



PAPER

Early prediction of tumour-response to radiotherapy in NSCLC patients

OPEN ACCESS

RECEIVED

19 March 2021

REVISED

23 September 2021

ACCEPTED FOR PUBLICATION

13 October 2021




PUBLISHED

5 November 2021

Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](https://creativecommons.org/licenses/by/4.0/).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.



Lameck Mbangula Amugongo^{1,2} , Eliana Vasquez Osorio^{1,2} , Andrew Green^{1,2} , David Cobben³, Marcel van Herk^{1,2} and Alan McWilliam^{1,2}

¹ Division of Cancer Sciences, University of Manchester, Manchester, United Kingdom

² Department of Radiotherapy Related Research, The Christie NHS Foundation Trust, Manchester, United Kingdom

³ The Clatterbridge Cancer Centre NHS Foundation Trust, United Kingdom

E-mail: lameckmbangula.amugongo@postgrad.manchester.ac.uk

Keywords: prediction, tumour changes, NSCLC, radiotherapy

Supplementary material for this article is available [online](#)

Abstract

Objective. In this study we developed an automatic method to predict tumour volume and shape in weeks 3 and 4 of radiotherapy (RT), using cone-beam computed tomography (CBCT) scans acquired up to week 2, allowing identification of large tumour changes. **Approach.** 240 non-small cell lung cancer (NSCLC) patients, treated with 55 Gy in 20 fractions, were collected. CBCTs were rigidly registered to the planning CT. Intensity values were extracted in each voxel of the planning target volume across all CBCT images from days 1, 2, 3, 7 and 14. For each patient and in each voxel, four regression models were fitted to voxel intensity; applying linear, Gaussian, quadratic and cubic methods. These models predicted the intensity value for each voxel in weeks 3 and 4, and the tumour volume found by thresholding. Each model was evaluated by computing the root mean square error in pixel value and structural similarity index metric (SSIM) for all patients. Finally, the sensitivity and specificity to predict a 30% change in volume were calculated for each model. **Main results.** The linear, Gaussian, quadratic and cubic models achieved a comparable similarity score, the average SSIM for all patients was 0.94, 0.94, 0.90, 0.83 in week 3, respectively. At week 3, a sensitivity of 84%, 53%, 90% and 88%, and specificity of 99%, 100%, 91% and 42% were observed for the linear, Gaussian, quadratic and cubic models respectively. Overall, the linear model performed best at predicting those patients that will benefit from RT adaptation. The linear model identified 21% and 23% of patients in our cohort with more than 30% tumour volume reduction to benefit from treatment adaptation in weeks 3 and 4 respectively. **Significance.** We have shown that it is feasible to predict the shape and volume of NSCLC tumours from routine CBCTs and effectively identify patients who will respond to treatment early.

1. Introduction

Non-small cell lung cancer (NSCLC) patients are a diverse population, with different characteristics, tumour sizes and tumour biology. For, inoperable NSCLC patients, radiotherapy (RT), chemotherapy or the combination are treatment options (Fried *et al* 2014, Postmus *et al* 2017). During RT of lung cancer, varying degree of tumour-response has been observed (Das *et al* 2010, Barker *et al* 2015, O'Connor *et al* 2015). Some tumours shrink during RT while others show no change in tumour size. Given this heterogeneity in tumour response, there is a clinical need to identify patients who are responding to RT as early in their treatment as possible.

Medical imaging is increasingly used as a non-invasive way to assess tumour response to treatment. Unlike, invasive techniques such as biopsies, imaging can provide spatial and temporal information about the tumour (Cook *et al* 2014). This information may be useful in improving RT by maximising tumour control and reducing

RT related toxicities. In recent years, novel treatment techniques such as adaptive radiotherapy (ART) have emerged in the treatment planning process (Guckenberger *et al* 2012). ART which was first described by Yan *et al* (1997), and aims to optimise RT by incorporating patient-level changes observed during RT in treatment plan modification. However, most previous studies have used computed tomography (CT) or positron emission tomography with CT to evaluate tumour regression for ART (Guckenberger *et al* 2011, Berkovic *et al* 2017, Roengvoraphoj *et al* 2018).

In lung RT, cone-beam computed tomography (CBCT) images are widely used to verify the patient's position before treatment. CBCT images can provide additional information about the tumour on the day of treatment. Previous studies have utilised CBCT images to monitor tumour response and demonstrated ART during the course of RT of lung cancer patients (Elsayad *et al* 2016, Møller *et al* 2016). Elsayad *et al* (2016) reported intrathoracic changes in 74% of patients occurring within the first three weeks of treatment. Adaptations were performed in 60% ($N = 43$) of the patients. However, in 25% of patients, new planning CT (pCT) for adaptation needed to be performed with urgency. In a larger study ($N = 177$) applying a strict adaptation criterion, Møller *et al* (2016) found that adaptation decreased lung dose in 75% of all the patients with ART. In both studies, adaptation was triggered by visually observed changes. Triggered adaptation is unpredictable and can constraint arrangements for resource allocation (re-scanning and re-planning). Contrarily, identifying patients who will benefit from ART early on during the course of RT will enable departments to plan, schedule and predict the resources that will need to be allocated.

This study posits that automatic analysis of CBCT images taken early in treatment (weeks 1 and 2) can be used to predict the shape and volume of the tumour in the later weeks (3 and 4) of RT so that adaptation can be scheduled timely.

2. Methods

2.1. Patients

In this retrospective study, 240 NSCLC patients who received 55 Gy in 20 fractions, with or without induction chemotherapy, treated at The Christie NHS Trust between 1 January 2013 and 31 October 2018 were collected. All images were collected with the Christie NHS Trust approval (UKCat, REC reference 17/NW/0060). Patient demographics are summarised in table 1 of supplementary material (available online at stacks.iop.org/PMB/66/225002/mmedia). All patients included met the following requirements; (1) CBCT images were available of the first three fractions, as well as CBCT images at fraction seven and fourteen, and (2) CBCT images were available in weeks 3 and 4. (3) CBCTs have sufficient image quality and no signs of non-tumour pulmonary changes such as atelectasis. A radiation oncologist reviewed all CBCTs identified to have insufficient quality and non-tumour pulmonary changes. These criteria resulted in a study cohort of 201 patients with a total of 1206 CBCT images. No other patient selection criteria were used.

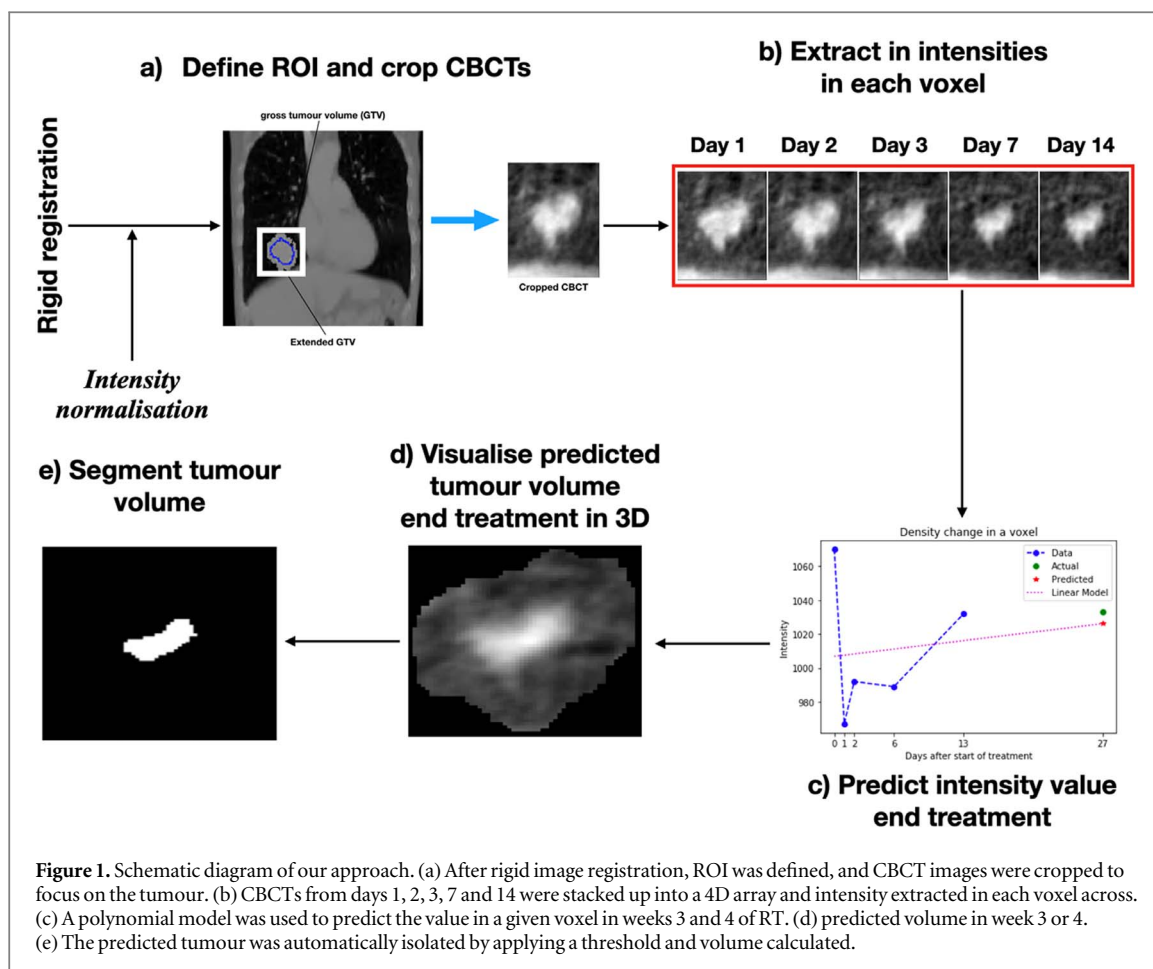
2.2. Image registration

The purpose of image registration is to find the appropriate transformation (T) that spatially maps corresponding pixels between pCT and subsequent CBCTs. Image registration was performed to automatically rigidly align follow up CBCT images to the pCT, using a two-step alignment process. The first step involved the rigid alignment of bony anatomy in 3D, correcting for patient rotations and translations along all three axes. This was achieved by creating a region of interest (ROI) including the vertebrae in all CT slices containing planning target volume (PTV). Next, registration was fine-tuned to align the soft tissue, using translations only. For this, a ROI containing the PTV was created, ensuring that the alignment of the soft tissue includes the gross tumour volume (GTV) and surrounding lung.

All registrations were visually checked for large errors. Image registration was performed using in-house software, World-match v9.0 (Wolthaus *et al* 2005), using the identical algorithm as used in XVI v5.6. No user interaction was involved at all.

2.3. Definition of region interest for pixel value analysis

All registered CBCT images were normalised to correct intensity information: to fix calibration differences between treatment machines and to allow comparison of CBCT between patients using the approach adapted from van Timmeren *et al* (2017), as described in our earlier work (Amugongo *et al* 2020). After, all CBCTs were cropped to focus on the tumour. A rectangular ROI was created using the GTV contour as defined on pCT, expanded by 1.3 cm, see figure 1(a)). This expansion is compatible with the extent of microscopic disease spread that rarely falls outside 1.3 cm from the radiographically defined GTV (Giraud *et al* 2000). In addition, an extension of 1.3 cm was used to include the immediately surrounding tissue in the prediction to allow enough surrounding tissue to be visible.



This part of the work was also carried out using World-match v9.0 (in-house software).

2.4. Prediction of density changes in each voxel

For all patients, intensity values were extracted in each voxel across all CBCT images from days 1, 2, 3, 7 and 14 (see figure 1(b)); these days correspond to the offline standard imaging protocol used in our institution at the time. In each voxel, four regression models were fitted, applying linear, quadratic and cubic polynomials, and a Gaussian process method. The Gaussian process is a generic supervised learning technique developed to solve regression and probabilistic classification problems (Rasmussen and Williams 2006). These models were used for extrapolation in the time domain and predict intensity values in each voxel for the CBCT image in week 3 (days 17–21) and 4 (days 24–28) of RT respectively, see figure 1(c)). The predicted intensity values for all voxels were visualised in 3D.

This part of the work was performed in python version 3.7, using polyfit regression, included in the NumPy package.

2.5. Tumour segmentation

Image segmentation by thresholding is a simple but powerful technique for segmenting images having bright intensities on dark background. In this study, a thresholding approach was used to segment the tumour volume. To get an approximate GTV segmentation in the predicted and the actual CBCT image acquired in weeks 3 and 4 of RT. Three threshold values were tested, and the threshold value of 374 HU provided the best segmentation results was chosen and applied. All pixel intensities above 374 HU were classified as tumour and included in the tumour segmentation. To guide the thresholding process, the GTV contour defined on the pCT expanded by 5 mm was used (see figure 1(d)), with thresholding performed inside this region only. This approach was preferred for its simplicity. The segmented tumour volumes were visually verified to exclude gross errors and exported as Neuroimaging Informatics Technology Initiative files. Sensitivity analysis was performed to see how changing the threshold value affects the segmented tumour volume. For this, segmentation threshold values of 281 and 468 HU were applied.

This part of the work was carried out in python version 3.7 using NumPy and NiBabel packages.

2.6. Model evaluation

For each model, the root mean square error (RMSE) was calculated in each voxel and summarised for each patient. RMSE measures the average magnitude of the errors between the predicted and actual measured values. Unlike other metrics, such as mean square error and mean absolute error, RMSE does not take the absolute value and gives high weight to large errors (Hyndman and Koehler 2006).

An independent metric, the structural similarity index metric (SSIM), was also used to assess visual similarities between the predicted and actual CBCT images acquired in weeks 3 and 4 of RT. Unlike RMSE which tries to estimate the perceived errors between the predicted and actual voxel intensities, SSIM is a perceptual metric that quantifies the similarities between two images by looking at the change in structural information; i.e. if voxel intensities in the two images have similar density values or line-up (Wang *et al* 2004). The range of SSIM is from -1 to 1 , and SSIM of 1 indicates that the two images (predicted and actual) to be identical and -1 means no similarity. SSIM was implemented as described by Wang *et al* (2004). Briefly, SSIM is used to compare the predicted and actual CBCT images. Therefore, quantify the visibility of errors (differences) between the predicted and actual CBCT images. SSIM and RMSE were calculated per model for all patients, comparing the observed actual CBCT images acquired in weeks 3 and 4 of RT, with the corresponding images generated by the models. The range was computed for each model.

For each patient, the tumour was segmented from the predicted and actual CBCT image acquired in week 3 and 4 of RT and tumour volume in cm^3 were computed. Next, the mean distance to agreement (DTA) in millimetres (mm) between the segmented tumour volume from predicted and actual CBCT images was calculated. In this study, we define DTA as the minimum distance from each point on the surface of the segmented predicted volume to the actual tumour volume (vice versa). We summarise these distances by calculating the mean and maximum values, following equations (1) and (2)

$$\text{mean DTA} = 0.5 * \left(\frac{\sum_i \text{dist}(p_i, T)}{M} + \frac{\sum_j \text{dist}(g_j, P)}{N} \right), \quad (1)$$

$$\text{max DTA} = \max(\max_i(\text{dist}(p_i, T)), \max_j(\text{dist}(g_j, P))), \quad (2)$$

where p and g correspond to points on the surface of the predicted and actual tumour, respectively; M and N are the number of points on the surface of the predicted and actual tumour, respectively; T and P the surface of the actual tumour and predicted volumes. Mean DTA represents the degree of mismatch between the two volumes, with a high value indicating the existence of regions of dissimilarity between the two sets, and a zero indicates that the volumes are identical. Unlike, the Hausdorff distance which measures the largest distance between two segmentations to characterise the worst matching region, and the Dice coefficient is dependent on volume. The mean DTA is the most appropriate metric as it measures the overall surface distance between the segmented predicted and actual tumour volume.

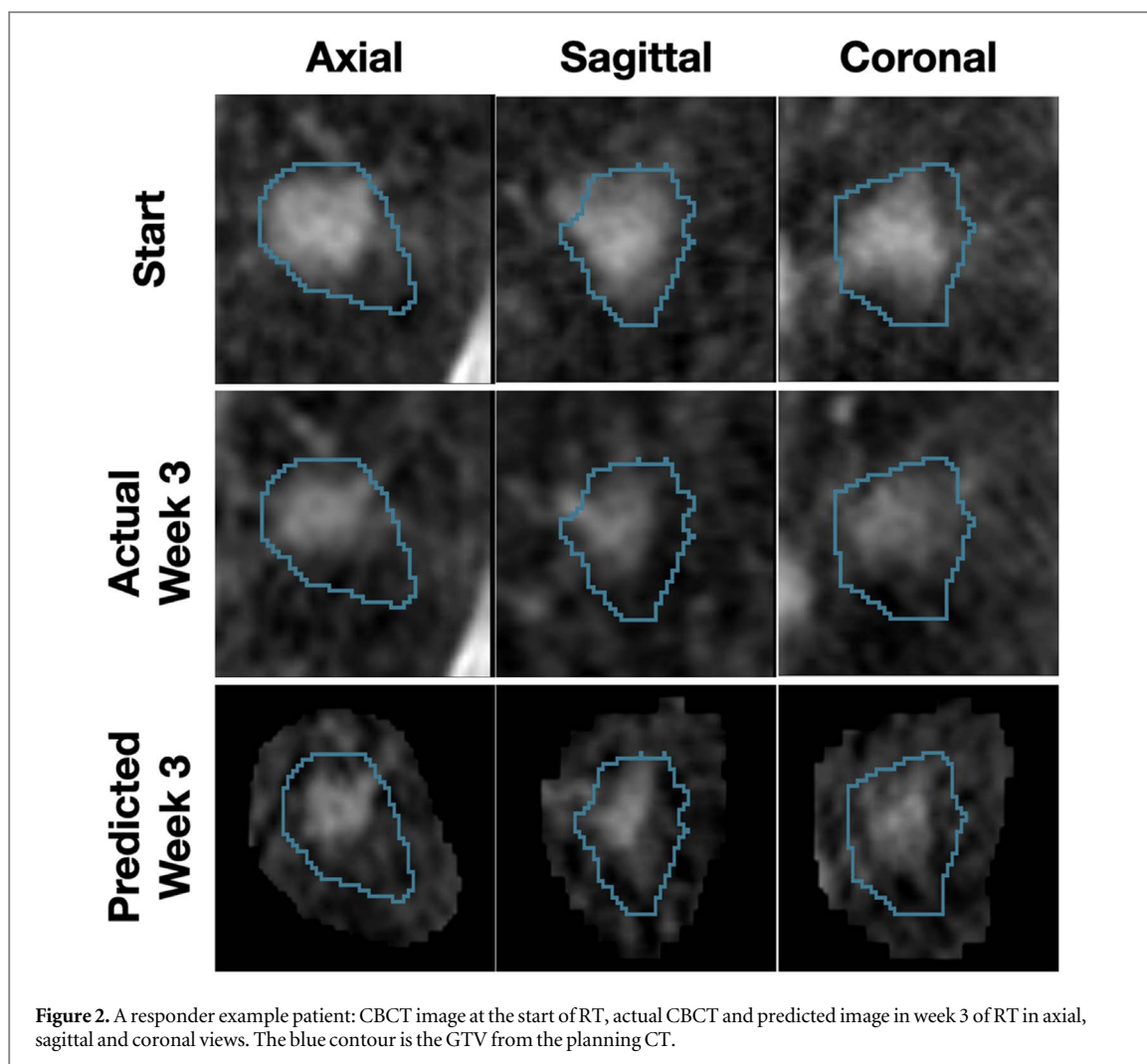
The mean DTA was calculated using World-match v9.0. The model evaluation was performed in python version 3.7, using sklearn and skimage packages respectively.

2.7. Tumour response classification

The actual and predicted percentage volume change at weeks 3 and 4 were computed for each patient for all models. A threshold of 30% was used to classify patients who will benefit from treatment adaptation in either week. The 30% reduction was derived from a clinical study that suggested that tumours with greater than 30% in tumour reduction are good candidates for ART (Woodford *et al* 2007). Receiver-operating characteristic analysis was performed to evaluate the overall performance of each model. We report the sensitivity (true positive rate), specificity (true negative rate), precision (ratio of correctly predicted positive cases over the total number of predicted positive cases), accuracy (ratio of correctly predicted cases to the total number of cases) and F1-measure (harmonic average of precision and sensitivity) in weeks 3 and 4 for each model. The formal definitions are provided in the supplementary material. The sensitivity is defined as the percentage of patients with a reduction of more than 30% that were correctly predicted to benefit from ART by the methods. The specificity represents the percentage of patients that were correctly predicted by the methods not benefit from ART. Additional experiments, using the threshold of 20% and 40% to patients who will benefit from ART are provided in the supplementary materials.

2.8. Rigid image registration uncertainties

When performing image registration, uncertainties are inevitable. These uncertainties can lead to errors in model predictions, depending on the nature of registration failure. The impact of rigid registration uncertainties was estimated by shifting each CBCT with a random offset sampled from a 3D Gaussian distribution ($\mu = 0$, $\sigma = 3$ mm). These parameters correspond to typically baseline shifts in lung cancer patients (Kanakavelu and Samuel 2016).



After these shifts, we calculated the volume, mean DTA, RMSE and SSIM. Simulations were repeated 100 times. Simulated results were compared to the initial prediction results using the original registration. Simulations were only performed for the predictions in week 3. This part of the work was carried out in python version 3.7, using the SimpleITK package.

3. Results

3.1. Image registration

Thirty-nine patients were excluded; three were because of unrelated pulmonary change (atelectasis), one because of inadequate image quality (structures not visible) and 25 patients were excluded because they did not have all the required CBCT images. Registration accuracy was visually assessed by examining whether the clip boxes were in the right positions and CBCTs well aligned to the pCT. Ten patients were excluded because of poor rigid image registration failures.

3.2. Prediction of density changes in each voxel

An example of a prediction in one voxel using a linear, quadratic, cubic and Gaussian model is shown in figure 1 of supplementary material. It indicates that the linear and Gaussian models performed better compared to the quadratic and cubic models. Figure 2 shows a patient CBCT series visualised in the axial, sagittal and coronal planes, predicted using the linear model.

It is evident that this patient displays visible tumour shrinkage on the CBCT acquired in week 3 of RT; a linear model predicts a visually similar image. Quantitatively, the GTV has shrunk from 12.85 cm^3 (tumour volume measured on pCT at the start of treatment) to 0.42 cm^3 (tumour volume on the CBCT at week 3). The predicted tumour volume in week 3 of RT for this patient was 0.32 , 1.52 , 0.91 and 2.71 cm^3 for the linear, Gaussian, quadratic and cubic models respectively.

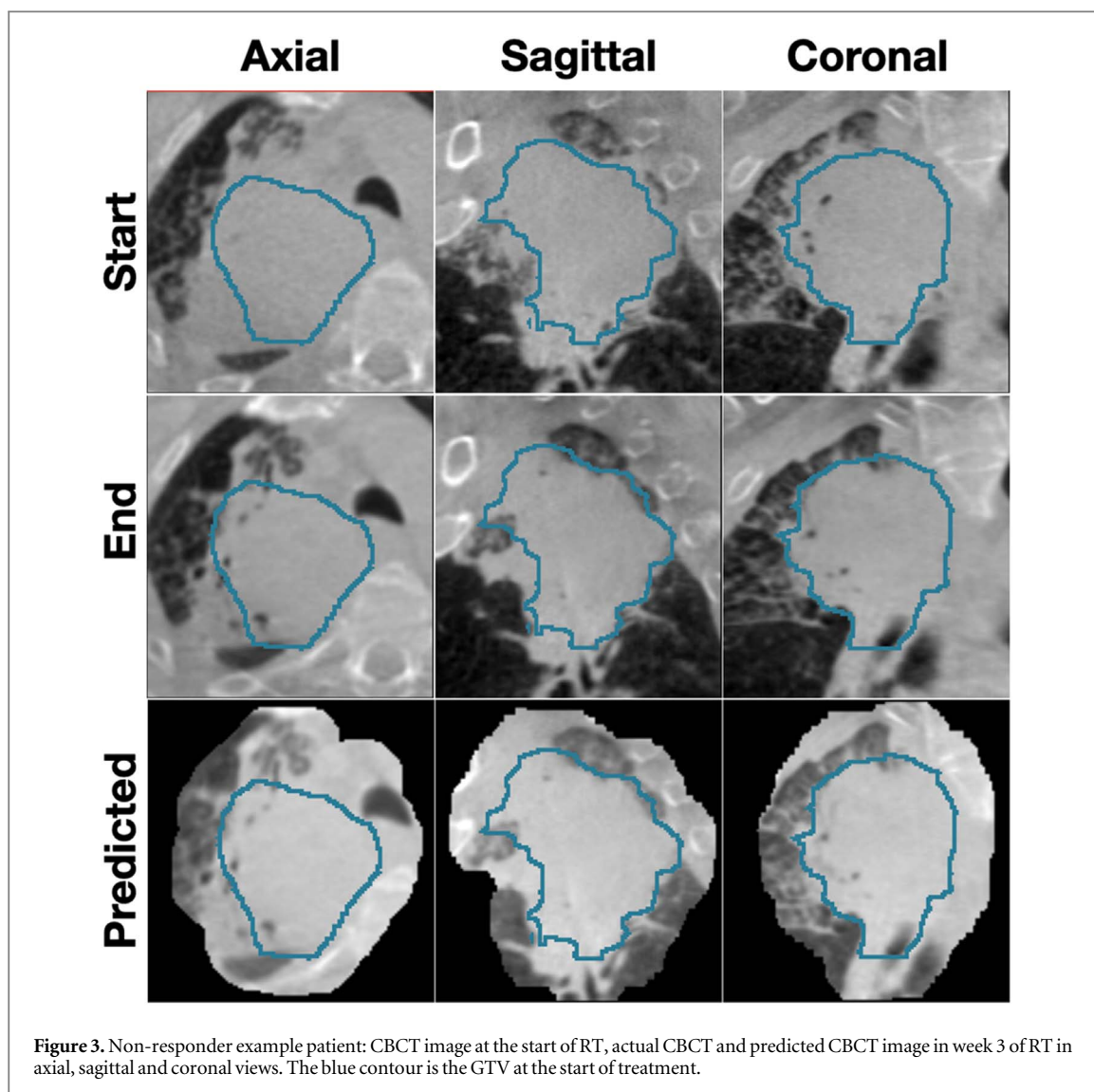


Figure 3. Non-responder example patient: CBCT image at the start of RT, actual CBCT and predicted CBCT image in week 3 of RT in axial, sagittal and coronal views. The blue contour is the GTV at the start of treatment.

Figure 3 shows an example case of a patient who with little tumour shrinkage. Both the predicted and actual image of the CBCT acquired in week 3 of RT displays little change. In this case, the tumour volume remains nearly the same throughout treatment, with a 4% actual reduction in tumour size. The volume change was -7.66 , -0.11 , -34.11 and -78.89 cm^3 , indicating a 4%, 0.1%, 18% and 41% decrease in tumour volume from the start of treatment as predicted by linear, Gaussian, quadratic and cubic models respectively.

3.3. Segmentation

After prediction, threshold-based segmentation was applied to identify the tumour. On visual inspection, gross segmentation errors were found in 3 cases and the majority of the patients (198 cases) had visually acceptable segmentations. Figure 4 shows an example of a segmented tumour volume for a representative patient.

3.4. Model validation

For all patients, the average RMSE score for the linear model is provided in figure 5. The average RMSE for the Gaussian, quadratic and cubic models is provided in figure 2 of supplementary material. The high order polynomial regression models performed worst. The linear model performed acceptably and was preferred. Figure 3 of supplementary material illustrates the performance of all models across all patients.

The SSIM for the linear model can be seen in figure 6. The SSIM for the Gaussian, quadratic and cubic models is shown in figure 4 of supplementary material. The average SSIM score for linear, Gaussian, quadratic and cubic models is 0.94, 0.94, 0.90 and 0.84 in week 3 of RT, respectively. In week 4, the average SSIM of 0.93, 0.92, 0.92 and 0.86 was observed for the linear, Gaussian, quadratic and cubic models respectively. These scores indicate most models predict tumours that look similar to the actual tumour, and the cubic model performed worst.

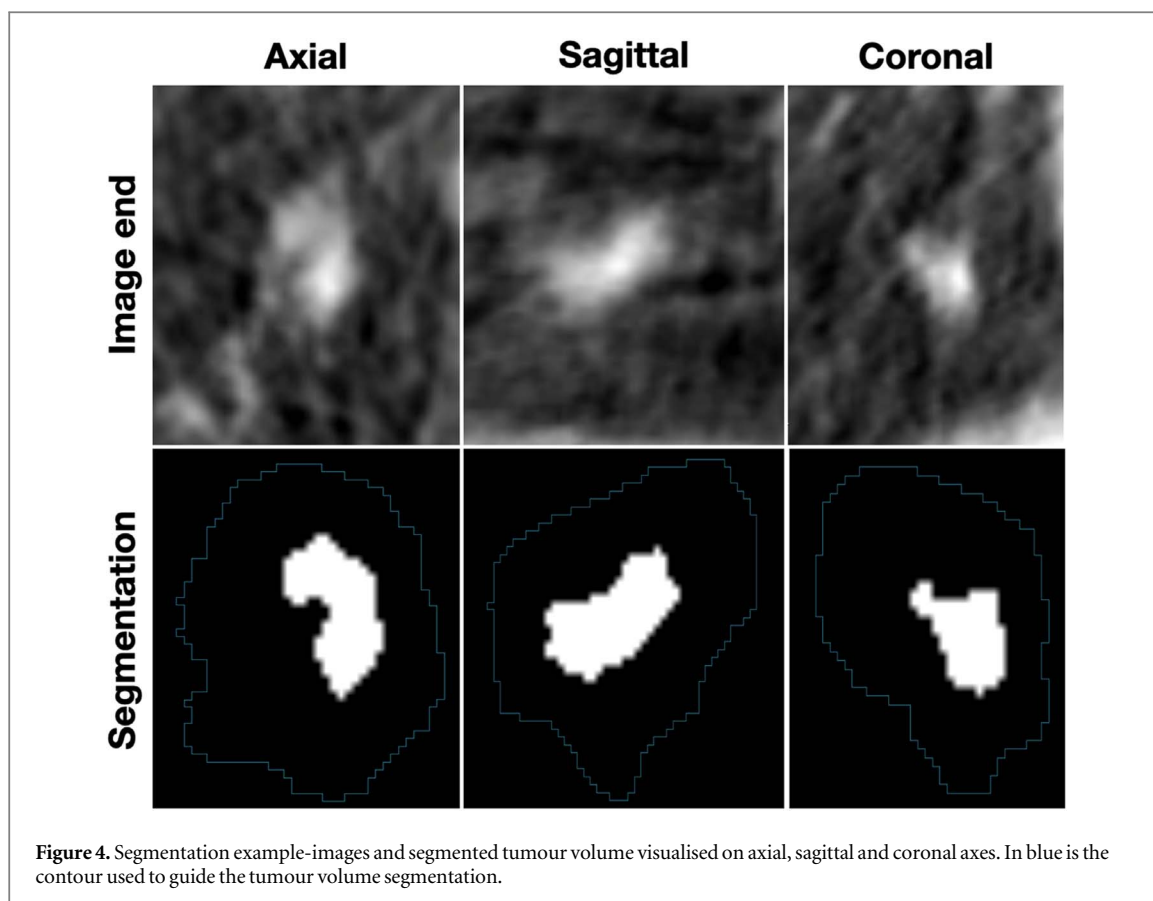


Figure 4. Segmentation example-images and segmented tumour volume visualised on axial, sagittal and coronal axes. In blue is the contour used to guide the tumour volume segmentation.

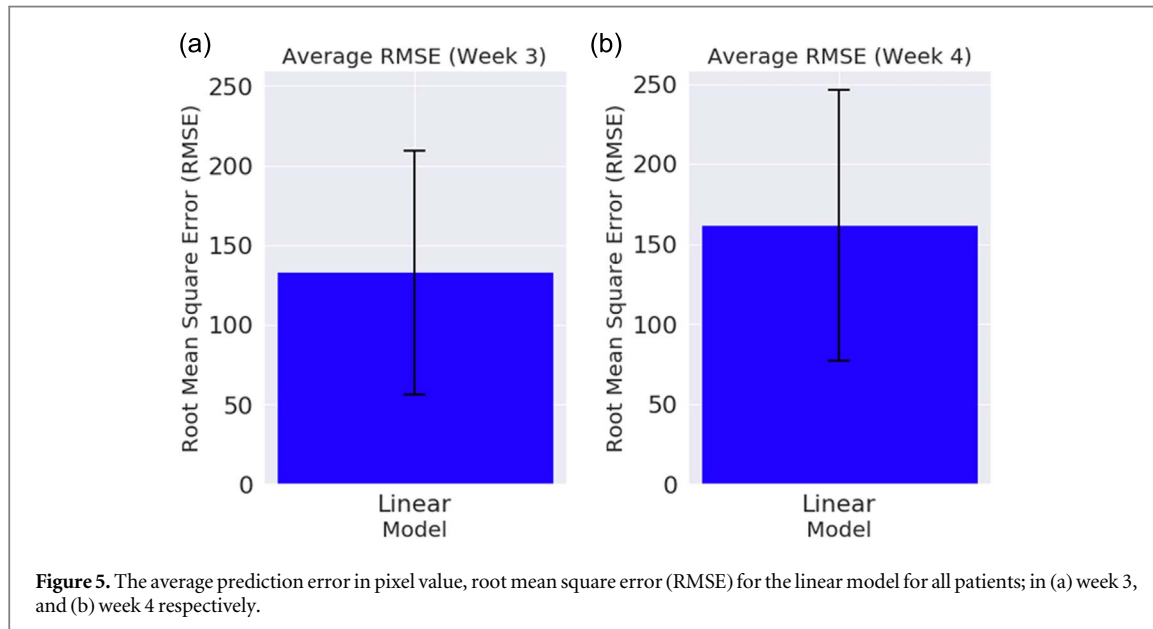


Figure 5. The average prediction error in pixel value, root mean square error (RMSE) for the linear model for all patients; in (a) week 3, and (b) week 4 respectively.

Figure 7 shows the correlation between the predicted tumour volume change for the linear model and the measured tumour volume change on the CBCT image acquired in week 3 of RT. The scatter plots for the Gaussian, quadratic and cubic models can be found in figure 5 of supplementary material. The best correlation between the predicted tumour volume change and the measured tumour volume change is obtained using a linear fit. The linear model predicted an average tumour volume of 52 cubic centimetres (cm^3) at week 3 compared to the measured tumour volume of 49 cubic cm^3 , an average range of 2 cm^3 . The ranges were 33.76 cm^3 , 58.88 cm^3 , 92.67 cm^3 and 171.50 cm^3 for the linear, Gaussian, quadratic and cubic models respectively. Scatterplots depicting the correlation between the predicted and measured tumour volume change in week 4 of RT for all models can be found in figure 6 of supplementary material. The Bland and Altman plots

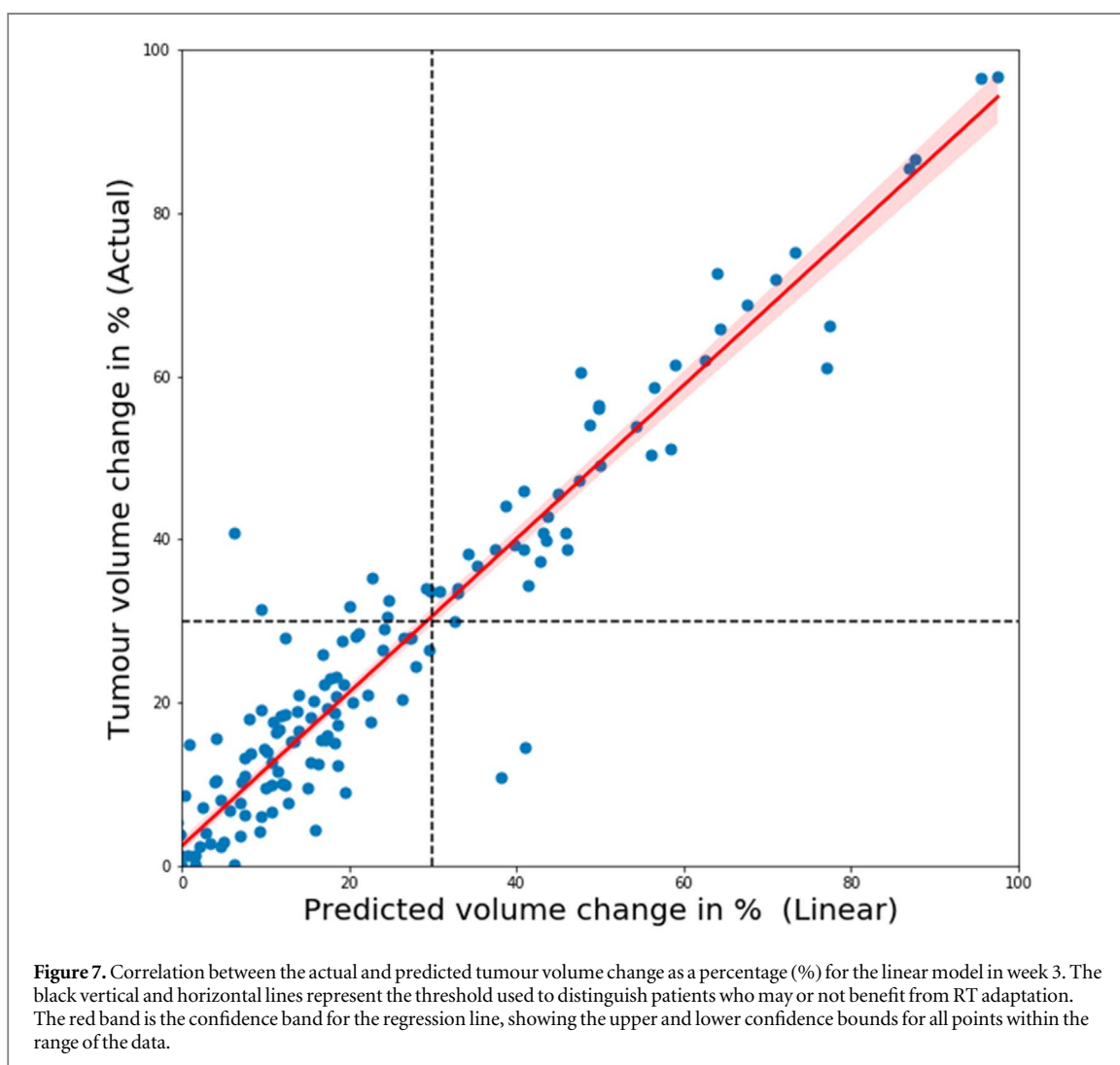
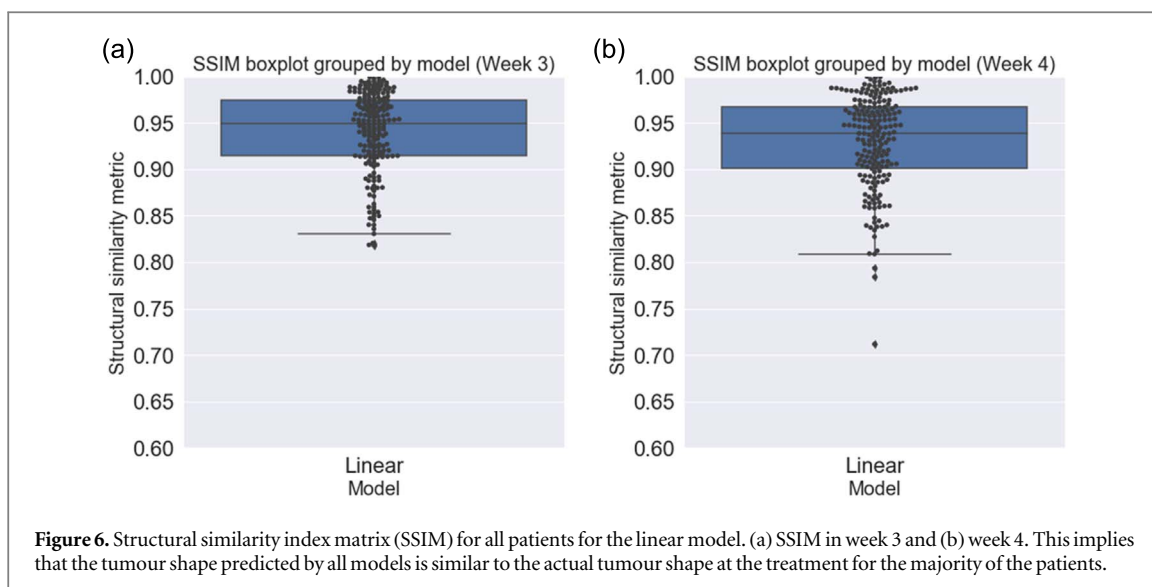


Table 1. The performance of the linear model at weeks 3 and 4. The threshold value for volume change and tumour response classification used was 0.3 (30%). The linear and Gaussian models perform better at identifying true positives and true negatives at week 3 than week 4.

	TP	FP	TN	FN	TPR	TNR	Precision	Accuracy	F-measure
Week 3	43	2	148	8	0.84	0.99	0.96	0.95	0.9
Week 4	46	9	135	11	0.81	0.94	0.84	0.9	0.82

Table 2. Mean, SD and maximum distance to agreement between the actual and predicted tumour volume for the initial prediction and after applying XYZ offset.

	Initial prediction	Prediction after applying random XYZ shifts, 100 simulations
	Linear	Linear
Mean DTA (mm)	0.55	0.60
SD (mm)	0.46	0.49
Max DTA (mm)	3.42	3.61

showing the percentage difference between the actual and predicted tumour volume in both weeks 3 and 4 are provided in figures 7 and 8 of the supplementary material.

3.5. Tumour response classification

In week 3, the sensitivity (true positive rate) for the linear model was 84%. Whereas, the specificity (true negative rate) was 99% for the linear model, see table 1. The sensitivity and specificity of the Gaussian, quadratic and cubic models are shown in table 2 of supplementary material. Overall, the linear fit performed best at identifying patients who will benefit from treatment adaption in week 3, while the quadratic model performed acceptably and the cubic model performed worst.

TP true positive, FP false positive, TN true negative, FN false negative, TPR true positive rate, and TNR true negative rate.

Table 1 present the sensitivity, specificity, accuracy, precision and *F*-measure for the linear model in week 4 of RT. The sensitivity for the Gaussian, quadratic and cubic models is provided in table 3 of supplementary material. Again, higher accuracy was observed using the linear fit compared to the other models.

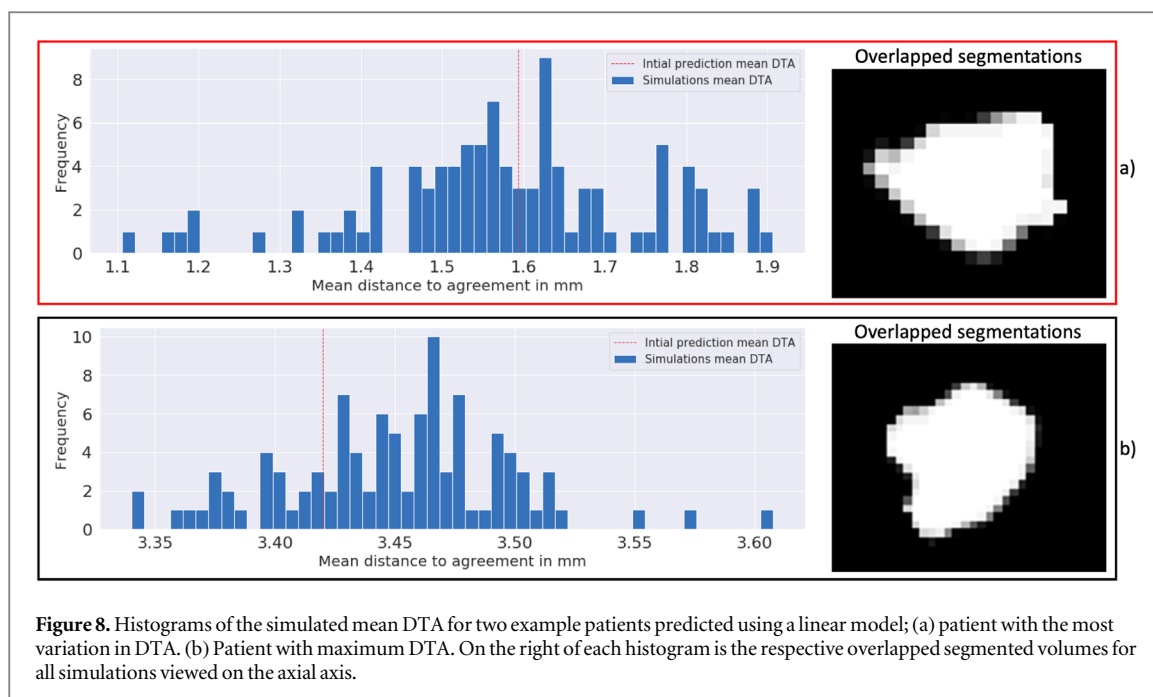
Compared to week 3, a lower sensitivity was observed in the gaussian model. This is a result of overestimation by the Gaussian model in week 4. In week 4 of RT, the false positive rate of 38% and 78% was observed for the quadratic and cubic models, respectively. The linear fit identified 21% and 23% of the patients to benefit from RT adaptation in weeks 3 and 4, respectively. Additional experiments, using the threshold of 20% and 40% to identify good responders for ART are provided in the supplementary materials, see figure 9. As expected, with a higher threshold, the number of patients eligible for ART decreases. The sensitivity analysis results are provided in tables 4 and 5 of supplementary material. As seen the higher-order polynomial models performed worse compared to the linear model.

3.6. Rigid image registration uncertainties

Table 2 shows the mean DTA, Standard deviation and max DTA between predicted and observed tumour segmentation for all patients; for initial prediction and after applying random XYZ shifts (100 simulations) using the linear model. The registration uncertainties for the Gaussian, quadratic and cubic models is illustrated in table 6 of the supplementary material. The results show that the predicted tumour volume was within 1mm in the majority of the patients. Plausible registration inaccuracies have only a very small impact on the results. The registration uncertainties for the Gaussian, quadratic and cubic models is illustrated in table 6 of the supplementary material.

DTA distance to agreement, GP gaussian process, mm millimetres, SD standard deviation.

The mean DTA for all simulations for two example patients are shown in figure 8; (a) a patient most affected by registration uncertainties and (b) the patient with the maximum DTA. The variations in mean DTA for each patient are also visible on the overlapped segmentation on the left. The mean DTA for all patients is shown in figure 10 of supplementary material. The standard deviation of the mean DTA for all patients is shown in figure 11 of the supplementary material. The mean and standard deviation of the RMSE and SSIM for all simulations is summarised in table 7 of supplementary material.



4. Discussion

To the best of our knowledge, this is the first study that has quantitatively predicted NSCLC tumour volume and shape at later weeks during RT using longitudinal CBCT images acquired up to mid-treatment of RT to proactively inform ART early. Our results show that individualised regression models built per-voxel using the intensity values from on-treatment CBCTs in the first two weeks of RT can predict tumour shape and volume in weeks 3 and 4 of RT, i.e. one and two weeks ahead.

Several studies have retrospectively investigated the need for ART in lung cancer patients due to tumour volume changes (Siker *et al* 2006, Britton *et al* 2007, Woodford *et al* 2007, Fox *et al* 2009) or intrathoracic changes (Kwint *et al* 2014, Møller *et al* 2016). In a cohort of 13 patients, Guckenberger *et al* (2011) illustrated that performing ART twice (week 3 or 5) during RT does not underdose the microscopic spread of the disease and can reduce the dose to the lung by 5%–8%. Another study by Guckenberger, Kavanagh and Partridge (2012) concluded that ART is safe and can improve tumour control probability. However, for ART to be effective sufficient regression has to be observed early on during treatment (Sonke and Belderbos 2010). Most of the studies in the literature, reported atelectasis making up a large percentage of all adaptations, 10% in Persoon *et al* (2015) and 12% in Møller *et al* (2014). In this study, patients with atelectasis were excluded, since we focused on tumour regression. Woodford *et al* (2007) found that 40% of patients in their cohort had a sufficient magnitude of tumour reduction to benefit from ART. Monitoring anatomical changes on daily CBCTs, Møller *et al* (2016) reported a false positive rate of 20% in patients requiring adaptation. In our study, the linear model achieved a false positive rate of 16% and 19% at weeks 3 and 4 respectively. However, inspecting figure 7(a) shows that only two patients were wrongly predicted to benefit from ART using the linear model.

Different studies have suggested the benefits of ART based on simulations. However, only a few studies have shown the clinical outcome of ART implementation. Tvilum *et al* (2015) demonstrated that ART can improve loco-regional control in a small cohort of 52 patients. The first study to report the clinical outcome of ART, the LARTIA trial reported marginal relapse in 6% of the patients and low toxicity (2% and 4% acute and lung damage) (Ramella *et al* 2017). However, in both cases the decisions to adapt are qualitative, for example, anatomical changes judged by radiation oncologists. Kwint *et al* (2014) proposed a semi-quantitative approach to evaluate intrathoracic changes. However, they did not include tumour volume changes in their criteria. Accounting for tumour position, lymph nodes, organs at risk and dosimetric impact, Møller *et al* (2016) described stricter adaptation criteria.

In the majority of studies that have implemented ART strategies, adaptation is triggered by the changes deemed detrimental as visually observed during RT. The unpredictable nature of triggered adaptation can constraint resource allocation because arrangements for re-scanning or re-planning is only made once the changes are detected (Sonke *et al* 2019). In busy departments, scheduling a re-scan and treatment re-plan may take several days (up to 5 days). Our findings demonstrate that we can predict tumour changes early during RT, suggesting that it is possible to identify patients who may benefit from ART early on. Early identification of

patients that benefit from ART will enable departments to allocate the required resources ahead of time. Our approach is not intended to replace radiation oncologists but to improve the ART clinical workflow by removing the manual inspection step in the selection of patients eligible for ART. The manual step is error-prone and subjective.

In our study, the linear model correctly predicted that 21% and 23% of patients may benefit from ART in weeks 3 and 4, respectively. Accepting the false positive rate would lead to 2 and 9 patients being re-planned unnecessarily at weeks (3 and 4), since their tumour shrinkage was wrongly predicted to be over 30%, while in reality, it ranged from 10.8 to 14.4% and 9.6%–29.1%, respectively. One patient had actual tumour regression of 29.1% and the linear model predicted a 34.4% reduction. Though the actual regression is not more than 30%, the model prediction was close. The 30% threshold used in this study is not intended to be a concrete principle but is useful as a criterion to identify or ‘flag’ patients for adaptive planning evaluation.

Although the results of this study are promising, limitations exist. First, the quality of CBCT images is poor compared to conventional CT. Intensity correction was applied according to the method described by van Timmeren *et al* (2017) to correct the distribution of intensities between CBCT images. However, image noise and artefacts remain a challenge; standard 3D CBCT images acquired in free-breathing were used in this study. Possible solutions to this have been proposed in the literature by optimising scatter correction (Poludniowski *et al* 2009) or performing motion compensation (Rit *et al* 2009).

Second, prediction accuracy could have been affected by the biases in the model, such as overfitting. To overcome this challenge, four regression models applying a linear, Gaussian, quadratic and cubic fit were used and their prediction results were compared. Large errors were observed in the higher-order (quadratic and cubic) models, and this can be a result of ill-conditioning, which causes the coefficients to be sensitive to small variations in the data. Lung cancer patients are now imaged daily. More information provided by daily CBCTs can potentially improve prediction accuracy, especially for the higher-order regression models.

In this study, a threshold value of 374 HU was used. However, the sensitivity analysis applying two different threshold values showed that the segmentation did not deteriorate the results, see tables 4 and 5 of supplementary material. The linear model still performed better at identifying patients with large tumour regression to benefit from ART. Different institutions should explore and determine the appropriate threshold values or perhaps apply an individual threshold value for each patient. Another limitation is that threshold selection and final evaluation of the model was performed on the same dataset. In addition, our segmentation approach uses an expanded region to guide the thresholding process. Though this is sensible for tumours surrounded by lung tissue, this can lead to tumour volume overestimation for tumours adjacent to the chest wall or mediastinal because normal surrounding tissue with similar density may be included in tumour volume. Consequently, underestimating relative volume reduction for these tumours. To reduce the impact of overestimation of tumour volume, the lung contour was then used to ensure the expanded region did not extend into other tissue outside the lung, i.e. mediastinum, diaphragm or chest wall. Also, the automatic segmentation used in this study is limited by qualitative assessment based on visual inspection only. Thus, further improvement in tumour volume segmentation may improve predictions.

Another issue that could have affected our results is image registration. In this study rigid image registration was used, rigid image registration cannot deal with complex deformations, for example, large anatomical changes, including weight loss and normal tissue response. To estimate the impact of image registration uncertainties, we simulated the prediction 100 times, applying a random 3D translation. We found that the average mean DTA was off by less than 3 mm in all patients for all four models implying that image registration uncertainties did not affect the results in the majority of the patients. Lastly, the lack of an external validation dataset can limit the clinical relevance of this study.

Because we are not extracting features of the segmented tumour on subsequent CBCT images, our approach maintains the spatial and temporal information of the tumour. The spatial information is important because tumours do not change homogeneously, different regions of the tumour are expected to respond differently to a treatment. Also, our approach applies fits in each voxel individually, thus can be applied to any patient. Clinically, our model will be useful in identifying tumours that are likely to change during treatment early. As part of a clinical workflow, a CT scan can be scheduled proactively for patients identified as good responders. Early identification of patients with large tumour regression will give radiation oncologists sufficient time to determine the safety of ART for each patient. However, to evaluate the safety of treatment adaptation, more information about changes around the surrounding tissues will be needed to distinguish the mode of tumour change, i.e. elastic and inelastic change. The former is favourable for the adaptation of treatment volumes, whereas the latter implies that treatment volumes should remain unchanged, at least for an intermediate dose level. A boost on the shrunken tumour may be more appropriate. Future work will explore complementary techniques such as deformable image registration to investigate changes happening to the surrounding tissue, therefore distinguish modes of tumour change. In addition, future work will also explore associating early regression with clinical data and explore the possibility to predict tumour control and patient’s outcome early.

We have shown that individualised regression models built per voxel on intensity values from on-treatment CBCT images can predict tumour volume and shape in weeks 3 and 4 of RT. If we can identify tumours responding to treatment early, patients that will benefit from plan adaption can be identified early.

5. Conclusion and future works

In this study, we have shown that it is possible to predict the tumour volume and shape in weeks 3 and 4 of RT, using intensity values extracted per-voxel longitudinally across CBCT images acquired in the first half of RT treatment. Image registration uncertainties did not impact the prediction accuracy. Finally, the linear model achieved good results at predicting the tumour volume and identifying patients who will or not benefit from RT adaptation early on during the course of RT.

Acknowledgments

Marcel van Herk, Andrew Green and Eliana Vasquez Osorio are supported by NIHR Manchester Biomedical Research Centre. Lameck Mbangula Amugongo is supported by Newton Fund as part of the Development in Africa with Radio Astronomy (DARA) Big Data project.

ORCID iDs

Lameck Mbangula Amugongo  <https://orcid.org/0000-0001-6468-2643>

Eliana Vasquez Osorio  <https://orcid.org/0000-0003-0741-994X>

Andrew Green  <https://orcid.org/0000-0002-8297-0953>

References

- Amugongo L M *et al* 2020 Identification of patterns of tumour change measured on CBCT images in NSCLC patients during radiotherapy *Phys. Med. Biol.* **65** 215001
- Barker H E *et al* 2015 The tumour microenvironment after radiotherapy: mechanisms of resistance and recurrence *Nat. Rev. Cancer* **15** 409–25
- Berkovic P *et al* 2017 Adaptive radiotherapy for locally advanced non-small cell lung cancer: dosimetric gain and treatment outcome prediction *Acta Oncol.* **56** 1656–9
- Britton K R *et al* 2007 Assessment of gross tumor volume regression and motion changes during radiotherapy for non-small-cell lung cancer as measured by four-dimensional computed tomography *Int. J. Radiat. Oncol. Biol. Phys.* **68** 1036–46
- Cook G J R *et al* 2014 Radiomics in PET: principles and applications *Clin. Transl. Imaging* **2** 269–76
- Das A K *et al* 2010 Radiogenomics predicting tumor responses to radiotherapy in lung cancer *Semin. Radiat. Oncol.* **20** 149–55
- Elsayad K *et al* 2016 Cone-beam CT-guided radiotherapy in the management of lung cancer *Strahlenther. Onkol.* **192** 83–91
- Fox J *et al* 2009 Quantification of tumor volume changes during radiotherapy for non-small-cell lung cancer *Int. J. Radiat. Oncol. Biol. Phys.* **74** 341–8
- Fried D V *et al* 2014 Prognostic value and reproducibility of pretreatment CT texture features in stage III non-small cell lung cancer *Int. J. Radiat. Oncol. Biol. Phys.* **90** 834–42
- Giraud P *et al* 2000 Evaluation of microscopic tumor extension in non-small-cell lung cancer for three-dimensional conformal radiotherapy planning *Int. J. Radiat. Oncol. Biol. Phys.* **48** 1015–24
- Guckenberger M *et al* 2011 Adaptive radiotherapy for locally advanced non-small-cell lung cancer does not underdose the microscopic disease and has the potential to increase tumor control *Int. J. Radiat. Oncol. Biol. Phys.* **81** e275–82
- Guckenberger M *et al* 2011 Potential of adaptive radiotherapy to escalate the radiation dose in combined radiochemotherapy for locally advanced non-small cell lung cancer *Int. J. Radiat. Oncol. Biol. Phys.* **79** 901–8
- Guckenberger M, Kavanagh A and Partridge M 2012 Combining advanced radiotherapy technologies to maximize safety and tumor control probability in stage III non-small cell lung cancer *Strahlenther. Onkol.* **188** 894–900
- Hyndman R J and Koehler A B 2006 Another look at measures of forecast accuracy *Int. J. Forecast.* **22** 679–88
- Kanakavelu N and Samuel E J J 2016 Accuracy in automatic image registration between MV cone beam computed tomography and planning kV computed tomography in image guided radiotherapy *Rep. Pract. Oncol. Radiother.* **21** 487–94
- Kwint M *et al* 2014 Intra thoracic anatomical changes in lung cancer patients during the course of radiotherapy *Radiother. Oncol.* **113** 392–7
- Møller D S *et al* 2014 Adaptive radiotherapy of lung cancer patients with pleural effusion or atelectasis *Radiother. Oncol.* **110** 517–22
- Møller D S *et al* 2016 Adaptive radiotherapy for advanced lung cancer ensures target coverage and decreases lung dose *Radiother. Oncol.* **121** 32–8
- O'Connor J P B *et al* 2015 Imaging intratumor heterogeneity: role in therapy response, resistance, and clinical outcome *Clin. Cancer Res.* **21** 249–57
- Persoon L C G G *et al* 2015 Is integrated transit planar portal dosimetry able to detect geometric changes in lung cancer patients treated with volumetric modulated arc therapy? *Acta Oncol.* **54** 1501–7
- Poludniowski G *et al* 2009 An efficient Monte Carlo-based algorithm for scatter correction in keV cone-beam CT *Phys. Med. Biol.* **54** 3847–64
- Postmus P E *et al* 2017 Early and locally advanced non-small-cell lung cancer (NSCLC): ESMO clinical practice guidelines for diagnosis, treatment and follow-up *Ann. Oncol.* **28** iv1–iv21

- Ramella S *et al* 2017 Local control and toxicity of adaptive radiotherapy using weekly CT imaging: results from the LARTIA trial in stage III NSCLC *J. Thoracic Oncol.* **12** 1122–30
- Rasmussen C E and Williams C K I 2006 *Gaussian Processes for Machine Learning* (Cambridge, MA: MIT Press) (<http://gaussianprocess.org/gpml/>) (Accessed: 16 July 2021)
- Rit S *et al* 2009 On-the-fly motion-compensated cone-beam CT using an *a priori* model of the respiratory motion *Med. Phys.* **36** 2283–96
- Roengvoraphoj O *et al* 2018 Analysis of primary tumor metabolic volume during chemoradiotherapy in locally advanced non-small cell lung cancer *Strahlenther. Onkol.* **194** 107–15
- Siker M L, Tomé W A and Mehta M P 2006 Tumor volume changes on serial imaging with megavoltage CT for non-small-cell lung cancer during intensity-modulated radiotherapy: how reliable, consistent, and meaningful is the effect? *Int. J. Radiat. Oncol. Biol. Phys.* **66** 135–41
- Sonke J-J, Aznar M and Rasch C 2019 Adaptive radiotherapy for anatomical changes *Semin. Radiat. Oncol.* **29** 245–57
- Sonke J-J and Belderbos J 2010 Adaptive radiotherapy for lung cancer *Semin. Radiat. Oncol.* **20** 94–106
- Tvilum M *et al* 2015 Clinical outcome of image-guided adaptive radiotherapy in the treatment of lung cancer patients *Acta Oncol.* **54** 1430–7
- van Timmeren J E *et al* 2017 Survival prediction of non-small cell lung cancer patients using radiomics analyses of cone-beam CT images *Radiother. Oncol.* **123** 363–9
- Wang Z *et al* 2004 Image quality assessment: from error visibility to structural similarity *IEEE Trans. Image Process.* **13** 600–12
- Wolthaus J W H *et al* 2005 Fusion of respiration-correlated PET and CT scans: correlated lung tumour motion in anatomical and functional scans *Phys. Med. Biol.* **50** 1569–83
- Woodford C *et al* 2007 Adaptive radiotherapy planning on decreasing gross tumor volumes as seen on megavoltage computed tomography images *Int. J. Radiat. Oncol. Biol. Phys.* **69** 1316–22
- Yan D *et al* 1997 Adaptive radiation therapy *Phys. Med. Biol.* **42** 123–32