Classification of fragmentation in lung cancer tumours using radiomic and fractal features

Rhydian Windsor
Performed in collaboration with George Needham.

Abstract

Some lung cancer tumours are believed to fragment as they shrink during radiotherapy. Here we continue on previous work exploring this phenomena [1], considering radiomic and fractal features of tumours. Feature selection algorithms are used to determine which of these features are important in determining whether fragmentation is occurring. Simulations of fragmenting and non-fragmenting tumours are then used to develop a classification system from the aforementioned important features. The classifier is found to be very accurate in determining whether fragmentation is occurring or not (96.5% 5-fold cross-validation across the simulated dataset). This classifier is then applied to real scan series of several lung cancer patients. We conclude that of all previous approaches, texture analysis is most likely to be useful in determining if fragmentation is occurring.

Keywords

Image-Guided Radiotherapy, Computer Vision, Tumour Recession, Radiomics, Fractal Analysis

Contents

1 INTRODUCTION

Adaptive radiotherapy (ART) is an advanced method of cancer treatment, motivated by a desire to reduce the toxicity [2] of traditional radiotherapy techniques whilst maintaining its treatment benefits. This treatment technique is of particular interest due to the recent rapid development of the field of computer vision, allowing for increasingly sophisticated analysis of image data ([3]–[5]). ART relies on exact knowledge of the location of tumours to ensure only cancerous tissue receives a high radiation dosage. For this reason, generating a robust method of determining the exact location of tumours from image analysis of scans, known as *automated delineation*, is a key area of research in the field currently ([6]–[9]). It is also important to monitor changes in a tumour during therapy so that the irradiated tissue volume can be adapted to accommodate any changes.

This report explores potential modes of shrinkage in lung cancer tumours. Specifically, the aim is to determine whether or not a tumour is *fragmenting* as it shrinks. This property is illustrated in Figure 1 and describes whether or not a tumour leaves behind small amounts of cancerous tissue in the region vacated by the main body of the shrinking tumour. It is important to know if a tumour is fragmenting in order to determine whether as it may be safe to reduce the irradiated volume in subsequent therapy. For example, for the non-fragmenting tumour B in Figure 1 it is safe to stop irradiating the volume marked in pink previously containing cancerous tissue however doing this for tumour C could result in 'fragments' of cancerous tissue being left untreated. These could then potentially cause a relapse at a later stage.

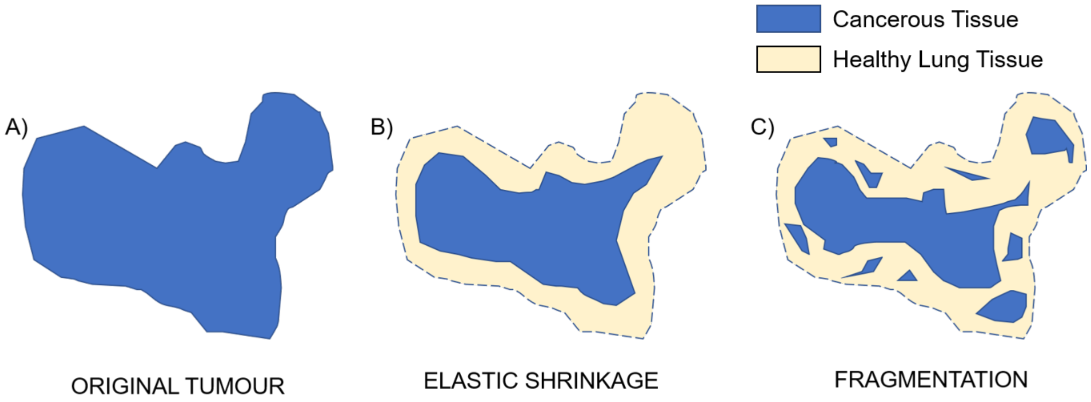


Figure 1. A graphic showing the difference between a fragmenting tumour and a non-fragmenting tumour. It should be noted the fragments which can be seen in tumour C are likely to be too small to observed by conventional CT scans. This figure is adapted from [1].

Very little previous work has been done in determining whether this fragmentation occurs and it remains unclear whether this effect occurs in tumours. One study focused on using the movement of surrounding tissues to determine the mode of shrinkage[1] [10], postulating that if no fragmentation occurs surrounding tissues will move to occupy the region recently-vacated by the tumour to a greater degree than if fragmentation occurs. A large variation in the amount of tissue movement was observed, providing evidence for fragmentation being an observable effect.

Other previous work has been conducted examining basic statistics of voxel data extracted from the region surrounding tumours throughout a scan series (shown in pink in Figure 1) in conjunction with manual visual analysis [1]. It was found that few significant changes were observed. This implies that any fragments left behind by tumours during shrinkage are too microscopic to be observed by conventional CT imaging.

Here, an attempt is made to approach the problem as a canonical supervised machine learning problem. Firstly, algorithms are developed which use a single scan as input and generate two scan series for that tumour, one corresponding to the tumour shrinking with fragmentation and one corresponding to it shrinking without fragmentation. The action of these algorithms are designed to closely follow the definition of fragmentation given above, considering previous work in the area. Once these simulated scan series of fragmenting and

[1] *It should be noted this work examined modes of shrinkage in head and neck tumours rather than lung tumours as used in this report.*

non-fragmenting tumours are defined the resulting images are analysed and several *radiomic* and *fractal* features are extracted (details of which are discussed later).

Features acting as key indicators on whether a scan series is fragmenting or not are determined by *feature selection algorithms* and the distribution of these features among real scan series are observed. In particular we observe similarities in the distributions of these features for the real and simulated data. Finally an attempt is made to train a classifier, which determines if a tumour is fragmenting. The classifier is trained using simulated data and then applied to the real data. The features selected as important are of particular interest as these could form a basis for observations in future studies concerning how tumours shrink.

The structure of the report is as follows; Section 2 gives a brief discussion of some of the relevant background theory for this study. In particular a basic description of some of the algorithms used is given. In section 3, the adopted methodology is outlined. The algorithms used to generate the simulated data are specified and details of how the fractal and radiomic features were extracted are given. The focus then turns to the results of the feature selection algorithms and the classifier developed for the given data. This is then followed in section 5 by a discussion of the significance of these results and ideas for future work in this area. Finally, section 6 summarises and concludes the report.

2 THEORY

2.1 Radiotherapy

Radiotherapy is a method of cancer treatment which relies on the use of intense electromagnetic radiation to kill cancer cells. It is most commonly administered using a linear accelerator (*linac*). The radiation dosage administered to patients is measured in *Grays* (Gy). 1 Gy is equivalent to delivering 1 Joule to a kilogram of matter (1 Gy $\equiv$ 1 Jkg^{-1}).

Radiotherapy was used to treat 27% of all patients diagnosed with all forms of cancer in the UK during 2013-14 [11]. However, a major issue with the treatment is the risk of *toxicities*. This refers to damage caused to otherwise healthy tissues by radiotherapy treatment [2]. Studies have shown toxicity to be an effect in around 37% of all patients treated with radiotherapy [12]. In order to reduce the risk of toxicities occurring in patients it is important to minimise the dose received by healthy tissue.

A recent development in radiotherapy which aims to tackle this problem is adaptive radiotherapy. Here patients receive scans throughout treatment in order to determine if there are any changes in the tumour volume. The irradiated volume is then adapted to account for these changes.

2.2 Computed Tomography Scanning

Computed tomography (CT) scanning is a diagnostic imaging technique using specialised x-ray equipment to generate cross sectional images of patients. It is usually credited to Allan Cormack and Godfrey Hounsfield who jointly won the Nobel Prize for Physiology and Medicine in 1979 for its inception [13].

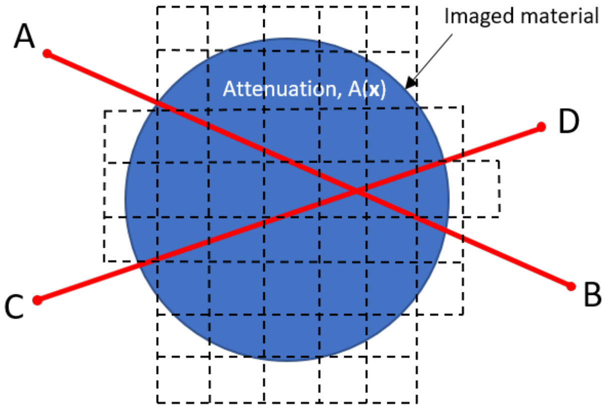


Figure 2. An illustrated of the CT scanning set up. X ray beams are emitted at points A and C and received at B and D respectively. By measuring the reduction in intensity for X-rays in many such configurations the attenuation co-efficient of the imaged material, $A(\mathbf{x})$, can be determined.

In this process x-ray beams of known intensity are directed through a patient at a variety of angles. The degree to which they are attenuated is recorded. This is illustrated in Figure 2. For the x-ray passing from A to B the measured intensity, I_1, is given by

$$I_1 = I_0 \exp^{\int_A^B A(\mathbf{x}) d\mathbf{x}} \tag{1}$$

where I_0 is the intensity of the x-ray emitted at A and $A(\mathbf{x})$ is the attenuation co-efficient at position vector x. Equation 1 can be re-arranged to get

$$g = \log \frac{I_1}{I_0} = \int_A^B A(\mathbf{x}) d\mathbf{x}. \tag{2}$$

Discretising the domain of equation 2, x into a rectangular grid as shown in Figure 2 yields the system of equations

$$\mathbf{g} = L\mathbf{a} \tag{3}$$

where a is a column vector of attenuation coefficients at each point in the grid, L is a system where each row denotes a path through the attenuated material and g is a column vector of measured changes in intensity [16]. Thus to determine the attenuation coefficients as a function of position the column vector a must be determined;

$$\mathbf{a} = L^{-1}\mathbf{g}. \tag{4}$$

Thus determining the attenuation co-efficient of the material as a function of position Ax is a matter of calculating the inverse of L. Problems such as these are referred to as *inverse problems* [17]. The exact methods of solving these problems can be highly non-trivial and are beyond the scope of this report.

2.3 Radiomics

Radiomics is a term penned in 2011 by Aerts et al. [18] [19] to describe the extraction of quantitative data from topological scans. The general method involves applying algorithms to scans to extract descriptive features. These features can be a wide range of metrics, from basic so-called *first order* features such as the mean and variance of voxel intensities in the scan to more complex features associated with texture analysis. The extracted features can then be used to make statistical inferences for a wide range of medical phenomena. A key concept is that radiomic analysis is *automated* and *reproducible*, allowing results to be easily and rapidly verified. It is of particular interest due to the current rise in popularity of machine learning in medical research [20], where radiomic features are often used as summary statistics for algorithms ([21], [22], [23], [24]).

Radiomic features are generally divided into seven classes of features

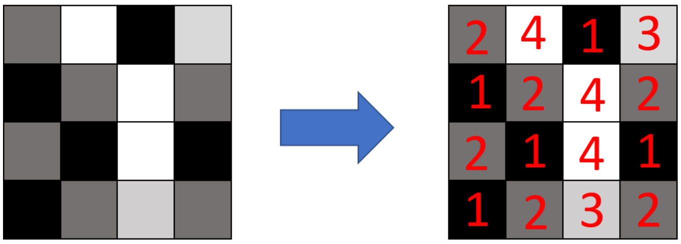
1) *First Order Statistics*
2) *Shape-based Features*
   as well as features based on the:
3) *Grey Level Co-occurrence Matrix*
4) *Grey Level Run Length Matrix*
5) *Grey Level Size Zone Matrix*
6) *Neighbouring Gray Tone Difference Matrix*
7) *Gray Level Dependence Matrix.*

Some of these classes are trivial. For example, first order statistics measures values such as the mean of voxel intensities and variance. Other classes (in particular 3-7 in the list above) are less obvious and characterise more subtle characteristics of the tumour such as its texture and homogeneity. A complete description of these classes is beyond the scope of this report but can be found in the documentation of the computational radiomics package described in [14]. Instead we discuss the two classes with features of importance to this experiment; the *Grey Level Size Zone Matrix* and the *Grey Level Co-occurrence Matrix*.

2.3.1 Grey Level Size Zone Matrix

The grey level size zone matrix was first proposed as a method of characterising the homogeneity of a texture [25]. Consider for example the greyscale image shown in Figure 3. The matrix representation of this image is

$$\begin{pmatrix} 2 & 4 & 1 & 3 \\ 1 & 2 & 4 & 2 \\ 2 & 1 & 4 & 1 \\ 1 & 2 & 3 & 2 \end{pmatrix} \tag{5}$$



2	4	1	3
1	2	4	2
2	1	4	1
1	2	3	2

Figure 3. An example greyscale image shown represented as matrix. Each pixel value is shown in red.

The grey level size zone matrix of the above matrix is

$$G = \begin{pmatrix} 0 & 0 & 2 & 0 \\ 1 & 0 & 0 & 1 \\ 2 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}. \tag{6}$$

Here, the element G_{ij} is the number of inter-connected volumes of pixels at grey level intensity i of size j in the matrix. Pixels are defined to be connected to each other if they are linked either diagonally, horizontally or vertically. For example the matrix in equation 5 has one interconnected volume of pixels at grey level 1 of size 4 and one of size 1. Therefore

$$G_{11} = G_{14} = 0. \tag{7}$$

It is important to note that the above arguments generalise easily to 3-dimensional greyscale images, such as CT scans, where pixels are replaced by 3-dimensional *voxels*.

2.3.2 Grey Level Co-Occurrence Matrices

The other class of radiomic features relevant to this study are statistics calculated using *co-occurrence matrices*. A co-occurrence matrix, $M(\delta, \theta)$ for an image is a matrix of size $N \times N$ (where N is the number of discrete intensity levels in the image). The element $M_{ij}(\delta, \theta)$ counts the number of times that elements of intensity levels i and j occur a distance δ from each other at an angle θ [26]. For example the co-occurrence matrix for image in Figure 3 with parameters $\delta = 1$ and $\theta = 0^\circ$ is

$$M(1, 0) = \begin{pmatrix} 0 & 3 & 1 & 3 \\ 3 & 0 & 2 & 0 \\ 1 & 2 & 0 & 0 \\ 3 & 0 & 0 & 0 \end{pmatrix}. \tag{8}$$

This co-occurrence matrix relates pixels located both horizontally ($\theta = 0^\circ$) and adjacent to each other ($\delta = 1$) . Note that by the definition, the co-occurrence matrix is always symmetric.

Like the grey level size zone matrix, the co-occurrence matrix generalises trivially to three dimensional images. For $\delta = 1$ there are two neighbours to each pixel for each of the 13 possible angles in 3-dimensions [14].

Co-occurrence matrices are frequently used in texture analysis, in particular for image segmentation and recognition. For example, metrics based on co-occurrence matrices have been used to classify rock type [27], face recognition [28] and land-cover classification [29].

2.4 Fractal Analysis and Lacunarity

In this report fractal features of lung cancer tumours are also explored. It has been shown in several studies that these features can act as an indicator for a range of behaviours in cancer [30]. Specifically, the *fractal dimension* and *lacunarity* of tumours are explored. These features are briefly outlined below.

2.4.1 Fractal Dimension

The fractal dimension of an image is a quantitative parameter which describes how an image's complexity varies with scale [31]. It is most popularly defined and measured by the *box-counting method*. This is illustrated in Figure 4.

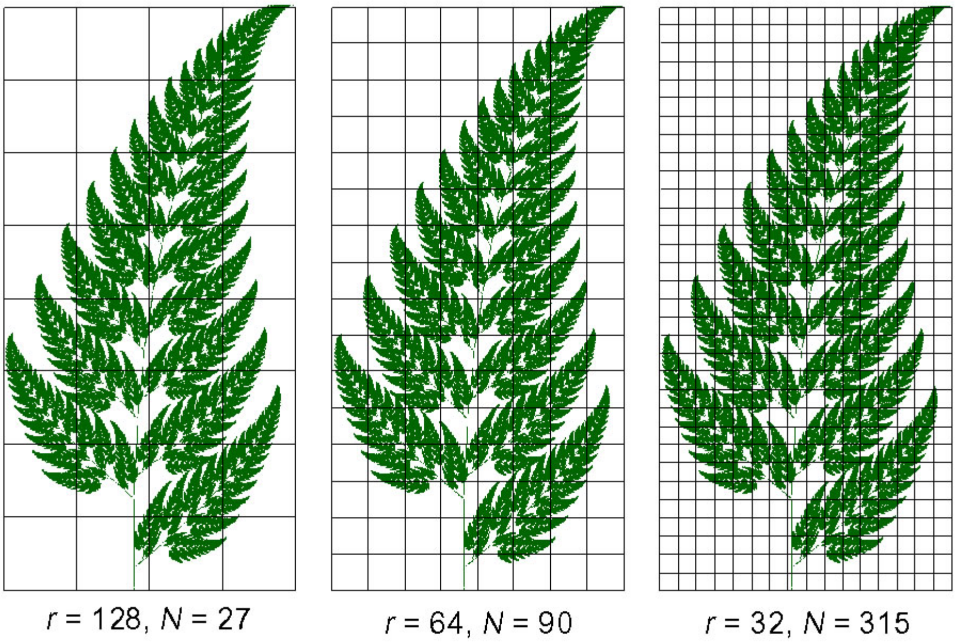


Figure 4. A figure showing the box-counting method of determining a fern leaf's fractal dimension, also known as the Minkowski–Bouligand dimension. [32]. This figure is adapted from [33]. Here r represents the length of the boxes used and N represents the number of boxes required to cover the fern.

For Figure 4, the number of boxes used to cover the fern is plotted as a function of the length of the overlaid boxes. A linear regression is performed and the gradient of the fitted line is the fractal dimension of the fern. Formally, the box-counting fractal dimension,D_B, is defined as

$$D_B = \lim_{r\to 0} \frac{\log N(r)}{\log r}. \tag{9}$$

Note that this equation only applies to a binary image. We can generalise the box-counting fractal dimension, D_B for a greyscale image such as a CT scan by summing the mean intensity in each box instead of the number of boxes the shape appears in,

$$D_B = \lim_{r\to 0} \frac{\log I(r)}{\log r}. \tag{10}$$

where $I(r)$ is the sum of the average pixel intensity in each box.

2.4.2 Lacunarity

Another similar yet distinct feature to fractal dimension is an image's *lacunarity*. Lacunarity, deriving from the Latin word *lacuna* meaning 'gap' or 'lake' is a measure of how an object fills space, with patterns containing more or larger gaps having higher lacunarity. In this study, lacunarity is determined quantitatively by a box-counting method. A set of N different box sizes is defined and indexed by n. The lacunarity at the box size denoted by n, λ_n, is given by

$$\lambda_n = \left(\frac{\sigma_n}{\mu_n}\right)^2, \tag{11}$$

where σ_n is the standard deviation in the boxes' mean intensities and μ is the mean of boxes' mean intensities. The total lacunarity, Λ, is then defined to be

$$\Lambda = \sum_{n=1}^{N} \frac{\lambda_n}{N}. \tag{12}$$

2.5 Feature Selection

As previously stated, this report attempts to formulate the problem of determining whether tumours are fragmenting or non-fragmenting as a canonical machine learning problem. Here an attempt to learn a function, F, to map a set of quantitative *features*, $\mathbf{X}$ onto a set of *labels*, $\mathbf{y}$;

$$F : \mathbf{X} \mapsto \mathbf{y}. \tag{13}$$

Here, for each patient, the features are radiomic and fractal metrics describing the patients scan series and the label is a binary value determining whether a tumour is fragmenting or not.

The methodology used results in a wide range of features being extracted from each patient, most of which are likely to be uninformative as to whether a tumour is shrinking or eroding. As such, a key challenge is to determine which extracted features are important in predicting the label. Several feature selection algorithms were applied in order to determine these important features. Ultimately, *greedy stepwise* feature selection was used; this is described in section 2.5.1. A key idea behind this feature selection is the average *leave one out cross validation* (LOOCV) score. This is defined as follows:

If we have n data points in our dataset, then the model is trained n times, each time leaving out a different data point. Once the model has been trained on the remaining $n-1$ points, it is evaluated on the remaining point. The error of the evaluation is the LOOCV score for the left-out data point. The average of this value across all points in the dataset is the average LOOCV score [34]. Note that the type of model used here and the error metric is left unspecified as cross-validation is general to both these choices in a machine learning framework.

2.5.1 Greedy Stepwise Feature Selection

In this algorithm, a large feature set with several redundant features, $\mathbf{X}$, and its associated labels $\mathbf{y}$ is reduced to a smaller dataset with the redundant features used, $\mathbf{X_r}$. The algorithm is as follows [35]:

- The feature and label set is shuffled and split in a test and training dataset.
- A one-feature model is trained for each feature on the training data and its average LOOCV is calculated, as previously described.
- The features are ranked in terms of the mean LOOCV score of their associated model from lowest to highest.
- Several models are then trained; one using the two features with the lowest mean LOOCV scores as input, one using the three features with the lowest mean LOOCV score, etc.
- Once the models have been defined, their mean LOOCV score is calculated.
- The features of the model with the lowest mean LOOCV score are selected as 'important' by the algorithm and their values for each data point become the reduced feature set $\mathbf{X}_r$.

Whilst not guaranteed to provide the set of features of all possible sets of features with the lowest LOOCV score, greedy stepwise evaluation provides a close approximation to the bulk search method whilst saving significantly on computational cost [36].

3 METHOD

3.1 Dataset Used

This study used a dataset consisting of CT and CBCT scans obtained from lung cancer patients during radiotherapy courses at the Christie NHS Foundation Trust in Manchester. Specifically, scan sets of patients undergoing two different types of treatment plan were used:

1) 8 non-small cell lung cancer (NSCLC) patients
2) 22 small cell lung cancer (SLC) patients

Treatment courses for all patients lasted from 4 to 46 days with corresponding scan series containing between 3 and 9 scans. Patients were scanned on an approximately bi-weekly basis.

Each patient was given a treatment planning (TP) scan on admission to the centre. The tumour and other key anatomical features were delineated manually by a clinician. An example of a delineated scan is shown in Figure 5.

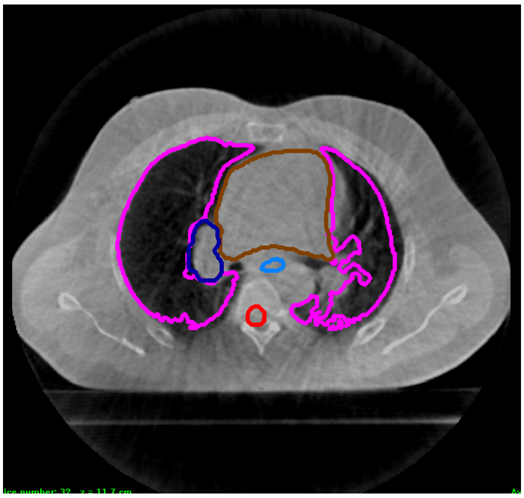


Figure 5. A slice of a delineated treatment planning scan from the dataset used. Several key anatomical features are delineated here; the internal treatment volume (ITV) is shown in navy blue) and the lung walls are shown in purple. Also delineated are the oesophagus (light blue), heart (brown) and spinal column (red). This figure is adapted from [1].

3.2 Generating Simulated Data

From the initial treatment planning scan of each patient in the dataset two simulated scan series were generated; one series corresponded to a fragmenting tumour and the other to a non-fragmenting tumour. This is illustrated in Figure 3.2. The algorithms used to generate these simulated scan series are described in sections 3.2.1 and 3.2.2.

3.2.1 Generating non-fragmenting tumours

Figure 6 shows a single iteration of the algorithm used to simulate non-fragmenting shrinking tumours. The action can be described as follows:

- In step A, the volume of the tumour (GTV) delineated by clinicians is extracted from the scans. The image is then converted into binary form, setting the intensity of all voxels above 500 CBCT units (units of greyscale voxel intensity in cone beam computed tomography scans) to 1 and all voxels below this threshold to 0. The resulting image is shown in step B in Figure 6. This is intended to isolate all voxels corresponding to healthy tissue.

- Small objects are removed from the image and edge detection is applied to the resulting image (step C). This is achieved using a Canny edge detector [37]. The exact details of how this are achieved are outlined in [1].
- Steps D and E show the resulting region isolated in step C being subtracted from the original tumour. The empty region created is then replaced by random Gaussian noise with a mean and variance corresponding to the mean and variance of all the non-zero pixels in the image below the threshold of 500 HU. The resulting simulation of the tumour after a single iteration of the algorithm is shown in step F.

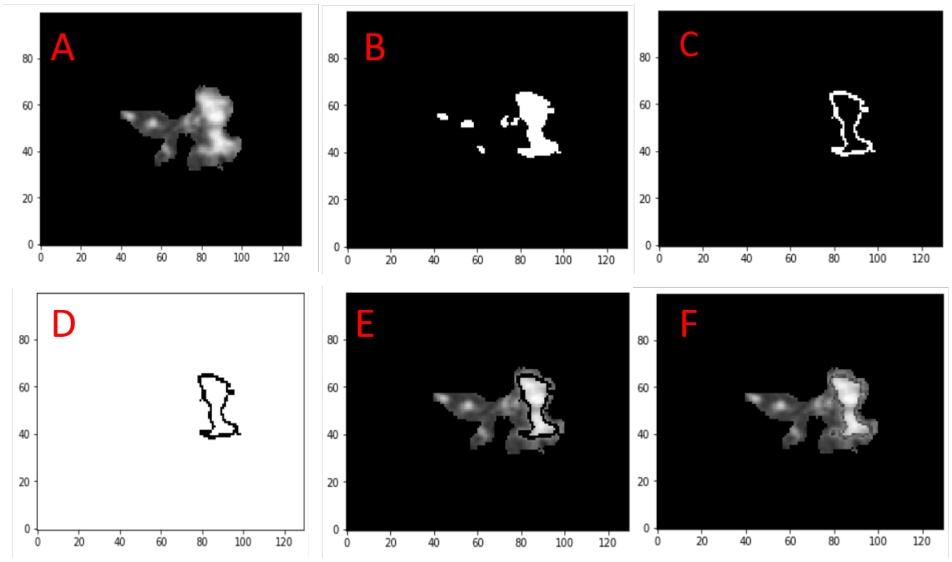


Figure 6. A figure showing the action of a single iteration of the algorithm used to simulate non-fragmenting tumours.

A simulated scan series generated from the tumour is shown in Figure 7. As expected, the tumour can be seen to be shrinking uniformly without fragmentation.

3.2.2 Generating fragmenting tumours

In order to develop simulations of fragmenting tumours, voxels were removed from the edge of the tumour stochastically. For a given image $I(\mathbf{x})$ is the intensity of the voxel at position vector x. A new image, $G(\mathbf{x})$, is defined by

$$G(\mathbf{x}) = |\nabla I(\mathbf{x})| \,. \tag{14}$$

An iteration of the algorithm used is illustrated in Figure 8. Steps A-C of the fragmenting algorithm defined in section 3.2.1 are applied to the CT scan. However, instead of being 1 voxel in radius, the ring around the tumour is 4 voxels. The selection criteria for which voxels in this region are replaced by 'healthy' tissue are specified as follows: For a given voxel in the

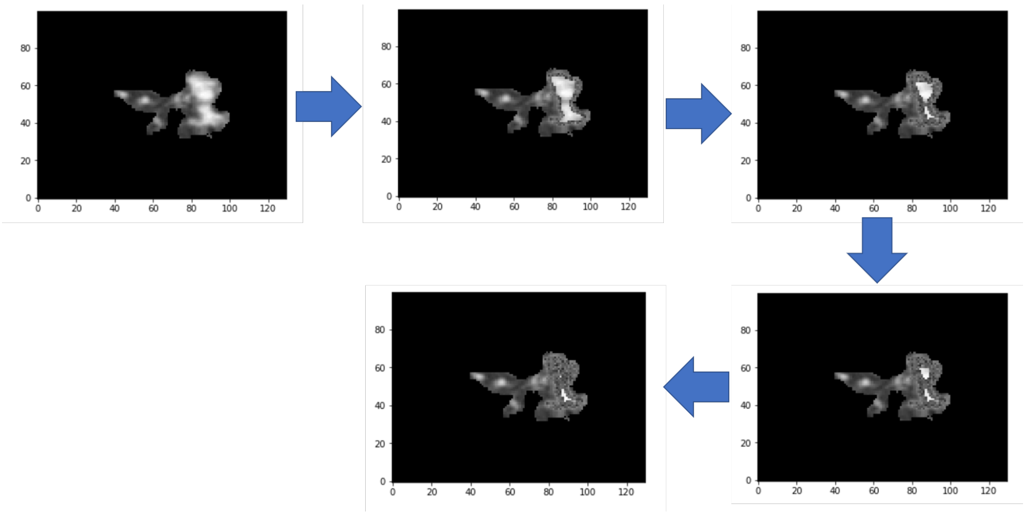


Figure 7. A simulated scan series generated from the tumour shown in Figure 6.

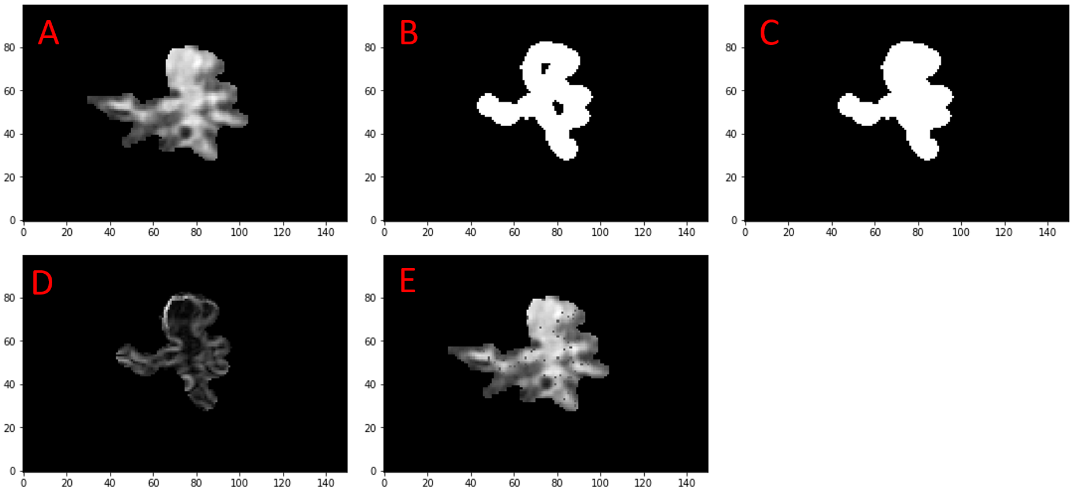


Figure 8. A graphic showing the main steps in the algorithm used to simulate scan series for fragmented tumours

region around the tumour, indexed by i, a random number, r_i, is drawn, from a triangular probability distribution. This is defined by

$$p(x) = \begin{cases} \frac{10x}{3} & \text{for } 0 \leq x < 0.6 \\ 5(1-x) & \text{for } 0.6 \leq x < 1 \\ 0 & \text{otherwise.} \end{cases} \tag{15}$$

This distribution was preferred to a Gaussian as it does not produce negative numbers. If the random number drawn for voxel i, r_i satisfies the condition

$$r_i < \frac{G(\mathbf{x}_i)}{G_{max}}, \tag{16}$$

where $G(\mathbf{x_i})$ is the gradient of the image at voxel i and G_{max} is the maximum gradient value in the image, the voxel is replaced. Like in the non-fragmenting algorithm, the voxels are replaced with random Gaussian noise with a mean and variance corresponding to the mean and variance of all non-zero voxels in the image below a threshold of 500 HU. A resulting scan series generated by the non-fragmenting algorithm is shown in Figure 9.

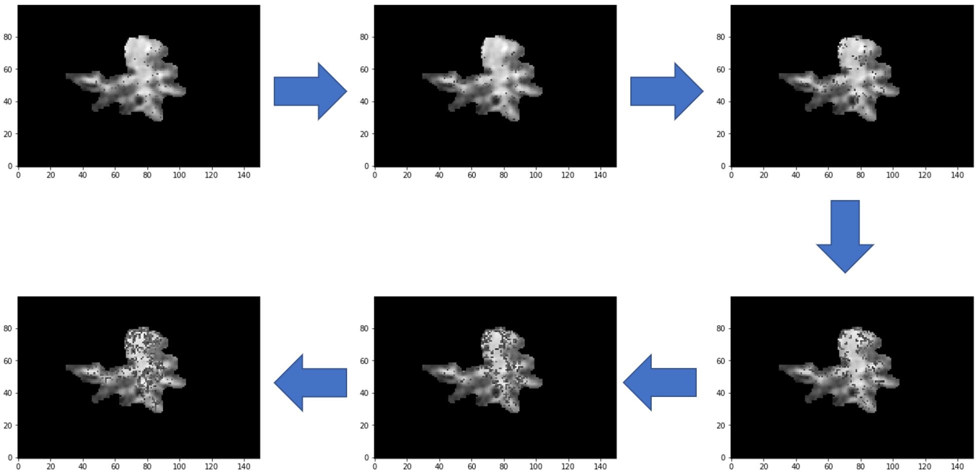


Figure 9. A slice of a simulated scan series generated by the fragmenting algorithm. Fragmentation can be seen as speckles on the tumour.

3.3 Radiomic and fractal feature extraction

Once the simulations of fragmenting and non-fragmenting tumours were generated, fractal and radiomic features were extracted from the scans. This was done using the `pyradiomics` [14] library for the radiomic features and `fraclac` [15] for ImageJ to extract the fractal features. On average, approximately 100 radiomic features were extracted per scan [14] as well as the mean of the box-counting lacunarity and fractal dimension across each 2-D slice of each scan. Once this was achieved, a straight line was fit through all the measured values of a feature through the scan series by simple linear regression. The resulting line was given by

$$\tilde{x} = g_x t + c_x, \tag{17}$$

where t is the number of days after the first scan, $\tilde{x}$ is the predicted value of feature x that time and g_x and c_x are free parameters to be determined. The gradients of these lines, $\{g_x\}$, were then used as features for each scan series. It should be noted that for the simulated data, the values of t for each scan corresponded to the most common radiotherapy plan in the dataset (scans on the 2nd, 4th, 8th and 10th day after the initial scan). Similar feature extraction was performed on the real scan series and the simulated scan series.

The next step was to determine a subset of these feature gradients which were key indicators as to whether a tumour was fragmenting or not. This was achieved by applying a greedy stepwise feature selection method as described in section 2.5.1 to the simulated scan series, using whether a given scan series was generated by the fragmenting or non-fragmenting algorithm as a label set.

4 Results

The feature selection chose four features as important. Specifically, three radiomic features as important, labelled by pyradiomics as:

- `original_glszm_GrayLevelNonUniformityNormalized`
- `original_glszm_SizeZoneNonUniformityNormalized`
- `original_glcm_Idmn`.

The box counting fractal dimension was also chosen as important. A discussion of the meaning of these features is given in section 5.

Principal component analysis was then performed on the selected features, compressing the feature set into two dimensions. This method is introduced in Appendix A. The resulting plot is shown in Figure 10.

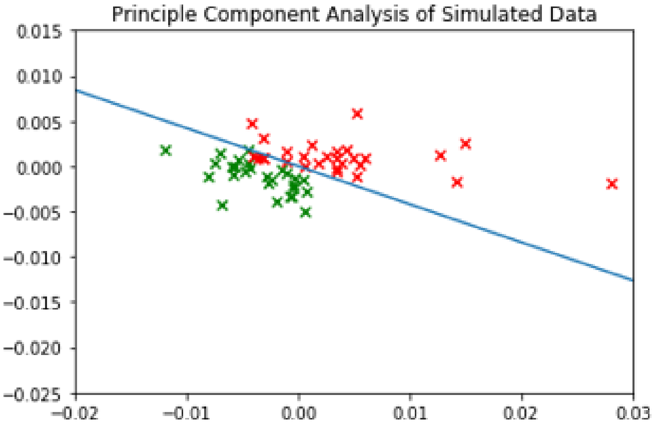


Figure 10. Principal component analysis of the simulated data features. Each point represents a different simulated scan series. Points plotted in red are simulated fragmenting tumours, where as points plotted in green are non-fragmenting tumours. A line roughly separating the two modes of shrinkage is shown in blue.

The variance of the dataset in the x- and y-axes are 4.06×10^{-5} and 4.03×10^{-6} respectively. A clear separation between the two modes of shrinkage simulated can be seen, indicating significant differences in the probability distributions of the important features. The same co-ordinate transform used to generate the co-ordinates used in Figure 10 was applied to the real dataset. The resulting plot is shown in Figure 11. It is interesting to note the reduced variance in scatter of the points for the real data in comparison with the simulated data. Potential interpretations of this are discussed in section 5.

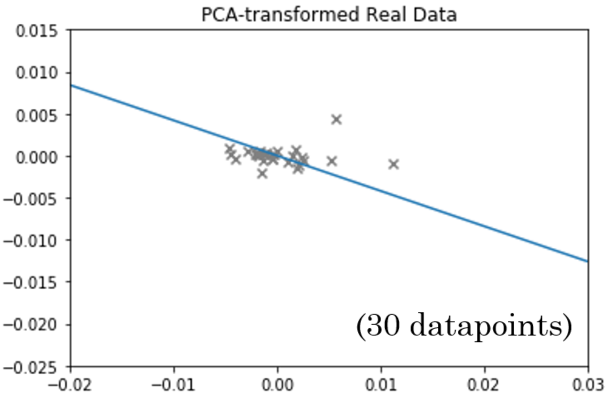


Figure 11. The feature set of the real data plotted in the co-ordinate system defined by the principal component analysis shown in Figure 10. The same line separating the two modes of shrinkage in Figure 10 is shown here in blue.

For Figure 11, the variances in the x- and y-axes are 1.06×10^{-5} and 1.03×10^{-6} respectively. The line shown in Figures 10 and 11 shows one possible classifier for determining whether a simulated tumour was shrinking or eroding. However, in the principal component analysis compression to two dimensions a large amount of information is discarded. An attempt was made to generate a classifier using the four important features specified at the beginning of this section. Specifically, a random forest classifier (see Appendix B) was trained on the simulated data and used to predict whether the real scans were fragmenting or non-fragmenting. The resulting classifier was found to have a 5-fold cross-validation error of 0.965 over the simulated scans. As only 57 simulated scans were used in this study, it is difficult to better determine the accuracy of this classifier without more data from which to generate simulations of tumour shrinkage.

5 Discussion

5.1 Significance of Extracted Features

As stated in section 4, three radiomic features were chosen by the stepwise evaluation as well as the fractal dimension. Of the three radiomic features chosen, two were based on the *grey level size zone matrix* of the image and one was based on the *grey level co-occurrence matrix.* These matrices are defined in section 2.3.1. The exact metrics classified as important were the normalised *grey level non-uniformity* and *size zone non-uniformity* of the size zone matrix and the normalised *inverse difference moment* of the co-occurrence matrix.

The grey level non-uniformity of the size zone matrix, or GLNN, is defined to be

$$GLNN = \frac{\Sigma_{i=1}^{N_g}\left(\Sigma_{j=1}^{N_s}\mathbf{P}(i,j)\right)^2}{N_z^2} \tag{18}$$

where N_g is the number of discrete intensity values in the image, N_s is the number of size zones in the matrix and N_z is the number of voxels in the region of interest (the tumour GTV). This is often interpreted to measure of the grey level uniformity in an image. The other extracted feature is the normalised size zone non-uniformity, or SZNN, of the size zone matrix. This is very similar to the GLNN and is given by

$$SZNN = \frac{\Sigma_{i=1}^{N_s}\left(\Sigma_{j=1}^{N_g}\mathbf{P}(i,j)\right)^2}{N_z^2}. \tag{19}$$

Notice here that the indices of the summation in equation 18 have been swapped. It is interesting to note that these two apparently similar statistics are uncorrelated enough to make them both useful features.

The other important feature was the normalised inverse difference moment (IDMN), a feature based on co-occurrence matrices. This is defined by

$$IDMN = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\frac{p(i,j)}{1+\left(\frac{|i-j|^2}{N_g^2}\right)}. \tag{20}$$

Here $p(i,j)$ is the mean of elements at position (i,j) of the normalised co-occurrence matrices with parameter $\delta = 1$. This can be considered another measurement of local homogeneity across an image.

The exact nature of the features considered important is not particularly significant. Many variables in the classes are similar and are therefore highly correlated with each other due to being based on the same matrix. Therefore, it is possible that several features in this class would act as descriptors of fragmentation. It is more interesting to note the classes from which these features are derived, specifically the size zone matrix, co-occurrence matrix for the radiomic features and the fractal dimension for the fractal features. It is not surprising that all of the selected features are normalised as the simulated scan series have significant variation in tumour size and fragmenting tumours of all size were made. It should also be noted that all of the selected radiomic features are traditionally associated with texture analysis. This implies that this is the best method to determine which mode of shrinkage is occurring in a simulated tumour.

5.2 Similarity to Real Data

In order to determine whether the simulations were accurate representations of the way tumours behaved, the distribution of features among the simulated and real data were compared. One measure of similarity is observing how the features appear in Figures 10 and 11. If the scattering is similar, both datasets can be assumed to belong to the same distribution and thus the simulations are accurate representations of reality.

It can be seen there is a some difference in the scatter of the datapoints. Specifically, there is a slightly larger variance in simulated data than in the real data in both axes of the principal component analysis. However, it is encouraging that the centroids of both datasets are located at the origin. This likely indicates that the simulated modes of shrinkage are more exaggerated than the modes which appear in reality. This is perhaps not surprising as it has been established that fragmentation is too microscopic to be observed by a CT scan and yet is readily observable in the simulation scans.

The other possibility not accounted for in the simulations is that different parts of a tumour may undergo different modes of shrinkage. Furthermore, the degree of fragmentation is much more likely to be continuous than a binary choice as presented here. More sophisticated simulation algorithms would be required to represent these phenomena.

5.3 Potential Future Work

The framework outlined here represents a promising way to explore the degree to which fragmentation occurs in tumours which, as discussed earlier, could have serious implications on the safety of image-guided radiotherapy. Whilst the work here is specific to lung cancer, in principle the same method could be applied to any localised tumour. However, several key weaknesses of the methodology have been discussed already. The focus now turns to how these limitations could be reduced in future work.

Firstly, as discussed in section 3.3, the algorithm currently uses linear regression to fit a straight line through the features with time. This assumes that the rate at which features change doesn't vary, discarding a great deal of information about the each feature's time series. It is likely that better results could be obtained by a more detailed analysis of the individual feature time series. Trivially, higher degree polynomials could be fit to the data. Another approximation made relating to the time series was that the simulated data did not have varied rates of shrinkage which could easily be implemented.

Another concern was that the fragmentation observed in the simulations was visible as can be seen in Figure 9. Previous research [10] has shown that fragmentation is too microscopic to be observed in CT scans, so this is an obvious over-simplistic assumption in the method. A potential solution to this would be generate the scan series at a higher resolution than the original data and then compress it to the same resolution as the scan series for feature extraction. This way microscopic fragmentation could be accounted for and its effect could be observed. The changes in features extracted would likely be more subtle and the discrepancy in variance between figure 10 and Figure 11 would be reduced.

A final possibility would be to link radiomic and fractal analysis with clinical data concerning whether a patient had a relapse at a later date. This is based on the assumption that patients with relapses of tumours in similar areas likely had fragmenting tumours left untreated. Such a dataset would be difficult to compile but would likely be very useful in this area of study.

6 Summary and Conclusion

In this report, a method for determining the degree of fragmentation in lung cancer tumours is discussed. This involves generating two sets of simulations of a based on a dataset of CT scans of lung tumours throughout treatment; one simulation set contains tumours fragmenting as they shrink and in the other set the tumours are not fragmenting. These simulations are not biologically motivated but rather match to a prespecified definition of fragmentation. The simulated scan series are analysed and a wide range of radiomic and fractal features are extracted. Feature extraction algorithms are used to determine which features are the best indicators of fragmentation. These are then used to generate a classifier which is shown to be very accurate in determining which mode of shrinkage is occurring, with a 5-fold cross validation of 96.5%.

A similar feature extraction is applied to the real scan series and a similar distribution of features is observed in the dataset although greater variance is observed in the simulated data (see Figures 10 and 11). It is concluded that the simulated algorithms have room for improvement and this could yield a more accurate classifier of fragmentation. Several potential improvements to these algorithms are then suggested.

The results of this report show that computational texture analysis is a strong candidate for properly determining whether fragmentation is occurring in tumours. This could have a significant impact on image-guided radiotherapy. Furthermore, the simulation algorithms used could likely be improved using inputs from oncology and biochemistry and could also inform research in these fields.

7 Acknowledgements

I'd like to thank Professor Anna Scaife and Dr. Alan McWilliam for designing and jointly supervising this project. I'd also like to express my gratitude Dr. William Beasley for several useful discussions and suggestions. Finally thanks goes to George for being a fantastic laboratory partner for the last two years and also for buying most of the coffee.

References

[1] Needham, G.R. & Windsor, R.L. (2018) 'Image-based data mining to analyse shrinkage versus erosion of lung cancer tumours treated with radiotherapy' *Master's project Report*, The University of Manchester, UK.

[2] (2011) Toxicity. In: Schwab M. (eds) Encyclopedia of Cancer. Springer, Berlin, Heidelberg.

[3] Chen, G.T.Y., Sharp, G.C. & Mori, S. Radiol Phys Technol (2009) 2: 1. https://doi.org/10.1007/s12194-008-0045-y

[4] Xing, L., Siebers, J. and Keal, P. (2007) "Computational Challenges for Image-Guided Radiation Therapy: Framework and Current Research" *Semin. Radiat. Oncol.*, 17 pp. 245-257.

[5] Jaffray D.A. et al (2008) "Applications of image processing in image-guided radiation therapy" *Medica Mundi*, 52(1), pp. 32-39.

[6] Schulz-Wendtland, R. et al. "Semi-automated delineation of breast cancer tumors and subsequent materialization using three-dimensional printing (rapid prototyping)." *J. Surg. Oncol.*,115(3), pp. 238-242.

[7] Ballangan, C. et al. (2011) "Automated Delineation of Lung Tumors in PET Images Based on Monotonicity and a Tumor-Customized Criterion", *IEEE Transactions on Information Technology in Biomedicine*, 15(5), pp. 691-702.

[8] Leibfarth, S. et al. (2015) "Automatic delineation of tumor volumes by co-segmentation of combined PET/MR data" *Physics in Medicine and Biology*, 60(14), pp.

[9] Kawata, Y. et al. (2017) "Impact of pixel-based machine-learning techniques on automated frameworks for delineation of gross tumor volume regions for stereotactic body radiation therapy" *Physica Medica*, 42, pp. 141-149.

[10] 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

[11] Cancer Research UK, http://www.cancerresearchuk.org/health-professional/cancer-statistics/diagnosis-and-treatment, Accessed 19/04/2018

[12] López Rodríguez, M., Cerezo Padellano, L. (2007) "Toxicity associated to radiotherapy treatment in lung cancer patients." *Clin Transl Oncol.*, 9(8), pp. 506-512.

[13] (2014) Computed Tomography. In Mikla, V. & Mikla, V. Medical Imaging Technology. Elsevier. Amsterdam.

[14] Griethuysen, J. J. M., Fedorov, A., Parmar, C., Hosny, A., Aucoin, N., Narayan, V., Beets-Tan, R. G. H., Fillon-Robin, J. C., Pieper, S., Aerts, H. J. W. L. (2017). Computational Radiomics System to Decode the Radiographic Phenotype. Cancer Research, 77(21), pp. 104-107.

[15] Karperien, A. (2001). FracLac for ImageJ; JavaDoc, source code, and jar, (Version 2.5) [Software]. Albury, NSW: Charles Sturt University. Retrieved October 1, 2011. Available from: US National Institutes of Health.

[16] Gourion, D. and Noll, D. (2002) "The inverse problem of emission tomography" *Inverse Problems*, 18(5), pp. 1435–1460.

[17] Kirsch, A. An Introduction to the Mathematical Theory of Inverse Problems, Springer, Applied Mathematical Sciences Series Vol. 120, 1996

[18] Lambin et al. (2012) "Radiomics: Extracting more information from medical images using advanced feature analysis" *European Journal of Cancer*, 48(4) pp. 441-446

[19] Aerts, H.J.W.L. et al. (2014) *Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach*, Nature Communications, Volume 5, Article 4006

[20] Kourou et al. (2015) "Machine learning applications in cancer prognosis and prediction" *Computational and Structural Biotechnology Journal*, 13, pp. 8-17

[21] Parmer, C. et al (2015) '"Machine Learning methods for Quantitative Radiomic Biomarkers" *Scientific Reports*, 5, Article number: 13087

[22] Lao, J. et al (2017) "A Deep Learning-Based Radiomics Model for Prediction of Survival in Glioblastoma Multiforme", *Scientific Reports*, 7, Article number: 10353

[23] Leger, S. et al (2017) "A comparative study of machine learning methods for time-to-event survival data for radiomics risk modelling", *Scientific Reports*, 7, Article number: 13206

[24] Wang, J. et al. (2017) "Machine learning-based analysis of MR radiomics can help to improve the diagnostic performance of PI-RADS v2 in clinically relevant prostate cancer", *European Radiology*, 27(10), pp. 4082–4090

[25] Thibault, G. et al. (2009). "Texture Indexes and Gray Level Size Zone Matrix. Application to Cell Nuclei Classification" Paper presented at the 13th International Conference on Pattern Recognition and Information Processing (PRIP), Minsk, Belarus.

[26] Haralick, R., Shanmugan, K. & Dinstein, I. (1973) 'Textural features for image classification' *IEEE Transactions on Systems, Man and Cybernetics*, SMC-3(3), pp. 610-621.

[27] Partio, M., Cramariuc, B., Gabbouj, M. and Visa, A. (2002) 'Rock texture retrieval using gray level co-occurrence matrix' [Online] Available at: https://www.researchgate.net/publication/228861769_Rock_texture_retrieval_using_gray_level_co-occurrence_matrix

[28] Eleyan, A. & Demirel, H. (2011) 'Co-occurrence matrix and its statistical features as a new approach for face recognition' *Turk J Elec Eng & Comp Sci*, 19(1), pp. 97-107.

[29] Marceau, D. J. et al. (1990) 'Evaluation of the grey-level co-occurrence matrix method for land-cover classification using SPOT imagery.' *IEEE Transactions on Geoscience and Remote Sensing*, 28(4), pp.513-519.

[30] Lennon, F.E. (2015) "Lung cancer— a fractal viewpoint",*Nat Rev Clin Oncol*, 20(11), pp. 664-675

[31] Da, Costa Fontoura Luciano, and Roberto Cesar Marcondes. *Shape Analysis and Classification: Theory and Practice*. CRC Press, 2001. pp. 442-446.

[32] Schroeder, M. Fractals, Chaos, *Power Laws: Minutes from an Infinite Paradise*. New York: W. H. Freeman, pp. 41-45, 1991

[33] Introduction to Fractal Dimension https://web4.wzw.tum.de/ane/dimensions/subsection3_4_2.html [Online] Accessed 08/05/18

[34] Rasmussen, C.E. & Williams, C.K.I. *Gaussian Processes for Machine Learning*, the MIT Press, 2006, pp. 111-112.

[35] Deng, K. (1998) 'Omega: On-Line Memory-Based General Purpose System Classifier', PhD thesis, Carnegie Mellon University, Pittsburgh PA, pp. 119-124.

[36] John, G.H., Kohavi, R. & Pfleger, K. (1994) "Irrelevant Features and the Subset Selection Problem" Paper presented at the International Conference of Machine Learning 1994, New Brunswick, NJ.

[37] Canny, J. (1986) "A Computational Approach To Edge Detection", IEEE Transactions on Pattern Analysis and Machine Intelligence 8(6) pp. 679–698

[38] Goodfellow, I., Bengio, Y. and Courville, A. (2016) "Deep Learning", *MIT Press*, Cambridge, MA, pp. 45-50.

APPENDIX

.1 Principal Component Analysis

The following description of principal component analysis is a summarises that given in "Deep Learning" by Ian Goodfellow and Yoshua Bengio [38]. The task is to map the set of vectors $\{\mathbf{x}_0, \mathbf{x}_1, ..., \mathbf{x}_n\}$ in $\mathbb{R}^l$ onto the set $\{\mathbf{y}_0, \mathbf{y}_1, ..., \mathbf{y}_n\}$ in $\mathbb{R}^q$ such that $q < l$. A given member of the first set, $\mathbf{x_i}$, is then recovered as accurately as possible from the corresponding member in the second, $\mathbf{y_i}$, by matrix multiplication such that, ideally,

$$\mathbf{x_i} = \mathbf{D}\mathbf{y_i}, \tag{21}$$

where $\mathbf{D}$ is a *decoding matrix*. To ensure a simple and unique solution, $\mathbf{D}$ is constrained to have unit norm and orthogonal columns. This can be shown to imply that

$$\mathbf{y_i} = \mathbf{D^T}\mathbf{x_i}, \tag{22}$$

In order to get optimal recovery of $\mathbf{x_i}$ after encoding then decoding, $\mathbf{D}$ is defined to be the solution to the problem

$$\underset{\mathbf{D}}{\operatorname{argmin}} \sqrt{\sum_{j=1}^{n} ||\mathbf{x}_j - \mathbf{D}\mathbf{D^T}\mathbf{x}_j||_2^2}. \tag{23}$$

The solution of this problem gives us an encoding function $\mathbf{D^T}$ such that we can compress any vector in our dataset $\mathbf{x}_i$ into $\mathbf{y}_i$ with minimal information loss and then recover $\mathbf{x}_i$ with high precision by using the decoding function $\mathbf{D}$. Principal component analysis is particularly useful for compressing high dimensional feature sets into two dimensions so that they can be visualised in graphs.

.2 Random Forest Classifiers

Random forest classifiers are a ensemble generalisation of random tree classifier. A random tree classifier decides on the class of features using a decision tree. Suppose a dataset where a given data point is represented by the feature set $\mathbf{x} = (x_1, x_2, x_3)^T$ and the label $y \in \{a, b, c\}$. We want to find a classifier, $C : \mathbf{x} \mapsto y$ for all points in the dataset. Random trees model achieve this by making a series of binary choices as shown in Figure 12.

Random trees have a tendency to over fit to data. This issue is rectified in a random forest classifier, where several random trees classifiers are trained. Each tree on a different selection of the points from the dataset. The class of a point is then decided by a majority vote of all the trees. It should be noted that over fitting is also often reduced in practice by limiting the number of decisions a single tree can make.

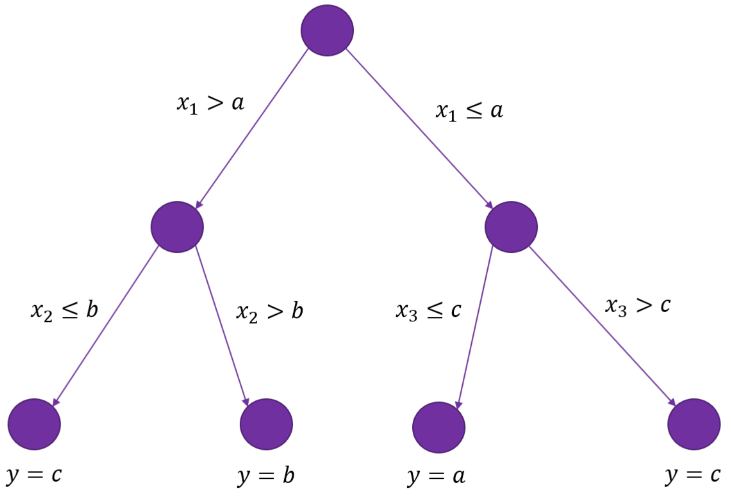


Figure 12. Starting at the top of the tree a series of binary decisions are made considering the value of a given feature. The feature selected to test at each node is the best separator of all the data points arriving at that note, as decided by an arbitrary objective function. As such, random trees attempt to separate the data into its possible classes (in this case a, b and c) in as few decisions as possible.