothed absolute residual plot for the nine patients.

Plotting mCT against $mCT–sCT$, we obtained the smoothed residual plot shown in Figure 5. Compared to HMM, our method improved both the underestimation and overestimation problem in CT image estimation.

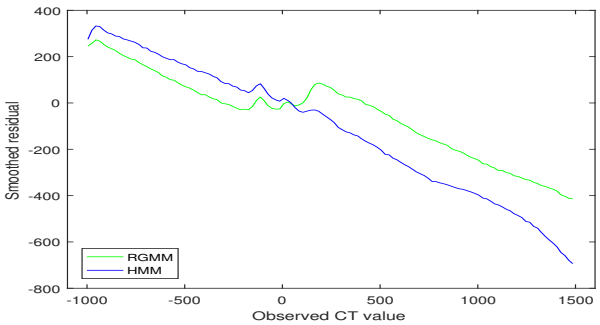

Figure 5. Smoothed residual plot for the nine patients.

To compare the robustness of our method with the existing model-based methods and results, we needed to further investigate CT image estimation on the remaining five datasets 1, 2, 4, 8, and 9. We obtained the optimal estimates of RGMM parameters for $J_0 = 6$ and $J_1 = 6$, which are the optimal underlying number of classes. According to Bayisa and Yu (2017), the optimal estimates of HMM, MRF, GMM, and GMM* parameters are obtained for an optimal underlying number of classes of 5 in HMM and MRF and an optimal underlying number of classes of 8 in GMM, and 6 in GMM* (on both bone and non-bone tissues). Using the leave-one-out cross-validation method, we have summarised the mean absolute errors of the models in Table 3.

Table 3. Summary of mean absolute errors of the estimation for the five patients.

Patient	RGMM	GMM*	HMM	MRF	GMM
1	135.72	144.58	133.56	137.95	145.67
2	134.31	152.04	138.68	137.34	153.40
4	140.01	161.28	142.92	158.99	163.85
8	125.35	146.62	139.17	145.05	146.78
9	133.48	160.62	143.89	165.36	158.85
mean	**133.77**	**153.03**	**139.65**	**148.94**	**153.71**

Table 3 shows that RGMM improved the estimation quality by approximately 6 HU on average and that RGMM is more stable. Most importantly, RGMM has much better performance on tissues dominated by bone (tissues that have CT image intensities greater than 100 HU). Table 4 summarises the mean absolute errors of the models for tissues dominated by bone. Table 4 shows that RGMM improved the estimation quality in bone tissue-dominated regions by approximately 62 HU on average. We also utilised a moving average to evaluate the performance of the

Table 4. Mean absolute errors for tissues dominated by bone.

Patient	RGMM	GMM*	HMM	MRF	GMM
1	265.49	315.16	324.94	307.95	314.41
2	273.84	348.05	360.03	328.89	365.12
4	236.20	301.78	331.16	322.48	328.01
8	216.67	269.33	296.04	280.63	292.78
9	282.27	349.28	366.13	357.22	359.56
mean	**254.89**	**316.72**	**335.66**	**319.43**	**331.98**

models on the tissues of the heads. Figure 6 shows the smoothed absolute residual plot for the five patients that was obtained in a similar manner as the smoothed plot for nine patients.

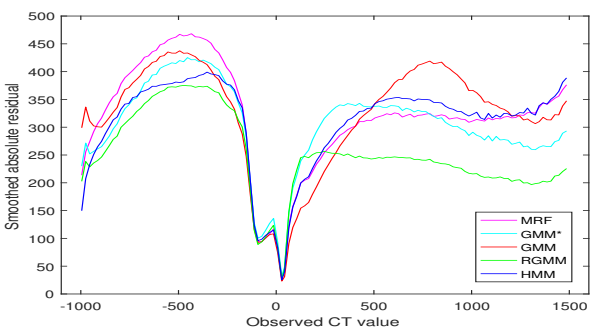

Figure 6. Smoothed absolute residual plot for the five patents.

The smoothed plot shows that RGMM has better performance than the other models. Specifically, our method has better bone tissue estimation quality. We also used the smoothed residual plot shown in Figure 7 to investigate the nature of the models for the tissues. The residual plot demonstrates that the models underestimated for the bone tissues and overestimated for the air spaces of the heads. However, it is evident from the plot that RGMM improved the underestimation problem for bone tissues compared to the other methods used in this work.

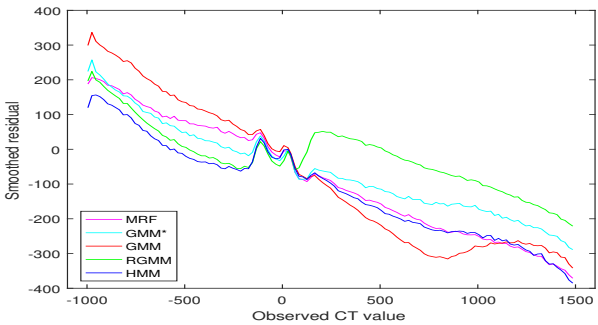

Figure 7. Smoothed residual plot for the five patients.

Existing methods have poor performance in the region of the throat cavity (Johansson et al., 2011; Hildeman et al., 2016), and we observed similar behaviour in our method. The results of our approach and HMM are shown in Figure 8 showing slices of a CT image, its predicted images, and the associated errors for a representative patient.

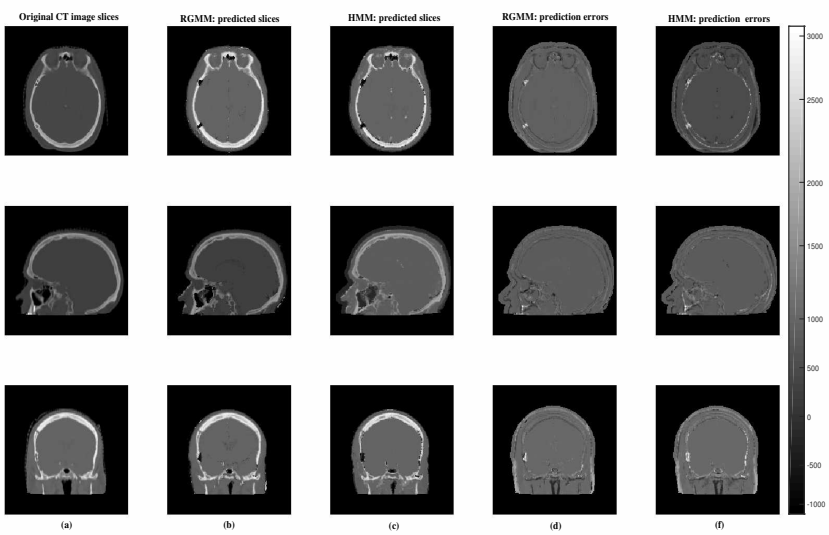

Figure 8. The first column (a) represents slices of CT image, the second and third column (b) and (c) shows the corresponding predicted slices of CT images, and the last two columns (d) and (e) show the errors of the prediction.

4 Discussion

We have presented a statistical learning method for CT image estimation from MR images and have evaluated the method using cross-validation on nine and five datasets.

Several methods have previously been utilised to investigate CT image estimations (Johansson et al., 2012; Nie et al., 2016; Huynh et al., 2016; Arabi et al., 2016; Kuljus et al., 2017; Bayisa and Yu, 2017). Our method is different from these works in that it combines supervised statistical learning and GMM to generate substitute CT images. Our approach also differs from the studies carried out by Keereman et al. (2010), Catana et al. (2010), and Hofmann et al. (2008) because we used UTE MRI sequences with two flip angles at two different echo times and used RGMM to estimate the CT images.

The proposed method has resulted in better accuracy of CT image estimation, especially for bone tissues, which is the most difficult task in the estimation of CT images from MR images (Karlsson et al., 2009). Figures 4 and 6 and Table 1–4 show that our method has better CT image estimation quality compared to the existing model-based methods. We observed that the absolute errors in CT image estimation are not only severe in the bone and air regions, but also at their interfaces with soft tissues. This problem might not be directly attributed to the model and might instead be related to patient motion during data acquisition, changes in the anatomy of the patients, and uncertainties during image registration.

Our method is more stable than the existing model-based methods. Table 1 shows that the variation in the mean absolute errors of the estimation across the heads for RGMM is lower than for HMM, and our method improved CT image estimation by approximately 23 HU on average. In particular, it provided better quality in estimating bone tissue-dominated regions and improved the estimation by approximately 107 HU on average (Table 2). For the nine datasets, the estimations by GMM and MRF are unstable due to ill- conditioning and the inverse problem of covariance matrices in GMM and MRF (Kuljus et al., 2017). This is also the reason that we did not compare our method with GMM*, that is, we did not expect good partitioning of the datasets based on the estimated CT image by GMM on the nine datasets. Furthermore, Table 3 shows that our method is better than GMM*, and it is reasonable to expect a similar result on the nine datasets. RGMM is less sensitive than HMM, MRF, and GMM on heads 3, 5, 6, and 7. The reason for the sensitivity of HMM, MRF, and GMM for the four datasets is that the MR images of the four heads obtained at 10 degree uniform flip angles have strange proton densities in certain regions of bone tissues.

To further investigate the robustness of our method, we compared it with the existing model-based methods and results and explored the estimation of CT images using the remaining five subset datasets. Tables 3–4 show that our method is more stable than the other models. Moreover, it has improved the estimation of bone tissue-dominated regions by approximately 62 HU on average. This shows that our method is relatively consistent and is a promising approach for generating substitute

CT images from MR images.

Johansson et al. (2011) used UTE MR images. We also used UTE MR images in this article. Recently, zero-echo-time (ZTE) has been shown to provide a sequence of MR images that show sufficient contrast between different tissues, especially air and bone. These sequences can be utilised to obtain better estimations of CT images.

An interesting extension of our work is to relax the boundary of the class labelling and to instead try to learn the boundary from the datasets. Another alternative to extend this work is to use supervised learning at the two stages that were utilised to estimate CT images from MR images.

5 Conclusions

We have shown that our method has better CT image prediction quality, especially in bone tissue estimation, compared to previous model-based methods. Evaluation of our method shows that it predicts CT image intensities accurately and that it is a promising method to obtain CT image substitutes for the implementation of fully MRI-based radiotherapy and for PET/MRI applications.

Acknowledgments

This work is supported by the Swedish Research Council grant (Reg. No. 340-2013-5342). The authors thank Adam Johansson and Thomas Asklund for providing us data and David Bolin for providing us Mathlab code for MRF. Moreover, the authors thank Kristi Kuljus for HMM results.

The computations were performed on resources provided by the Swedish National Infrastructure for Computing (SNIC) at High Performance Computing Center North (HPC2N).

Disclosure of Conflicts of Interest

The authors have no relevant conflicts of interest to disclose.

References

References

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.*

Bayisa, F. L. and Yu, J. (2017). Model-based estimation of computed tomography images. *arXiv:1705.03799*.

Boettger, T., Nyholm, T., Karlsson, M., Nunna, C., and Celi, J. C. (2008). Radiation therapy planning and simulation with magnetic resonance images. *Proc. SPIE*, 6918.

Boydev, C., Demol, B., Pasquier, D., Saint-Jalmes, H., Delpon, G., and Reynaert, N. (2017). Zero echo time MRI-only treatment planning for radiation therapy of brain tumors after resection. *Physica Medica*, page http://dx.doi.org/10.1016/j.ejmp.2017.04.028.

Buckland, M. and Gey, F. (1994). The relationship between recall and precision. *American Society for Information Science*, 45(1):12–19.

Burgos, N., Cardoso, M., Thielemans, K., Modat, M., Pedemonte, S., Dickson, J., Barnes, A., Ahmed, R., Mahoney, C., Schott, J., Duncan, J., Atkinson, D., Arridge, S., Hutton, B., and Ourselin, S. (2014). Attenuation correction synthesis for hybrid PET-MR scanners: application to brain studies. *IEEE Trans Med Imaging*, 33(12):2332–2341.

Catana, C., Drzezga, A., Heiss, W., and Rosen, B. (2012). PET/MRI for neurologic applications. *Nucl Med*, 53(12):1916–1925.

Catana, C., Guimaraes, A., and Rosen, B. (2013). PET and MR imaging: the odd couple or a match made in heaven? *Nucl Med*, 54(5):815–824.

Catana, C., Van der Kouwe, A., Benner, T., Michel, C., Hamm, M., Fenchel, M., Fischl, B., Rosen, B., Schmand, M., and Sorensen, A. (2010). Toward implementing an MRI-based PET attenuation-correction method for neurologic studies on the MR-PET brain prototype. *J. Nucl. Med.*, 51:1431–1438.

Chen, K. T., Izquierdo-Garcia, D., Poynton, C. B., Chonde, D. B., and Catana, C. (2017). On the accuracy and reproducibility of a novel probabilistic atlas-based generation for calculation of head attenuation maps on integrated PET/MR scanners. *Nucl Med Mol Imaging*, 44:398–407.

Chen, L., Price, R. A., Jr., Wang, L., Li, J., Qin, L., McNeeley, S., Ma, C. M., Freedman, G. M., and Pollack, A. (2004). MRI-based treatment planning for radiotheraphy: domestic verification for prostate IMRT. *Int. J. Radiation Oncology, Biol., Phys.*, 60(2):636–647.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM-algorithm. *Journal of Royal Statistical Society*, 39(1):1–38.

Edmund, J. M. and Nyholm, T. (2017). A review of substitute CT generation for MRI-only radiation therapy. *Radiation Oncology*, pages 1–15.

Frantzen-Steneker, M. (2015). Sp-0022: Routine QA of an MR-only workflow. *Radiotherpy Oncol.*, 115.

Galar, M., Fernandez, A., Barrenechea, E., Bustince, H., and Herrera, F. (2012). A review on ensembles for the class imbalance problem: bagging, boosting, and hybrid based approaches. *IEEE Trans. Syst., Man Cybern*, 42(4):463–484.

Heesters, M. A., Wijrdeman, H., Struikmans, H., Witkamp, T., and Moerland, M. (1993). Brain tumor delineation based on CT and MR imaging. implications for radiotherapy treatment planning. *Strahlenther Onkol.*, 169(12):729–733.

Herzog, H., Pietrzyk, U., Shah, N., and Ziemons, K. (2010). The current state, challenges and perspectives of MR-PET. *Neuroimage*, 49(3):2072–2082.

Hildeman, A., Bolin, D., Wallin, J., Johansson, A., Nyholm, T., Asklund, T., and Yu, J. (2016). Whole-brain substitute CT generation using markov random field mixture models. *arXiv:1607.02188v2*.

Hofmann, M., Steinke, F., Scheel, V., Charpiat, G., Farquhar, J., Aschoff, P., Brady, M., Scholkopf, B., and Pichler, B. J. (2008). MRI-based attenuation correction for PET/MRI: A novel approach combining pattern recognition and atlas registration. *J. Nucl. Med.*, 49:1875–1883.

Hsu, S., 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(23):8419–8435.

Huynh, T., Gao, Y., Kang, J., Wang, L., Zhang, P., Lian, J., and Shen, D. (2016). Estimating CT image from MRI data using structured random forest and auto-context model. *IEEE transactions on medical imaging*, 35(1):174–183.

Johansson, A., Karlsson, M., and Nyholm, T. (2011). CT substitute derived from MRI sequences with ultrashort echo time. *Medical Physics*, 38(5):2708–2714.

Johansson, A., Karlsson, M., Yu, J., Asklund, T., and Nyholm, T. (2012). Voxel-wise uncertainty in CT substitute derived from MRI. *Medical Physics*, 39(6):3283–3290.

Joshi, M. V., Kumar, V., and Agarwal, R. C. (2001). Evaluating boosting algorithms to classify rare classes: Comparison and improvements. *Proceedings 2001 IEEE International Conference on Data Mining*, pages 257–264.

Just, M., Rösler, H. P., Higer, H. P., Kutzner, J., and Thelen, M. (1991). MRI-assisted radiation therapy planning of brain tumors-clinical experiences in 17 patients. *Magn. Reson. Imaging.*, 9:173–177.

Karger, C. P., Hipp, P., Henze, M., Echner, G., Höss, A., Schad, L., and Hartmann, G. H. (2003). Stereotactic imaging for radiotherapy: accuracy of CT, MRI, PET and SPECT. *Phys. Med. Biol.*, 48:211–221.

Karlsson, M., Karlsson, M. G., Nyholm, T., Amies, C., and Zackrisson, B. (2009). Dedicated magnetic resonance imaging in the radiotherapy clinic. *Int. J. Radiation Oncology, Biol., Phys.*, 74(2):644–651.

Keereman, V., Fierens, Y., Broux, T., De Deene, Y., Lonneux, M., and Vandenberghe, S. (2010). MRI-based attenuation correction for PET/MRI using ultrashort echo time sequences. *J. Nucl. Med.*, 51:812–818.

Kemppainen, R., Suilamo, S., Tuokkola, T., Lindholm, P., Deppe, M. H., and Keyrilainen, J. (2017). Magnetic resonance-only simulation and dose calculation in external beam radiation therapy: a feasibility study for pelvic cancers. *Acta Oncologica*, 56(6):792–798.

Korhonen, J. (2015). Magnetic resonance imaging-based radiation therapy- methods enabling the radiation therapy treatment planning workflow for prostate cancer patients by relying solely on MRI-based images throughout the process. *Aalto University publication series*.

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(1).

Kuljus, K., Bayisa, F. L., Bolin, D., Johansson, A., Lember, J., and Yu, J. (2017). Comparison of hidden Markov chain models and hidden Markov random field models in estimation of computed tomography images. *arXiv:1705.01727*.

Maspero, M., Van den Berg, C. A. T., Landry, G., Belka, C., Parodi, K., Seevinck, P. R., Raaymakers, B. W., and Kurz, C. (2017). Feasibility of MR-only proton dose calculations for prostate cancer radiotherapy using a commercial pseudo-CT generation method. *Phys. Med. Biol.*, 62(24):9159–9176.

Mathews, J. D., Forsythe, A. V., Brady, Z., Butler, M. W., Goergen, S. K., Byrnes, G. B., Giles, G. G., Wallace, A. B., Anderson, P. R., Guiver, T. A., McGale, P., Cain, T. M., Dowty, J. G., Bickerstaffe, A. C., and Darby, S. C. (2013). Cancer risk in 680 000 people exposed to computed tomography scans in childhood or adolescence: data linkage study of 11 million australians. *BMJ*, 346:f2360.

Nie, D., Cao, X., Gao, Y., and Wang, L., S. D. (2016). Estimating CT image from MRI data using 3D fully convolutional networks. *Proceedings*, 10008 LNCS:170–178.

Powers, D. M. (2011). Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation.

Praveena, M. and Jaiganesh, V. (2017). A literature review on supervised machine learning algorithms and boosting process. *International Journal of Computer Applications*, 169(8):32–35.

Schmidt, M. A. and Payne, G. S. (2015). Radiotherapy planning using MRI. *Phys. Med. Biol.*, 60:R323–361.

Seiffert, C., Khoshgoftaar, T., Hulse, J. V., and Napolitano, A. (2010). Rusboost: A hybrid approach to alleviating class imbalance. *IEEE Trans. Syst., Man, Cybern.*

A, Syst., Humans, (1):185–197.

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. *Phys. Med.*, 42(10):6090–6097.

Suit, H. and du Bois W. (1991). The importance of optimal treatment planning in radiation therapy. *Int. J. Radiat. Oncol. Biol. Phys.*, 21:1471–1478.

Tanner, S. F., Finnigan, D. J., Khoo, V. S., Mayles, P., Dearnaley, D. P., and Leach, M. O. (2000). Radiotherapy planning of the pelvis using distortion corrected MR images: The removal of system distortions. *Phys. Med. Biol.*, 45:2117–2132.

Wiesinger, F., Sacolick, L. I., Menini, A., Kaushik, S. S., Ahn, S., Veit-Haibach, P., Delso, G., and Shanbhag, D. D. (2016). Zero TE MR bone imaging in the head. *Magnetic Resonance in Medicine*, 75:107–114.

Yu, C., Petrovich, Z., Apuzzo, M. L. J., and Luxton, G. (2001). An image fusion study of the geometric accuracy of magnetic resonance imaging with the leksell stereotactic localization system. *Applied Clinical Medical Physics*, 2(1):42–50.