

Image-based data mining to analyse shrinkage versus erosion of lung cancer tumours treated with radiotherapy

Rhydian Windsor

Performed in collaboration with George Needham.

Abstract

Lung cancer tumours are believed to regress in two modes during radiotherapy; by shrinking elastically and by fragmenting. A method is proposed applying image-based data mining techniques to cone-beam computed tomography (CBCT) scans in order to determine which mode of recession is occurring during a course of radiotherapy. This method then is applied to 8 non-small cell lung cancer (NSCLC) patients. The proposed method is compared to an alternative, manual method of classifying the mode of recession using the movement of neighbouring tissues. No clear separation between the two modes of tumour recession is observed. Finally, several suggestions are made to improve the proposed methodology in order to distinguish between the two modes.

Keywords

Image-Guided Radiotherapy, Computer Vision, NSCLC, Tumour Recession, Automated Delineation



CONTENTS

1	Introduction and Previous Work	2
2	Background Theory	4
2.1	Lung Cancer	4
2.2	Radiotherapy	4
2.3	Computed Tomography Scanning	5
3	Method	5
3.1	Data Set	6
3.2	Mask Generation	7
3.2.1	Thresholding and removing other dense tissues	7
3.2.2	Filling cavities and removing small bodies	8
3.2.3	Edge detection and blur	9
3.3	Superposing the mask onto CBCT scans	12

4	Analysis and Results	12
4.1	Analysing voxel information	12
4.2	Comparison of peaks corresponding to healthy and cancerous tissue	14
4.3	Overlaid markers method	15
4.4	Comparing height of cancerous tissue peak in annulus to equivalent peak for the whole tumour	16
5	Uncertainties and Potential Further Research	17
5.1	Sources of Uncertainty	17
5.2	Areas Highlighted for Further Research	18
6	Conclusion	18
7	Acknowledgements	19
	References	19

1 INTRODUCTION AND PREVIOUS WORK

50% of all people born in the UK after 1960 will be diagnosed with some form of cancer during their lifetime [1]. 13% of these cases will be lung cancer [2], making it the third most common cancer type after breast (15%) and prostate (13%) [3]. It is also the most fatal, accounting for 22% of cancer deaths in the UK in 2014 [4]. Therefore developing effective and efficient treatment for this disease is an important goal of medicine today.

Image-guided radiotherapy provides us with one such method of treatment. In this technique, 3-dimensional X-ray computed tomography scans are used to generate images of tumours and surrounding regions during a course of radiotherapy. This allows clinicians to monitor anatomical changes and adapt the applied radiation dose accordingly. Of particular concern is the *toxicity of treatment*; the Encyclopedia of Radiation Oncology defines this as

“...side effects that occur on treatment or in the immediate post treatment period.”

[5]

Studies have linked lung radiotherapy to a wide range of these toxicities, including pneumonitis (inflammation of lung tissue), pulmonary fibrosis (scarring of lung tissue) [6], [7], [8] and recently cardiac disease [9]. The effect is particularly pronounced in elderly patients, where combined chemotherapy and radiotherapy can lead to severe toxicity in almost half of all patients [10].

It is thus a key challenge to minimise the radiation applied to healthy tissue whilst ensuring cancerous tissue is sufficiently irradiated. This can be effectively achieved by monitoring how tumours change during radiotherapy. For example, if a tumour is shrinking, the area irradiated during radiotherapy could be reduced or altered to account for tumour changes.

To explore this, several studies have been conducted into how the volume of tumours shrink ([11], [12], [13], [14], [15]) and how their *baseline position* (the average position a of tumour

during a respiratory cycle) moves during radiotherapy ([11], [13], [14], [15]). However, most current work in this area assumes that tumours shrink elastically rather than fragmenting when radiation is applied, as shown in Figure 1. This is a potentially dangerous assumption as reducing the irradiated area to the main body of a fragmented tumour could lead to fragments being left untreated. These could then be responsible for a relapse if given time to grow.

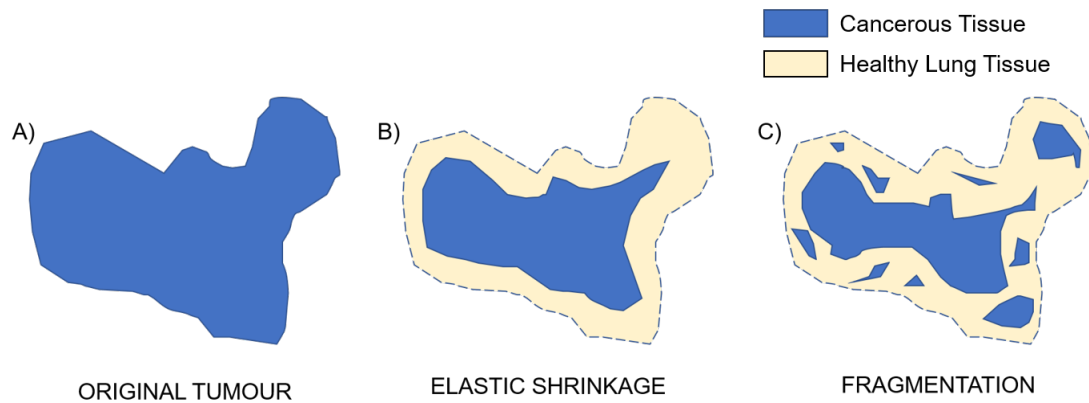


Figure 1. A diagram demonstrating the difference between elastically shrinking and fragmenting tumours. A shows the tumour before radiotherapy. B is an elastically shrunk version of A, whereas C is a fragmented version.

Despite this, there has been very little research into the phenomena of lung cancer shrinkage versus fragmentation. In 2017, Hamming-Vrieze et al. attempted to observe whether head and neck tumours were shrinking elastically or dissolving using the changes in the surrounding tissue and observed several markers which could be indicative of differing modes of tumour recession [16]. Here we propose another method of determining the modality of shrinkage, by measuring the tissue composition at the tumour borders. This is done using image-processing techniques on computed tomography scans obtained throughout radiotherapy courses for several patients. This is advantageous to the method used by Hamming-Vrieze et al. as it only requires one manual input, a delineation of the tumour volume and is otherwise completely automated. Thus the method can be applied to scan series for large numbers of patients rapidly.

The structure of this report is as follows. Section 2 covers some essential background theory related to the project. No specialist knowledge in radiotherapy is assumed; lung cancer, radiotherapy and computed tomography are all covered in sufficient depth to justify the methodology. Section 3 discusses steps taken to extract voxel information from the edges of the tumour. This also includes relatively comprehensive summaries of the algorithms used to achieve this. In section 4 the methods used to analyse the voxel information are outlined and the results for the 8 NSCLC patients included in this study are presented. Potential weaknesses of the methodology are covered in section 5 and possible ways of improving it are discussed. Finally, section 6 summarizes the key outcomes from this study.

2 BACKGROUND THEORY

2.1 Lung Cancer

Cancer is a disease where a genetic mutation causes abnormally rapid cell growth which can spread across the body. In their seminal paper published in 2000, “The Hallmarks of Cancer” [17], Hanahan and Weinberg defined six traits indicative of cancer cells:

- 1) Cancer cells generate their own growth.
- 2) Cancer cells are unaffected by signals from the body to stop their growth.
- 3) Normal cell death does not occur in cancer cells.
- 4) There is no upper limit on the number of cancer cell divisions.
- 5) Cancer cells stimulate blood vessels to grow such that they supply nutrients to tumours.
- 6) Cancer cells invade local tissue and spread across the body.

Lung cancer refers to any tumour in the lungs which possesses these traits. There are two main types: *small-cell lung carcinoma* (SCLC) and *non-small-cell lung carcinoma* (NSCLC) [18]. SCLCs account for 12% of all cases and are generally treated using chemotherapy. Cells of this variant are so named because they appear small under a microscope. In this experiment we examine NSCLC patients. This is an all-encompassing term for lung cancers which are not SCLCs and thus there is great variation in the class. However, most NSCLCs belong to one of three main variations; lung adenocarcinoma (38.5% of all NSCLC cases), squamous cell lung carcinoma (20%) and large cell carcinoma (2.9%) [19]. Whilst an in-depth pathology of the individual types is outside the scope of this report, it is important to note that NSCLCs tend to be less receptive to chemotherapy than SCLCs and thus tend to be treated by a combination of radiotherapy and chemotherapy. [20]

2.2 Radiotherapy

Treatment of cancer for patients in this study was achieved by radiotherapy, either individually or in combination with chemotherapy. Radiotherapy is defined [21] to be

“The treatment of diseases by exposing the affected part to radioactive rays such as X-rays or gamma rays”.

Today, this is one of the most common ways of treating lung carcinomas. Cancer Research UK estimates that 28% of all patients diagnosed with NSCLC tumours are treated using this method. [22].

The underlying concept is to use electromagnetic radiation to ionise DNA molecules in cancerous cells, damaging them such that they can not reproduce. This ionisation occurs by two mechanisms, labelled as *direct* and *indirect* [23]. Direct ionisation is the dominant mode in high energy radiotherapy, whereby photons interact with the DNA molecules causing molecular bonds to break. In indirect ionisation, the photons instead react with an intermediary molecule (such as water). This generates free radicals which then ionise the DNA. This mode is more dominant in low-energy radiotherapy.

Conventionally, the dose strength in radiotherapy is measured using *Gray* (Gy). 1 Gy corresponds to 1 Jkg^{-1} , or one joule of energy deposited by treatment x-rays per kilogram of tissue irradiated.

2.3 Computed Tomography Scanning

In order to diagnose tumours and monitor how they change over a course of radiotherapy, *computed tomography* (CT) imaging is used. This method uses variability in x-ray attenuation of different tissues to create detailed internal images of the patient showing differences in tissue density. The method relies implicitly on the fact the attenuation of tissues is dependent on their electron density which in turn strongly correlates to the physical density of the tissue [24].

The basic principle is as follows. The patient lies on a bed whilst a motorized X-ray source rotates around them on a helical path, called a *slip ring gantry*. The attenuation of the emitted X-rays are measured as they pass through the patient at different angles. Once the degree of attenuation is measured for all angles, a non-trivial mathematical transform (*filtered back projection* [25]) generates an image. An example of a CT image is shown in Figure 2.

In this method a combination of traditional computed tomography (CT) scans and modern cone-beam computed tomography scans (CBCT) scans are used. CBCT differs from traditional CT scanning in that x-rays diverge, creating a cone [26]. CBCT is advantageous as the scanner can be mounted on radiotherapy treatment machines, allowing image guidance at the time of therapy. However because CBCT irradiates 3-D cones of tissue rather than 2-D slices it is more susceptible to noise from surrounding tissue and produces less accurate images.

CT and CBCT scans differentiate between different types of tissue according to how they attenuate X-rays. The degree to which this occurs is generally measured in *Hounsfield Units* (HU) ¹. This is a linear transform of the classical attenuation coefficient which allows easy comparison to the attenuation of water, specified by

$$HU = 1000 \times \frac{\mu - \mu_{water}}{\mu_{water} - \mu_{air}}. \quad (1)$$

3 METHOD

In the proposed methodology, voxel information from CT and CBCT scans is used to determine changes in density in the regions surrounding lung cancer tumours. Voxels are points in 3-dimensional grids containing image information, analogous to pixels in 2-dimensional images. The challenge presented here is defining these regions of interest to analyse; tumours do not have uniform attenuation coefficient throughout their volume and can be similarly attenuating to surrounding healthy tissue, making them difficult to delineate.

1. Due to the additional noise effects in CBCT the calibration of attenuation coefficients and HU is much more challenging. Instead an approximation to HU called 'CBCT number' is often used. A review of this can be found in [27]. However, for the purposes of this report CBCT scans can be considered to measure HU.

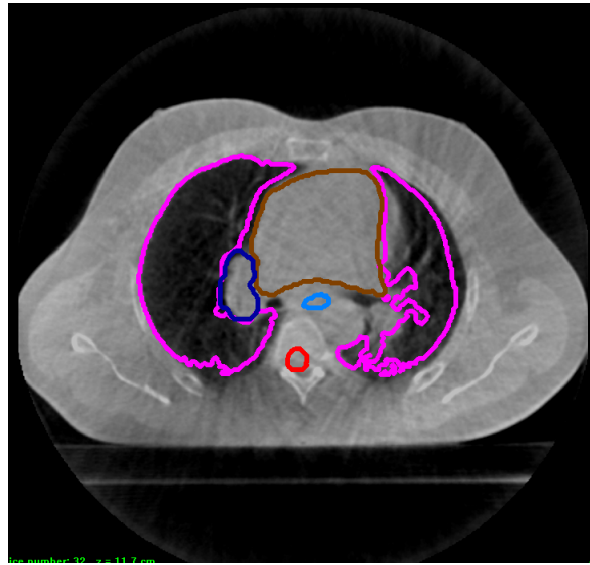


Figure 2. An example slice of a delineated CT treatment planning scan. The density of the tissue is shown by its brightness with the most dense tissue in white and the least dense in black. A variety of delineations are shown here; ITV (dark blue), heart (brown), oesophagus (light blue), lung walls (purple) and spinal column (red).

3.1 Data Set

This experiment used datasets of CT and CBCT scans across radiotherapy courses for 8 NSCLC patients treated at the Christie NHS Foundation Trust in Manchester. The radiotherapy course for each patient was defined as follows. On admission to the centre, patients were given an initial treatment planning (TP) CT scan. This was used by clinicians to delineate the volume containing the tumour, known as the *gross tumour volume* (GTV). Other anatomical regions of interest such as the spinal cord, heart and lung walls were delineated. On some scans an *internal target volume* (ITV) was delineated instead of a GTV. This, similarly to a GTV, delineates the area occupied by the tumour however leaves an extra set-up margin around the tumour to account for variability in patient positions and inaccuracies in alignment of the CT scanner beams [28]. Figure 2 shows an example of a TP CT scan used.

The TP scan was analysed and a radiotherapy course was planned for the patient. Each patient received 55 Gy of radiation over 20 daily treatments (known as fractions). The patients underwent regular CBCT scans during radiotherapy (approximately every 3 to 7 days) to monitor daily setup variability. This meant clinicians could monitor developments in tumour anatomy and make changes to radiation doses if necessary. The scans obtained from these were aligned with each other in a method known as *registration*. This was done by comparing mutual information between the TP scan and subsequent scans for regions containing the tumour. Figure 3 shows a complete scan series for one patient.

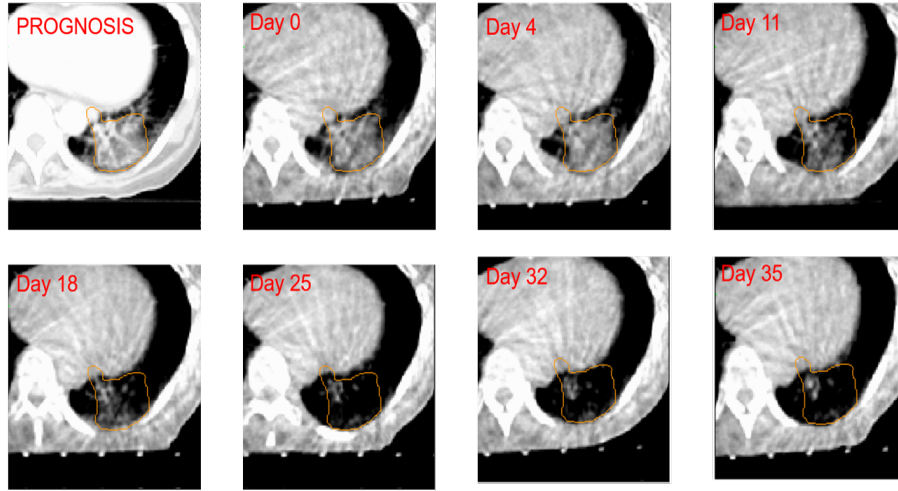


Figure 3. A complete scan series for one patient used in this experiment, labelled NSCLC7. The orange delineation on the scan shows the GTV delineated by a clinician on the TP scan (top left picture). Subsequent scans show tumour changes over the course of radiotherapy. It should be noted that this figure only shows 1 of 109 axial slices of the lung cavity taken by the scan.

3.2 Mask Generation

To analyse the voxel information of the region surrounding the tumour, it was necessary to define a consistent automated method of creating a mask for this region. This was achieved by applying an image-processing pipeline to the Day 0 CBCT scan for each patient. This pipeline is described by Figure 4. The only manual input to this process is the tumour delineation in the TP scan. Masks were generated from the scan done at the beginning of treatment rather than from the TP scan in order to account for any anatomical changes in the tumour that may have occurred between treatment planning and treatment beginning.

3.2.1 Thresholding and removing other dense tissues

Delineating tumours effectively is a major challenge in radiotherapy treatment today. Automated delineation is a very active field of research ([30], [31], [32]) however there is still no definitive way of achieving this task. Furthermore, manual delineation by clinicians is believed to have a great deal of uncertainty associated with it [33]. It was observed that many of the tumours in the dataset appeared to be rough, capturing a significant amount of healthy tissue as well as the tumour. An example of a GTV delineation in our dataset which appears to contain healthy tissue is shown in Figure 5.

In this study co-ordinates of all voxels contained by the manually delineated region are extracted. Then, to remove healthy tissue remaining in the delineation volume, these regions are thresholded and all voxels below 500 HU are removed from the volume.

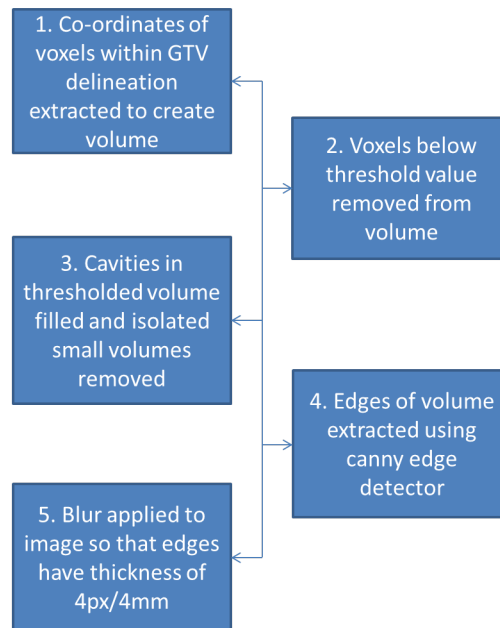


Figure 4. The data pipeline applied to each Day 0 CBCT scan to generate masks defining regions to be analysed. This process was implemented using the Python 3 programming language [29]

Another potential issue was dense tissues other than tumours appearing in the region around the edge of the tumour to be analysed. Examples of such dense tissues are bone or heart muscle tissues. This could have had a confounding effect as it may have been mistaken for cancerous tissue. To resolve this, delineations of other regions containing dense tissues such as the heart or lung walls were removed from masks of the GTV.

3.2.2 Filling cavities and removing small bodies

Due to variability in density of tumours and noise from surrounding tissues affecting CBCT scans, cavities often appeared in the tumour volume after thresholding. This is shown in Figure 6. Small isolated cancerous regions also appeared after thresholding. However, we are interested in the region immediately surrounding the main body of the tumour. Therefore the cavities are filled and any small bodies at the edge of the volume are removed. An example of this process is shown in Figure 6.

This process is achieved using the `scipy` Python package's [34] `ndimage.binary_fill_holes` (to remove the cavities in each slice) and `skimage's` [35] `morphology.remove_small_objects` (to remove small bodies at the edge of the main tumour volume).

The key to both these operations is labelling all interconnected spaces in the region. This is achieved using the *flood-fill* algorithm, often used to make 'bucket-fill' tools on graphic

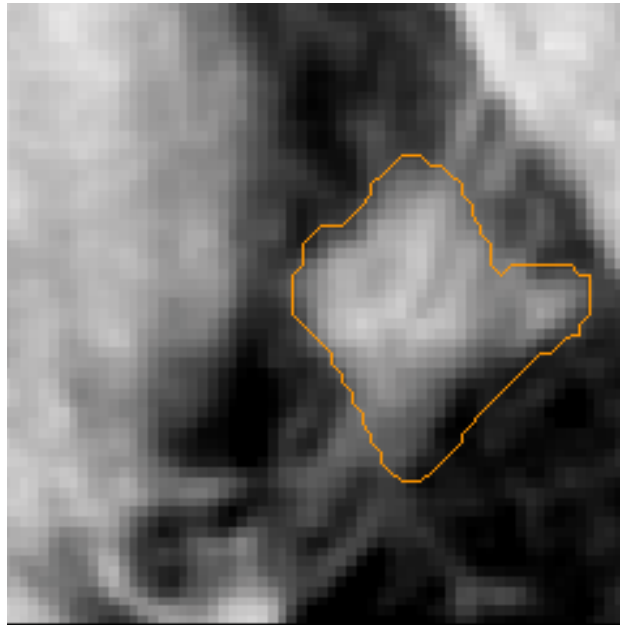


Figure 5. The initial scan on treatment for one patient in the study, NSCLC6. The manual delineation of the tumour is outlined in orange. This appears to capture less dense, healthy tissue, shown by the darker regions at the bottom of the volume.

editing software. The algorithm labels interconnected spaces by specifying that each voxel must be connected by some connection function to at least one other voxel in that space to be considered a member. The connection function used defines any voxels neighbouring each other as connected. Diagonally touching voxels are not considered connected.

This labelling is achieved as follows.

- 1) Starting from any voxel, find the closest unconnected voxel, v .
- 2) Label all voxels connected to v . This is done by determining $\{V\}$, the set of all neighbouring voxels connected to v and then recursively calculating the connections to each element of $\{V\}$ until all voxels have no further unlabelled connections.
- 3) Repeat the above steps until no more unlabelled voxels remain in the array.

In `ndimage.binary_fill_holes` the flood-fill algorithm is used to find interconnected volumes of value 0 inside interconnected volumes of value 1 and in `morphology.remove_small_objects` it is used to remove all interconnected volumes of value 1 with less than a specified number of voxels. After these operations, a single, interconnected and filled volume is left, defining the main body of the tumour.

3.2.3 Edge detection and blur

Once the mask representing the main body of the tumour was generated, it was used to create a region of interest (ROI) at the borders of the tumour to analyse throughout the scan series.



Figure 6. A figure showing the process of thresholding and removing cavities and fragments from the initial manually delineated region for one CBCT slice. After thresholding, the mask is converted to binary by setting all remaining non-zero voxels to 1. Any cavities or small bodies created as a result of thresholding are then removed.

It was decided this ROI would be defined as an annulus 2mm inside and 2mm outside of the tumour. Since each voxel in the scans occupies 1mm^3 , the annulus was a 4 voxel thick shell surrounding the main tumour volume. Creating this mask involved a two-stage process; finding the edges of the volume defining the main body of the tumour and then blurring the edges so that the annulus had the desired thickness.

The edge detection was achieved using a Canny edge detector [36]. This detector was chosen due to its speed and robustness to noise. `skimage's feature.canny` implementation of the algorithm was used. This multi-stage algorithm is as follows:

- 1) Noise is reduced by applying Gaussian blur to the image. Each voxel becomes a weighted average of its neighbourhood, where the weight of each neighbouring voxel is a Gaussian function centered on the initial voxel. After consideration of the noise caused by uncertainties in CBCT scanning, a σ value of 2 voxels was chosen.
- 2) A Sobel-Feldman filter [37] was applied to the blurred image slice-by-slice, in order to determine the gradient intensity and direction at each pixel. This is achieved by convolving the image with G_x and G_y to get the x and y components of the gradient respectively where

$$G_x = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix} \quad (2)$$

and

$$G_y = \begin{bmatrix} +1 & +2 & +1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}. \quad (3)$$

The images generated after the convolutions are A_x and A_y . For the pixel at co-ordinates (i,j) , the intensity of its gradient $|A_{ij}|$ is calculated by

$$|A_{ij}| = \sqrt{A_{xij}^2 + A_{yij}^2} \quad (4)$$

and the direction of the gradient, θ , is given by

$$\theta = \arctan \left(\frac{A_{yij}}{A_{xij}} \right). \quad (5)$$

- 3) Once the gradient field has been extracted from the image, *non-maximal suppression* is applied to detect edges in the image; each pixel in each slice is checked to see if it is a local maximum of gradient intensity in the direction of the gradient and if not, the pixel is *suppressed* (set to zero). [38]. Conventionally, a final step of *hysteresis thresholding* is used to remove soft edges in non-binary pictures. However because the masks are binary, this step is not essential here.

The algorithm results in a one pixel thick annulus defining the edge of the main body of the tumour. The final step to create the mask to analyse was to increase its thickness so that it extended 2 mm (or 2 px) inside the tumour and 2 mm outside the tumour. This was achieved using `skimage's filters.rank.maximum` function whereby every pixel in a specified neighbourhood is set to the local maximum of pixel value. The neighbourhood used was a disc of radius 2 pixels implemented by `skimage.morphology.disc`. The edge detection process applied to the main tumour body shown in Figure 6 is demonstrated in Figure 7.

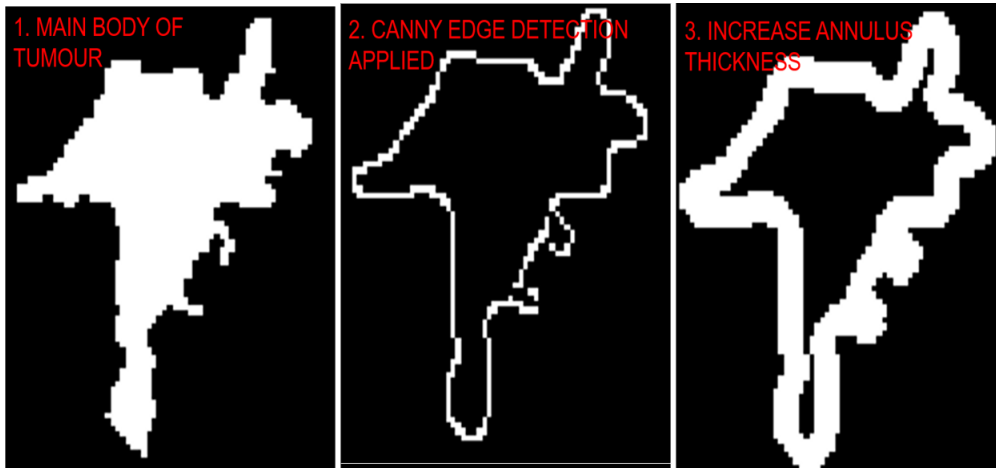


Figure 7. A figure showing the process of edge detection for the main body of the tumour generated in Figure 6. Step 3 shows the blurring of the edges detected in step 2.

3.3 Superposing the mask onto CBCT scans

The mask generated denoted the location of the edge of the tumour in the first scan. The tissue composition of this region of interest (ROI) was analysed throughout the course of radiotherapy by superposing the mask with subsequent, registered scans. This was achieved by performing a pixel-by-pixel multiplication of the pixels from the mask (either 0 or 1) and the scans. The newly generated images for one slice of a scan series are shown in Figure 8.

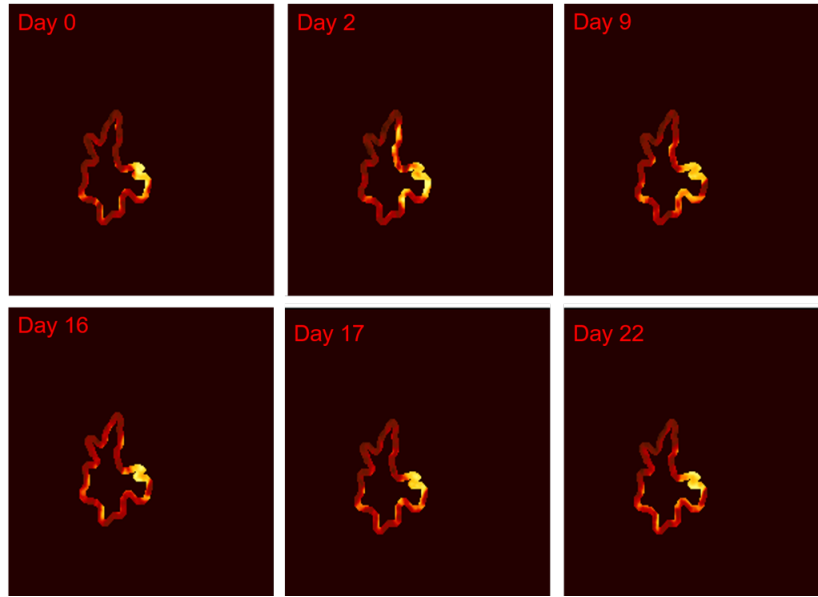


Figure 8. A demonstration of how tissue densities change throughout the scan for one slice of one patient (NSCLC1). Brighter regions show denser tissue, likely to be cancerous. Darker regions are less dense, healthy tissue.

4 ANALYSIS AND RESULTS

4.1 Analysing voxel information

Once the scans had been combined with the masks as demonstrated in Figure 8, the voxel information of the ROI was analysed. A series of histograms were created for each patient, showing the change in distribution of HU values in the ROI throughout the radiotherapy course. As an example, Figure 9 shows the histograms created for the patient shown in Figure 8.

Most histograms generated exhibited a bimodal distribution with the lower HU peak corresponding to healthy lung tissue captured by the mask and the higher HU peak corresponding to cancerous tissue. In order to extract features from these histograms and analyse how they changed during therapy, regression was used to fit a function to each histogram. This was also beneficial to lessen the effects of noise in the data.

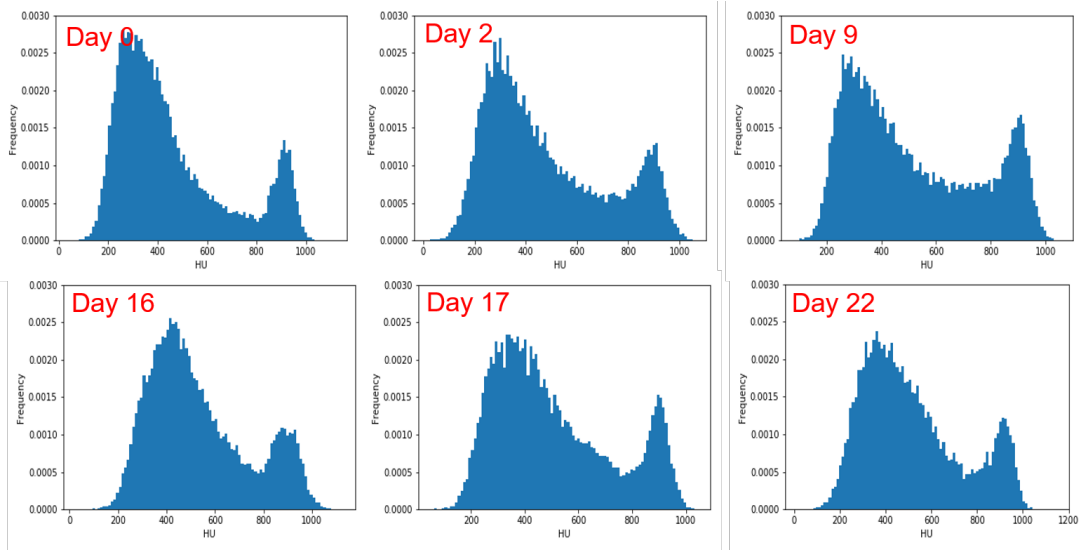


Figure 9. Normalised histograms showing how the distribution of HU values throughout the region of interest changes during treatment. We can see two clear peaks in the distribution; one at around 300 HU and one at 900 HU. The 300 HU peak corresponds to healthy lung tissue, whereas the 900 HU peak likely indicates cancerous tissue.

This was achieved by fitting multimodal guassian distributions to the data using Poisson maximum log-likelihood regression. It is assumed that the data is modelled by a bimodal distribution,

$$F(x) = A_1 \exp\left(\frac{-(x - \bar{x}_1)^2}{\sigma_1^2}\right) + A_2 \exp\left(\frac{-(x - \bar{x}_2)^2}{\sigma_2^2}\right) \quad (6)$$

where $F(x)$ is the normalised frequency of voxels at HU value x , A_1 and A_2 are the amplitudes of each peak, \bar{x}_1 and \bar{x}_2 are the maxima values and σ_1 and σ_2 are their standard deviations.

Each histogram is fully specified by a set of tuples $\{(x, y)\}$, where x is the median HU of a histogram bin and y is the number of voxel counts in that bin. The probability of a bimodal distribution of the form shown in Equation 6 generating a histogram specified by the set of tuples $\{(x, y)\}$ is correlated to the *log likelihood*,

$$\ln(p(y_1, \dots, y_m | x_1, \dots, x_m)) = \sum_{i=1}^m y_i \ln(F(x_i)) - F(x_i) - \ln(|\Gamma(y_i + 1)|), \quad (7)$$

where

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt \quad (8)$$

When the log likelihood is maximised for a given set of parameters (A_1 , A_2 , \bar{x}_1 , \bar{x}_2 , σ_1 and σ_2) the distribution is most likely to produce the data given by a histogram. Using this

method and gradient descent achieved with the L-BFGS-B algorithm ([39], [40]), a bimodal Gaussian distribution was fitted to each histogram. This was implemented using `scipy's optimize.minimize` package (the negative log-likelihood was minimised). It was found however, that a trimodal Gaussian distribution tended to give a better fit. Accordingly, in practice

$$F(x) = A_1 \exp\left(\frac{-(x - \bar{x}_1)^2}{\sigma_1^2}\right) + A_2 \exp\left(\frac{-(x - \bar{x}_2)^2}{\sigma_2^2}\right) + A_3 \exp\left(\frac{-(x - \bar{x}_3)^2}{\sigma_3^2}\right) \quad (9)$$

was fitted to the data.

4.2 Comparison of peaks corresponding to healthy and cancerous tissue

An attempt was made to observe the change in relative heights of the two peaks corresponding to the cancerous and healthy tissues. Figure 10 shows the how the value

$$r(t) = \frac{A_3(t) - A_1(t)}{A_3(t) + A_1(t)} \quad (10)$$

changes over the course of radiotherapy. If $\frac{r(t)}{r(0)} > 0$ this means the amount of cancerous tissue is the ring has increased since the beginning of therapy and if $\frac{r(t)}{r(0)} < 0$ the amount of cancerous tissue in the ring has decreased.

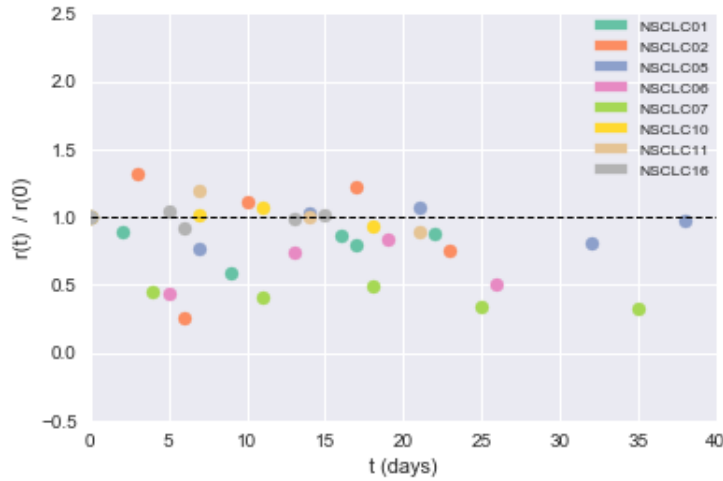


Figure 10. A graph plotting how the value of r (defined in Equation 10) changes during therapy for each of the eight patients in the study. As expected, there is a negative correlation between time subjected to radiotherapy and amount of cancerous tissue in the ROI.

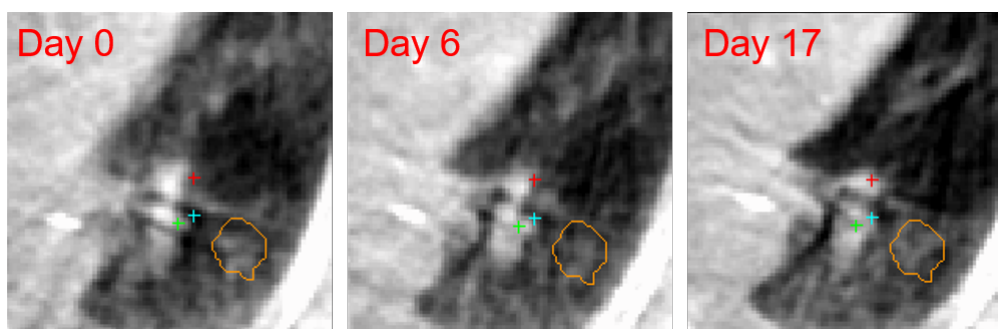


Figure 11. A figure showing the markers used to determine manually whether a tumour is shrinking elastically or non-elastically. The tissue marked out by the fiducial markers (crosses) appears to move slightly towards the ITV (orange delineation), into the space vacated by the recessing tumour. It was thus concluded that this tumour was recessing elastically.

4.3 Overlaid markers method

In order to determine what features may be inherent to time series of the form shown in Figure 10 for each mode of tumour recession, an attempt was made to classify the mode manually. This relied on observing the movement of tissue around the tumour, similar to the method used by Hamming-Vrieze et al. in [16]. Markers were overlaid on each aligned scan for each patient to see whether key features of the surrounding tissue such as bronchioles moved during the course of treatment. If surrounding tissue is found to be moving towards the tumour into areas vacated by the shrinking tumour it is assumed to be elastic. Otherwise the tumour is assumed to be non-elastic (fragmenting). Figure 11 shows three scans from the time series for one patient.

As this method relies on the interpretation of an individual, to ensure reproducibility the scans were classified twice independently. Both classifications were found to broadly agree with each other. The results are shown in Figure 12.

Patient Name	Mode of Recession
NSCLC1	Elastic
NSCLC2	Non-Elastic
NSCLC5	Non-Elastic
NSCLC6	Non-Elastic
NSCLC7	Elastic
NSCLC10	Non-Elastic
NSCLC11	Non-Elastic
NSCLC16	Elastic

Figure 12. A table showing the results of the overlaid marker method for each patient. 5 of the 8 patients are believed to be shrinking non-elastically, compared to 3 which are shrinking elastically

This classification was used in conjunction with information from Figure 10 to generate Figure 13. This shows the time series for the patients split into two graphs; one showing tumours believed to be recessing elastically and one showing tumours believed to be recessing non-elastically (fragmenting). Whilst it appears the elastically fragmenting patients had a greater and more consistent decrease in r during therapy, the difference was not found to be statistically significant due to patient numbers available. It was also observed that patients with a greater decrease in r over the course of therapy appeared to have a greater variance from a linear time series. A linear regression was applied to the time series and a χ^2 test was performed on each fit. The correlation between the gradient of each fit and its χ^2 value was compared using Pearson's R value, found to be 0.8860.

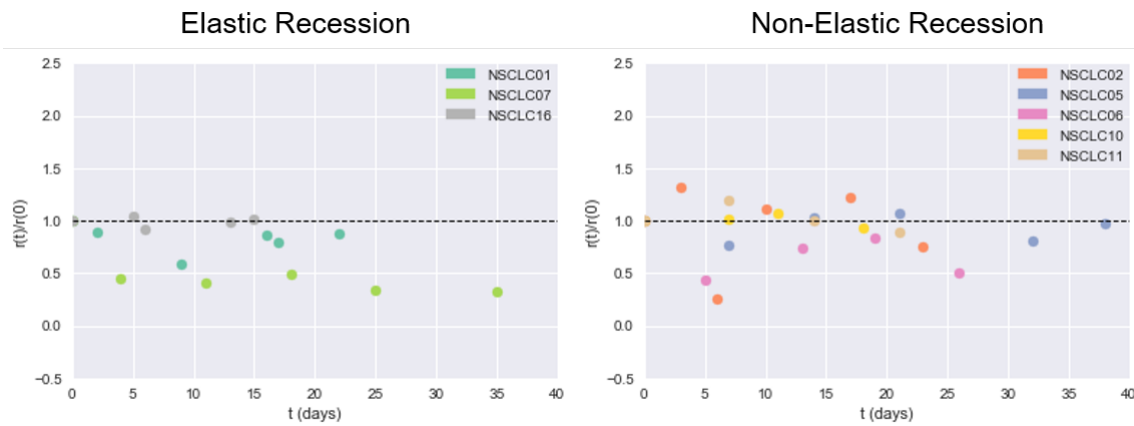


Figure 13. The time series for the patients were split according to how they were classified using the fiducial markers technique.

4.4 Comparing height of cancerous tissue peak in annulus to equivalent peak for the whole tumour

After the method described above did not create a clear separation between the modes, it was suggested that the degree to which a tumour is fragmenting can be measured by comparing the HU composition of tissue at the tumour borders to the HU composition for the whole tumour. If the proportion of cancerous tissue in a ring around the edge of the tumour decreased rapidly compared to the proportion of cancerous tissue in the whole tumour, it is likely to be shrinking elastically, whereas if the degree to which the proportion of cancerous tissue decreases is similar for the tumour borders and the whole tumour, it may be evidence for fragmentation.

In order to make this comparison, histograms were generated showing the HU distribution across the whole tumour. This was done using the same method of mask generation described in section 3 without the edge detection step. A trimodal gaussian distribution was then fit to the data and the relevant parameters were extracted. The ratio of the amplitudes of the peaks

corresponding to cancerous tissue for the annulus and the whole tumour, x , was measured;

$$x(t) = \frac{A_3^{ROI}(t)}{A_3^{tum}(t)} \quad (11)$$

where $A_3^{ROI}(t)$ is the amplitude for the peak corresponding to cancerous tissue in the annulus around the tumour and $A_3^{tum}(t)$ the corresponding peak for the whole tumour. The time series of this quantity for each patient is shown in Figure 14.

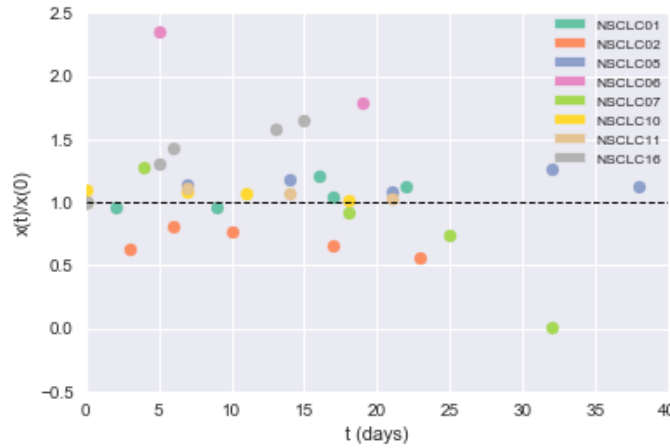


Figure 14. A graph showing the change in x throughout therapy. Like Figure 10, there is no clear separation between the two modes of recession.

However, while a range of time series shapes were observed there was no clear separation of the suspected elastically shrinking and fragmenting tumours. Thus it was concluded another methodology to separate the two modes is required.

5 UNCERTAINTIES AND POTENTIAL FURTHER RESEARCH

This section concerns limitations of the methodology used, as well as speculation on how it may be extended to generate a clear separation.

5.1 Sources of Uncertainty

The major source of uncertainty was the mask generation step. This is a difficult challenge and while thresholding removed some inaccuracies of the manual delineations, it was likely not entirely accounted for. Accurately and reproducibly delineating tumours remains one of the key challenges of image-guided radiotherapy today.

Another problem was quantitatively analysing the generated histograms, achieved by fitting a model to them using regression. This was done to identify cancerous tissues and healthy tissue in the tumour. Analysis was done on the amplitude of the peaks. However the area

under the peaks should have been analysed instead as this is the quantity that corresponds to the number of voxels. A further issue is that the analysis was often confounded by other high density tissue such as bone being near the tumour. This was accounted for by removing any part of the tumour masks near the lung wall or heart.

Also of concern was the registration (alignment) of the scans. If scans in a series were not properly registered, this could result in different areas of the tumour being analysed in each scan, leading to different tissue distributions.

Only 8 patients were used in the course of this experiment. It is possible that with more to analyse, a distinction between the two modes of recession could be made. Another possibility not explored here is that recession could occur by a combination of both modes, shrinking for the most part with slight fragmentation or changing modes midway through therapy.

5.2 Areas Highlighted for Further Research

The methodology used did not distinguish between the two methods of decay, however here several suggestions are made for an improved method which may tell the difference.

One approach suggested was the use of *radiomics*. This is a method of converting CT scans into an extensive feature set of potentially several hundred elements. Examples of features include several descriptors of image histograms, as well as descriptors of relationships between certain voxels [41]. Coupled with a larger dataset and potentially using unsupervised machine learning techniques, this high dimensional data could yield further insights into the recession modes.

Another area to explore would be applying an automated structure identification algorithm to better identify tumour topology. One proposed algorithm was DisPerSE (DIScrete PERsistent Structures Extractor) [42]. A particular area of interest would be the identification of filaments protruding from tumours. These have been shown to perform an integral role in the spread of cancers ([43], [44]) and identification could be useful for determining aggressive tumour growth.

Finally, it has been suggested that the value of the cost function used in image registration could be used as a feature to determine whether tumours were shrinking or fragmenting. Due to topological differences between tumours shrinking elastically and fragmenting, the registration algorithm would perform better on one mode than the other. By analysing how this algorithm performs, differences between the two modes of recession could be observed.

6 CONCLUSION

An image-based data mining methodology was developed, analysing voxel information around the edge of tumours in CT scans. An attempt was made to use this methodology to classify whether lung cancer tumours were shrinking elastically or fragmenting during the course of therapy. This was tested on 8 NSCLC patients and compared with a method of determining whether cancers were shrinking elastically or fragmenting using fiducial markers. No clear distinction between the two modes were observed using the suggested method. Therefore,

classification of tumour regression mode remains an open problem. Several suggestions were made to improve the methodology in order to generate an effective system to solve this problem.

7 ACKNOWLEDGEMENTS

George and I would like to express our gratitude to those contributed in particular to this project. Firstly, thanks to Professor Anna Scaife of the Jodrell Bank Centre for Astrophysics for her guidance and support in particular on image-processing side of the project. We would also like to thank Dr. Alan McWilliam and Dr. William Beasley of the Christie Cancer Research Centre for their help with the medical side of the project, contribution of software, procurement of the dataset and, most importantly, patience when we were confused.

REFERENCES

- [1] Cancer Research UK, <http://www.cancerresearchuk.org/health-professional/cancer-statistics/risk/lifetime-risk> , [Accessed 10/12/2017]
- [2] Cancer Research UK, <http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/lung-cancer/incidence> [Accessed 10/12/2017]
- [3] Cancer Research UK, <http://www.cancerresearchuk.org/health-professional/cancer-statistics/incidence/common-cancers-compared> [Accessed 10/12/2017]
- [4] Cancer Research UK, <http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/lung-cancer> [Accessed 10/12/2017]
- [5] Dragun A.E. (2013) Acute Radiation Toxicity. In: Brady L.W., Yaeger T.E. (eds) Encyclopedia of Radiation Oncology. Springer, Berlin, Heidelberg
- [6] Einhorn, L. et al. (1976) "Enhanced pulmonary toxicity with bleomycin and radiotherapy in oat cell lung cancer" *Cancer* 37(5) pp. 2414-2416
- [7] Abratt, R.P., Morgan, G.W. (2002) "Lung toxicity following chest irradiation in patients with lung cancer." *Lung Cancer* 35(2) pp. 103-109
- [8] Müller, G. et al. (1994) "Radiation pneumonitis and pulmonary fibrosis after the CT-planned radiotherapy of bronchial carcinoma" *Strahlenther Onkol* 170(7) pp. 400-407
- [9] Wang, K. et al. (2017) "Cardiac Toxicity After Radiotherapy for Stage III Non-Small-Cell Lung Cancer: Pooled Analysis of Dose-Escalation Trials Delivering 70 to 90 Gy" *Journal of Clinical Oncology* 35(13) pp. 1387-1394.
- [10] Kale, M.S. et al. (2017) "Treatment Toxicity in Elderly Patients With Advanced Non-Small Cell Lung Cancer." *Am J Clin Oncol* 40(5) pp. 470-476.
- [11] Bosmans, G. et al. (2006) "Intra-patient variability of tumor volume and tumor motion during conventionally fractionated radiotherapy for locally advanced non-small-cell lung cancer: A prospective clinical study" *Int. J. Radiation Oncology Biol. Phys.* 66(3) pp. 748-753
- [12] Kupelian P.A. et al. (2005) "Serial megavoltage CT imaging during external beam radiotherapy for non-small-cell lung cancer: observations on tumor regression during treatment" *Int. J. Radiation Oncology Biol. Phys.* 63(4) pp. 1024-1028
- [13] Lim G. et al. (2011) "Tumor Regression and Positional Changes in Non-small Cell Lung Cancer During Radical Radiotherapy" *Journal of Thoracic Oncology* 6(3) pp. 531-536
- [14] Knap, M.M. et al. (2010) "Daily cone-beam computed tomography used to determine tumour shrinkage and localisation in lung cancer patients" *Acta Oncologica* 49 pp. 1077-1084
- [15] Juhler-Nøttrup, T. et al. (2008) "Interfractional changes in tumour volume and position during entire radiotherapy courses for lung cancer with respiratory gating and image guidance" *Acta Oncologica* 47 pp. 1406-1413
- [16] Hamming-Vrieze, O., et al. (2017) "Analysis of GTV reduction during radiotherapy for oropharyngeal cancer: Implications for adaptive radiotherapy" *Radiotherapy and Oncology* 122(2) pp. 224-228
- [17] Hanahan D., Weinberg R.A. (2000). "The hallmarks of cancer". *Cell* 100(1) pp. 57-70
- [18] lungcancer.org, "Lung Cancer 101", https://www.lungcancer.org/find_information/publications/163-lung_cancer_101/268-types_and_staging [Online] Accessed 17/12/2017.

- [19] Rodriguez-Canales, J., Parra-Cuentas, E., and Wistuba, I.I. "Diagnosis and Molecular Classification of Lung Cancer" *Lung Cancer Treatment and Research* Editor: Reckamp, K.L., Springer International Publishing Switzerland, 2016 pp. 25-26.
- [20] Arsen'ev A.I. et al. (2008) "A comparison of efficacy of different modalities of treatment for locally-advanced and metastatic non-small lung cancer" *Vopr Onkol* 54(3) pp. 281-286
- [21] Anon (2004) "Radiotherapy". Dictionary of Medical Terms. 4th ed. London: A C Black. pg. 341
- [22] Lung Cancer UK, <http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/lung-cancer> [Accessed 19/12/2017]
- [23] Steel, G.G. (2002) Basic Clinical Radiobiology, 3rd ed. Oxford: Oxford University Press.
- [24] Kamalian, S., Lev, M.H. and Gupta, R. (2016) "Computed tomography imaging and angiography – principles" *Neuroimaging Part I, Handbook of Clinical Neurology*. Volume 135, pp. 3-20
- [25] N. Chetih and Z. Messali (2015) "Tomographic image reconstruction using filtered back projection (FBP) and algebraic reconstruction technique (ART)", *2015 3rd International Conference on Control, Engineering & Information Technology (CEIT)*, Tlemcen, 2015, pp. 1-6.
- [26] Hatcher, D.C. (2010) "Operational principles for cone-beam computed tomography" *Journal of the American Dental Association* 141 (Suppl. 3) pp. 3S–6S.
- [27] Srinivasan, K., Mohammadi, M., Shepherd, J. (2014) "Applications of linac-mounted kilovoltage Cone-beam Computed Tomography in modern radiation therapy: A review" *Polish Journal of Radiology* 79 pp. 181-193.
- [28] Berthelsen, A.K. et al. (2007) "What's new in target volume definition for radiologists in ICRU Report 71? How can the ICRU volume definitions be integrated in clinical practice?" *Cancer Imaging* 7(1) pp. 104–116
- [29] Python Software Foundation. Python Language Reference, version 3.5. Available at <http://www.python.org>
- [30] Ballangan, C. et al. (2011) "Automated delineation of lung tumors in PET images based on monotonicity and a tumor-customized criterion" *IEEE Trans Inf Technol Biomed.* 15(5) pp. 691-702
- [31] Gu, Y. et al. (2013) "Automated Delineation of Lung Tumors from CT Images Using a Single Click Ensemble Segmentation Approach" *Pattern Recognit.* 46(3) pp. 692–702
- [32] Velazquez, E.R. et al. (2013) "Volumetric CT-based segmentation of NSCLC using 3D-Slicer" *Scientific Reports* 3, Article Number 3529.
- [33] Segedin, B. and Petric, P. "Uncertainties in target volume delineation in radiotherapy – are they relevant and what can we do about them?" *Radio Oncol.* 50(3) pp. 254–262.
- [34] Jones, E., Oliphant, E., Peterson, P. et al. "SciPy: Open Source Scientific Tools for Python" 2001-, <http://www.scipy.org/> [Online] Accessed 18/12/2017
- [35] van der Walt, S. et al. (2014) "scikit-image: image processing in Python"
- [36] Canny, J. (1986) "A Computational Approach To Edge Detection", *IEEE Transactions on Pattern Analysis and Machine Intelligence* 8(6) pp. 679–698
- [37] Duda, R. and Hart, P. (1973) *Pattern Classification and Scene Analysis*, 1st ed. Hoboken, New Jersey: John Wiley and Sons. pp. 271-272
- [38] OpenCV, "Canny Edge Detection", https://docs.opencv.org/3.3.1/da/d22/tutorial_py_canny.html [Online] Accessed 19/12/2017
- [39] Byrd, R. H., Lu, P. and Nocedal, J. (1995) "A Limited Memory Algorithm for Bound Constrained Optimization" *SIAM Journal on Scientific and Statistical Computing* 16 (5) pp. 1190-1208.
- [40] Byrd, R. H., Lu, P. and Nocedal, J. (1997) "L-BFGS-B: Algorithm 778: L-BFGS-B, FORTRAN routines for large scale bound constrained optimization" *ACM Transactions on Mathematical Software* 23 (4) pp. 550-560.
- [41] Gillies, R.J., Kinahan, P.E. and Hricak, H. (2016) "Radiomics: Images Are More than Pictures, They Are Data" *Radiology* 278(2) pp. 563-57
- [42] Sousbie, T. (2013) "DisPerSE: robust structure identification in 2D and 3D" arXiv:1302.6221 [astro-ph.CO]
- [43] Barberis, L. et al. (2009) "Leukocyte transmigration is modulated by chemokine-mediated PI3K γ -dependent phosphorylation of vimentin." *Eur J Immunol* 39 pp. 1136–1146.
- [44] Tzivion, G. et al. (2000) "Calyculin A-induced vimentin phosphorylation sequesters 14-3-3 and displaces other 14-3-3 partners in vivo." *J Biol Chem* 275 pp. 29772–29778