Using Machine Learning Applied to Radiomic Image Features for Segmenting Tumour Structures

Henry Clifton*, Alanna Vial*, Andrew Miller[§†], Christian Ritz*, Matthew Field[‡¶],
Lois Holloway[‡§¶], Montserrat Ros*, Martin Carolan[§†] and David Stirling*
* School of Electrical, Telecommunications and Computer Engineering, University of Wollongong, Wollongong, NSW, AUSTRALIA
‡ Liverpool and Macarthur Cancer Therapy Centres and Ingham Institute, Liverpool, NSW, AUSTRALIA
§ Centre for Medical Radiation Physics, University of Wollongong, NSW, AUSTRALIA
¶ South West Clinical School, University of New South Wales, NSW, AUSTRALIA
† Illawarra Cancer Care Centre, Wollongong, NSW, AUSTRALIA

Abstract— **Lung cancer (LC) was the predicted leading cause of Australian cancer fatalities in 2018 (around 9,200 deaths). Non-Small Cell Lung Cancer (NSCLC) tumours with larger amounts of heterogeneity have been linked to a worse outcome. Medical imaging is widely used in oncology and non-invasively collects data about the whole tumour. The field of radiomics uses these medical images to extract quantitative image features and promises further understanding of the disease at the time of diagnosis, during treatment and in follow up.**

It is well known that manual and semi-automatic tumour segmentation methods are subject to inter-observer variability which reduces confidence in the treatment region and extent of disease. This leads to tumour under- and over-estimation which can impact on treatment outcome and treatment-induced morbidity.

This research aims to use radiomic features centred at each pixel to segment the location of the lung tumour on Computed Tomography (CT) scans. To achieve this objective, a Decision Tree (DT) model was trained using sampled CT data from eight patients. The data consisted of 25 pixel-based texture features calculated from four Gray Level Matrices (GLMs) describing the region around each pixel. The model was assessed using an unseen patient through both a confusion matrix and interpretation of the segment.

The findings showed that the model accurately (AUROC = 83.9%) predicts tumour location within the test data, concluding that pixel based textural features likely contribute to segmenting the lung tumour. The prediction displayed a strong representation of the manually segmented Region of Interest (ROI), which is considered the ground truth for the purpose of this research.

I. Introduction

Cancer is a well known prominent cause of illness caused by the uncontrolled growth of mutated cells [1]. It contributed to three of every ten cancer related deaths in Australia with similar figures worldwide [2], [3]. Lung cancer (LC) was predicted to lead the most cancer related deaths in 2018 (9,198 deaths). Overall the five-year survival rate at 15.8%, is the 2nd highest mortality rate of the ten most commonly diagnosed cancers [2].

Radiomics is a field of research-based medical image analysis. It utilises computer vision techniques, Machine Learning (ML) and data mining to quantify insightful features [4], [5]. Several studies indicate that including these features from imaging data can lead to an improved predictive accuracy in prognostic models [5]–[8]. The field promises to contribute significantly toward personalised medicine.

This research aims to use radiomics to improve the post-diagnosis and treatment success of patients and subsequently increase the survival rate of LC. To achieve this, a model was developed focused on pixel based texture features.

Pixel based texture features are computed per pixel and produce a value that describes the relationship between the pixel and its surrounds. The model was trained by analysing all the pixels inside the clinician identified GTV and using these resultant values to train a DT. The outcome was to apply the classification of either 'tumour' or 'non-tumour' to each pixel within the slice. Once trained, the model requires no manual intervention. This data can then be displayed as a visible structure and compared statistically and visually to a clinicians delineation.

A simplified, trained auto-segmentation model can potentially assist clinicians in assessing the extend of disease, as well as providing an expert physician with a second opinion. Auto-segmentation models can provide opportunities to include more data to serve large data ML algorithms to further contribute to the radiomic analysis. This research uses pixel based textural features combined with machine learning to predict tumour structures in CT scans.

This paper is structured as follows, Section I introduces this work, Section II contains a discussion on current radiomic and ML techniques and literature, Section III outlines the method followed and key processes undertaken. Section IV outlines the outcomes of the applied model with selected inputs, Section V discussess the results and Section VI concludes this work.

II. Review of Radiomics Literature

Radiomics in oncology combines computer vision and machine learning with oncology data to devise a largely image-based personalised approach to medicine. Research suggests image features can provide additional information when predicting outcomes when compared to current methods [5]–[8]. The paper by Aerts et al. [6] presented 440 radiomic features across 1,019 patients that mapped a prognostic radiomic signature capturing intratumoural heterogeneity.

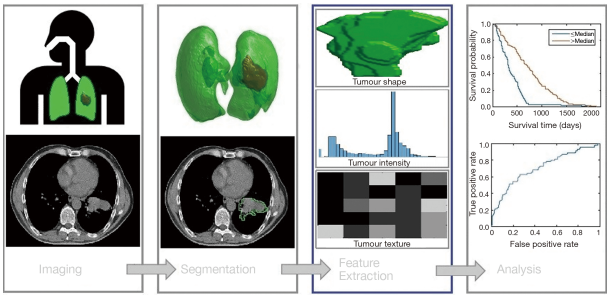


Fig. 1: Radiomics Workflow comprising of four key stages; Imaging, Segmentation, Feature Extraction, and Analysis [10].

Radiomics research aims to support doctors and patients interpret images in the diagnostic and prescription process. A computer-based decision support system (DSS) can consider many more factors that describe the patient and the disease. This method has been applied and successfully implemented to many other scientific fields [9] and has the potential to reduce misdiagnosis and improve treatment ineffectiveness.

The radiomics work flow comprises of four core pillars, Imaging, segmentation, feature extraction and analysis [6] as outlined in Figure 1.

Medical imaging captures an instance in patient anatomy or functional processes. Common imaging modalities used clinically include CT, Positron Emission Tomography (PET), and Magnetic Resonance Imaging (MRI). With regards to capturing anatomical information in diagnosing NSCLC, CT is routinely collected [11]. The most widely available complete data type is CT scans [12], [13], and is therefore is utilised in this research.

The Gross, or visible, tumour volume (GTV) is usually manually segmented by an oncologist for the purposes of radiotherapy. A GTV that is too small or in the wrong location can lead to 'geographical miss', and the failure of therapy to control the tumour. A GTV that is too large can lead to unnecessary normal tissue loss or damage, with subsequent unnecessary loss of quality of life and treatment-induced morbidity [14]. The manual delineation GTV and lack of clear information on the tumour introduces inter-observer variability [15]–[19].

Semi-automatic tumour segmentation software tools are available to assist oncologists with delineations [19]–[21] while no automatic methods for tumour segmentation are used clinically [10]. Existing radiomic studies [22]–[28], consider a large number of image features including GLM to predict patient survival. Success of GLMs within radiomic literature presents opportunities of its application in other avenues such as automatic tumour segmentation. Various techniques are suggested to contribute to an auto-segmentation process [15], [29]–[31]. A popular approach is patching, which breaks the image into smaller images to be processed. Pixel-based features capture the relationship between one and the surrounding pixels thus providing pixel specific features for every location within a patient. In this study, we assess if this information can contribute to a segmentation model.

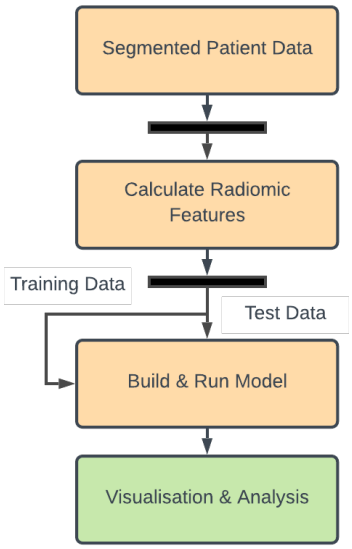


Fig. 2: Flow Chart outlining the method used for this research. Beginning with segmented patient data, each patient has radiomic features calculated. Following this training and testing data are used to build the model and predict outcomes respectively. Results are then visualised and assessed.

III. Methodology

The field of radiomic research has a core method proposed by Lambin et al. [32] and has been adopted as a common approach to comparing algorithms and features [4], [6], [8], [33]. This research is an adapted version of this approach and is summarised in Figure 2.

The process commenced with CT image data segmented by an expert physician to produce an ROI outlining the lung tumour. The image data bounded by the ROI had radiomic features analysed to establish the inherent characteristics of the disease. This radiomic data was sampled and then passed to a classifier which categorised each pixel as "tumour" or "non-tumour". The classifier predicted the outcome on an unseen test patient which was then compared to the manually segmented ROI.

The primary dataset used in this research was anonymised "NSCLCRadiomics" dataset [34] from the Cancer Imaging Archive (TCIA)[35] which contains 422 NSCLC pretreatment CT scans with GTV manually delineated by a radiation oncologist. The patients had varied demographics, tumour stage and tumour position. The dataset was first published with Aerts et al. [6] and was used to build their prognostic radiomic signatures.

A subset of this dataset includes 8 patients from NSCLCRadiomics that were delineated by a clinical radiation oncologist (AM). Subsequently, the subset dataset has been calculated with the same tools and techniques which minimise inter-observer variability between patients. This process aims to

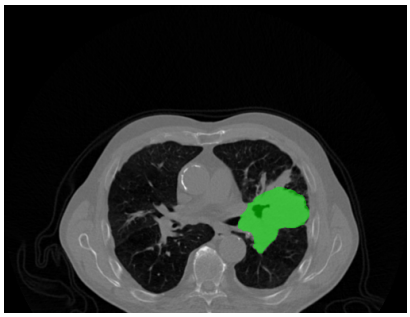


Fig. 3: Slice no. 76 of patient LUNG1-001

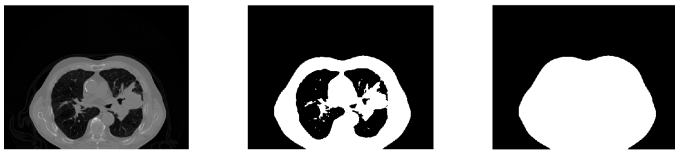
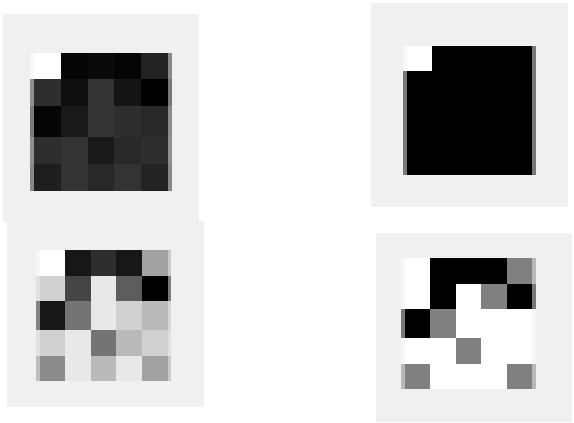


Fig. 4: Process of defining the patient mask

(a) Example 25 pixel image with outlier

(b) Image (a) reduced to 3 GLs



(c) Image (a) with tone removed

(d) Image (c) reduced to 3 GLs

Fig. 5: Visual comparison of texture loss due to tone

reduce variability in the technique between each GTV calculation.

Patient data was limited to a single slice originally at 512×512 pixels. Each image was masked and cropped to eliminate the bed captured within the CT. The subset of 8 patients are sampled at a ratio 1:5 (tumour:non-tumour) and were used as training data. The test case LUNG1-001 included all pixels of slice 76 is taken from the original dataset. The CT data overlayed with the original GTV for patient LUNG1-001 is displayed in Figure 3.

A. Radiomic Features

Radiomic features provides localised texture information that can be visualised and analysed with similar methods to the original CT scan. Previously published texture features [6] obtain a gray-level matrix and summate all data in accordance to various formulae, which results in a loss of anatomical detail.

B. Preprocessing

Preprocessing incorporates processes to convert the data from intensity (HU) to texture described as a specified interval of Gray Levels (GL). Preprocessing minimises computational time and complexity while providing an option to adjust detail resulting in diversity in feature images.

Firstly, unwanted pixel data is eliminated through masking then cropping. A simple binary mask approach is used to segment the patient scan from the surrounding background which improves the precision in future calculations. Figure 4 shows an example, based on a threshold of -400 HU.

The resulting data is split into texture and tone. Texture refers to the description of patterns or arrangement of intensities within an image while tone is a range of values from light (maximum HU) to dark (minimum HU) within the dataset. For the purpose of this research, tone refers to the range of unassigned HU values that are unused in GLM textural comparison. Tone is removed by reassigning pixels to a rank from one to the length of unique pixel values.

Large ranges of tone can lead to texture loss when adjusting the GL range as outlined in the example Figure 5. Figure 5c shows the image with tone eliminated leaving the original texture. Figure 5d shows this image reduced to three GL. The effect of reducing tone can be appreciated in the difference between Figures 5b and 5d.

Although discretisation methods such as binning are common for textural features, this research rescales the data to a specified GL range. Rescaling is a simple method and can be easily interchanged when assigning new ranges. Minor loss in detail occurs but key traits can still be mapped as appreciated in the comparison of Figures 5c and 5d. This process establishes a standardised GL range and GLM dimensions which optimises textural comparison and minimises the intervention of thresholding.

C. Gray-Level Matrices (GLM)

Each Matrix value captures the probability of a pixel relationship. Each matrix focuses on different characteristics of the image and provide insight into its texture. The four Gray Level Matrices that are developed in this paper to generate pixel based radiomic features include;

- Gray-level co-occurrence matrix (GLCM)
- Gray level run-length matrix (GLRLM)
- Gray level size zone matrix (GLSZM)
- Gray level dependency matrix (GLDM)

1) Gray-level co-occurrence matrix (GLCM): Initially proposed by Haralick [36], the GLCM is defined as the second order joint probability function of an image described as $\mathbf{P}(i, j; \delta, \alpha)$. The (i, j)th element of the $N_g \times N_g$ size matrix is the number of times that combination of intensity levels occurs in α direction at a distance of δ. where N_g represents unique GL.

This matrix combines opposing directions ($\alpha = 0^o$ = 180^o) halving the calculations from eight neighbouring pixels to four matrices ($0^o, 45^o, 90^o$ and 135^o). Each neighbouring pixel was calculated at a distance of $\delta = 1$.

A GLCM calculated at 5GL has a size of 5×5, such that $N_g = 5$. An excessively large GL range produces poor results as pixel relationships are rarely the same. Due to the size representing $N_g \times N_g$, images with larger GLs produce a excessive GLCM and subsequently will large computational costs and less occurrences of the same pixel relationship.

2) Gray-level run-length matrix (GLRLM): Gray Level Run Length Matrices was first proposed by Galloway [37]. A gray level run is the set of consecutive points with the same gray level value [38]. The matrix is a summation of the run length and gray level for (i, j)th elements in the image $\mathbf{P}(i, j; \alpha)$. The size $N_g \times N_r$, and can be calculated in α directions, where N_r represents largest run length. Four directions of α are calculated like the GLCM.

When GL is reduced, run lengths are more likely to increase, in turn depicting distinctions between rough and smooth textures.

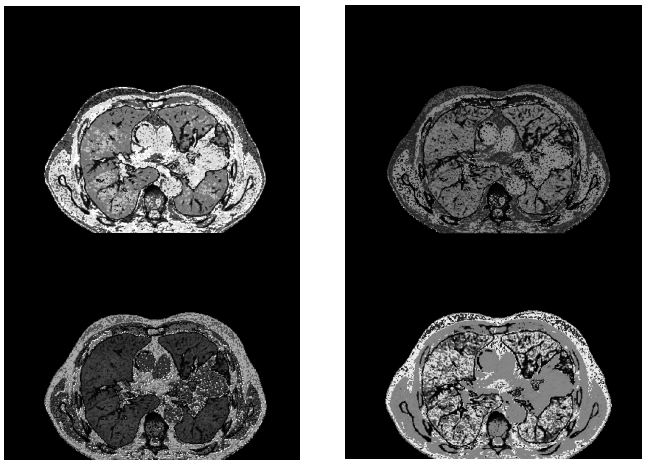
3) Gray-level size zone matrix (GLSZM): Gray Level Size Zone Matrix was first proposed by Thibault [39], utilising GLRLM fundamentals to produce a matrix representing all directions. Size zone is the count of elements with the same GL, connecting in any direction $\mathbf{P}(i, j)$. The matrix represents each GL and the count of masses with an exact size within the data. The matrix size $N_g \times N_s$, where N_s is the maximum size.

4) Gray-level dependency matrix (GLDM): First proposed as Neighboring Gray Level Dependence Matrix by Sun in 1983 [40], the GLDM is the summation of GLs with the number of connected pixels at a distance δ, dependent on the center pixel $\mathbf{P}(i, j; \delta)$. A pixel j is dependent on the center pixel i if $|i-j| \leq \alpha$. The matrix size is represented by $N_g \times N_d$, where N_d is the maximum dependency sizes. Commonly, in large images, $N_d = 9$, representing each direction surrounding the pixel.

D. Feature Image

Tumour segmentation requires pixel features and thus the GLM values are substituted into a feature image. Each pixel within the original image $I(x, y) = i$, was recalculated according to the GLM technique. The new feature image substitutes each pixel value i as the normalised GLM probability $P(i, j)$, such that $I_r(x, y) = P(i, j)$. Where $P(i, j) = \mathbf{P}(i, j)/\mu$ and μ is the GLM mean. For each GLM a respective feature image is calculated. E.g. The GLCM feature image results in a noisy image that highlights common pixel relationships and darkens the uncommon, which can be observed in Figure 6. Various directions provide diverse descriptors for each pixel in comparison to its surrounding pixels.

(a) GLCM-32GL-0deg (b) GLCM-32GL-90deg



(c) GLCM-32GL-45deg (d) GLCM-32GL-135deg

Fig. 6: GLCM feature image in 4 directions for 32 GL

E. Building and Testing the Model

The classification of features calculated was used to correlate with a numerical result or an outcome. Outcomes have a distinct medical value (survival rate, tumour stage, differentiation of tumour from non-tumour). To achieve this result, supervised learning compares training data to an outcome and creates a model that categorises inputs according to the outcome. Therefore, this outcome (tumour or non-tumour) is included with the data to compare projected estimation with the actual result.

The supervised learning for model building used the open source software tool Waikato Environment for Knowledge Analysis (Weka) 3.8.3 [41]. Weka is a platform used for data mining with various machine learning algorithms and this research used the data preparation, classification and visualisation tools.

Constraints for many ML algorithms, such as CNN and deep learning include limited explanations when discussing outcomes. Complex statistical or mathematical computations, act as a black box and have little meaning to medical professionals. Decision Trees (DT) are a common decision support tool that provide a simple explanation behind each outcome [42]. The algorithm comprises of observations or decisions, represented as branches and conclusions or outcomes represented as leaves. An input passes down the tree until reaching an outcome and when discussing results each decision can be tracked back through the tree.

A downfall of DT is overfitting, where a model trains on a dataset too closely and reacts poorly to new data. In DTs overfitting is represented with large complex trees. Pruning restricts the leaves and subsequently leads to a smaller more robust model. The DT pruning process limits the minimum number of objects per leaf, or training pixel classifications. Highly influential features contribute to large amounts of pixel classifications while less influential features are cut from the final model through pruning.

IV. Results

Several specifications within the patient data, radiomic features, and model characteristics are selected to assess the model and its result against the hypothesis.

A. Radiomic Features

This research includes four GLM features calculated in both 32 GL and 5 GL with the GLCM and GLRLM further calculated in four directions and averaged. The HU intensity, and the GTV pixel outcome are also included. Table I provides all 25 features calculated for each patient image, which are then used to train the decision tree.

TABLE I: Features imported into the model

Features		
HU	GLRLM-32GL-45deg	GLRLM-32GL-90deg
GLCM-32GL-0deg	GLRLM-32GL-135deg	GLCM-5GL
GLCM-32GL-45deg	GLRLM-32GL	GLRLM-5GL-0deg
GLCM-32GL-90deg	GLSZM-32GL	GLRLM-5GL-45deg
GLCM-32GL-135deg	GLDM-32GL	GLRLM-5GL-90deg
GLCM-32GL	GLCM-5GL-0deg	GLCM-5GL-135deg
GLRLM-32GL-0deg	GLRLM-5GL-135deg	GLRLM-5GL
GLCM-5GL-45deg	GLCM-5GL-90deg	GLSZM-5GL
GLDM-5GL	GTV	

Through pruning, many features are not included in the model, as they have little influence over pixel classifications. 25 features are imported yet only 5 features are included in the pruned DT. These features determined in order of importance are namely GLSZM-32GL, HU, GLRLM-5GL-0deg, GLCM-32GL-0deg and GLSZM-5GL.

B. Model Characteristics

The Weka implementation of a pruned or unpruned C4.5 DT [42] is calculated through the J48 algorithm. The most influential parameters when pruning the tree is the minimum number of objects per leaf (M). Small values of M lead to larger, less restricted DTs where large values of M produce highly pruned models. Due to variability in the data from various tissue types, considerable pruning is conducted with an M value of 88. The resultant model, contains 14 leaves and a total size of 27 nodes.

The training data imported into Weka contains 5229 instances (single pixels each with 25 features) sampled from the eight training patients. The sampling was random with a ratio of 1:5 (tumour:non-tumour), based on a related approach utilised in the determination of spatial geo-hazards and landslip susceptibility mapping [43], [44]. This was included to limit sampling bias in favour of non-tumour classification.

The model was trained using five fold cross-validation, where 20% of the data was reserved for reduced-error pruning. This further improves the robustness of the model. To predict outcomes the dataset was split into eight training patients and one test patient, such that predictions were performed on unseen data. The unseen test patient predictions was reconstituted into an image to providing visual comparison to the original GTV.

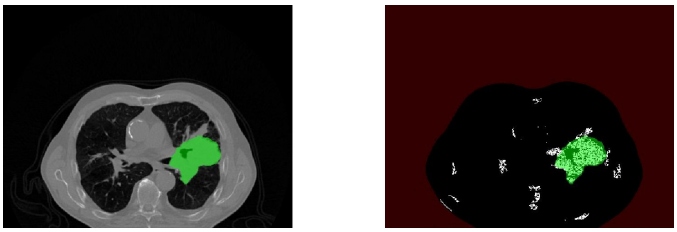


Fig. 7: Visualised model result. Left, original CT slice, right, predicted GTV slice

Figure 7 is an example, where the original CT image is on the left. The GTV is overlaid as the green area, and considered ground truth for "tumour" pixels. Figure 7 also shows the model prediction on the right. The model has classified all white pixels as "tumour". Once again the GTV is overlaid in green.

Several important factors were considered in relation to the statistical measures of model performance (Table II). The performance is firstly evaluated by finding the number of True Positives (TP) to evaluate the accuracy of correctly classifying cancerous pixels, the True Negatives (TN) to find the correct classification of non-tumour pixels, the False Positives (FP) to find how many cancerous pixels were incorrectly classified and the False Negatives (FN), to find how many non-cancer pixels were incorrectly classified.

TABLE II: Confusion matrix

		Manual Segmentation	
		Tumour	Non-Tumour
Predicted	Tumour	2306	1206
	Non-Tumour	2009	73270

Since the ratio of non-tumour:total instances is 94.4%, a model that incorrectly predicts all pixels as "non-tumour" would have an accuracy (ACC), which is all true classifications over the entire sample, of 94.4%. The ACC of the example texture model was 95.9% suggesting that the model correctly classified the tumour (Figure 7).

We calculated other parameters of model validity. Precision or positive predictive value (PPV) which refers to the probability of a tumour value being correct 65.7%. The sensitivity, or true positive rate (TPR), was 53.4%. The false positive rate (1 – specificity, FPR) which correctly reject non-tumour 1.6%.

Figure 8, displays the Receiver Operating Characteristic (ROC) curve of the model plotted in blue, based on the TPR and FPR. The orange line represents a model where all outcomes are the same resulting in an Area Under the Curve (AUROC) of 50%. The calculated AUROC for the pruned model results in 83.9%, where 50% is poor and 100% is perfect fit.

V. Discussion

Although the AUROC measure demonstrates a significant measure invariant from sample distribution, other values, such as precision, 65.7%, and sensitivity, 53.4% leave room for improvement.

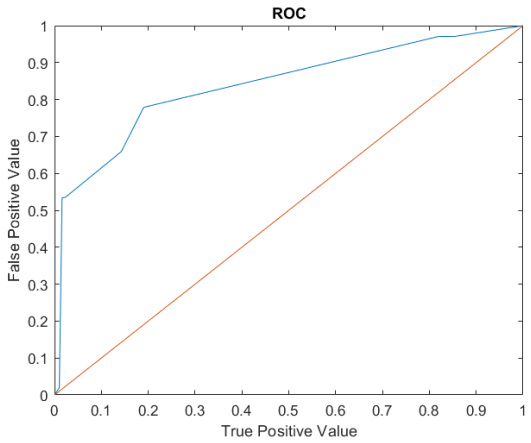


Fig. 8: The Area under the ROC curve for the texture

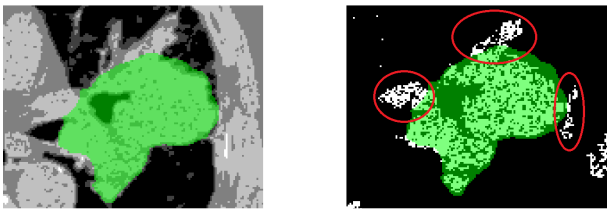


Fig. 9: Visualised model result. Left, original CT slice 5 GL. right, predicted GTV with branches highlighted

Reflecting on precision of the model to correctly classify tumour, it is worth reassessing the images in Figure 9. The right image shows three areas of protruding "tumour" in continuity with the original GTV. If these areas are a tumour, but incorrectly classified by the oncologist, the precision of the model will be adversely impacted.

The aim of this research has been to enable improved decision making by oncologists and so there will always be a dialogue between the expert and the decision system's advice. A radiation oncologist (AM) provided a discussion of the probable causes for the protruding masses.

While the non-GTV "tumour" region to the left of the GTV (9 o'clock) could be a pulmonary vessel, no other vessel in the CT image is highlighted. There could be infiltration of cancer into mediastinum along the hilum, but this should be assessed on superior and inferior CT slices.

The non-GTV "tumour" area above the tumour (12 o'clock) extending to the right could be a blood vessel. When compared with the left image, it can be seen that adjacent to the left there is another blood vessel which is narrower, and which does not produce a "tumour" signal. Its thickness indicates that it is either infiltrated with the tumour or possibly occluded/consolidated by the upstream mass. The "tumour" signal raises the possibility that it is infiltration by the tumour.

Finally, the non-GTV "tumour" region to the right of the GTV (3 o'clock), is situated over the lung pleural. While the tumour does reach the pleura on the CT image, infiltration cannot be seen. The presence of the "tumour" signal extending anteriorly and posteriorly is in keeping with the known behaviour of lung tumours involving the visceral and parietal pleura. The "tumour" signal raises the possibility of infiltration and may affect the T staging.

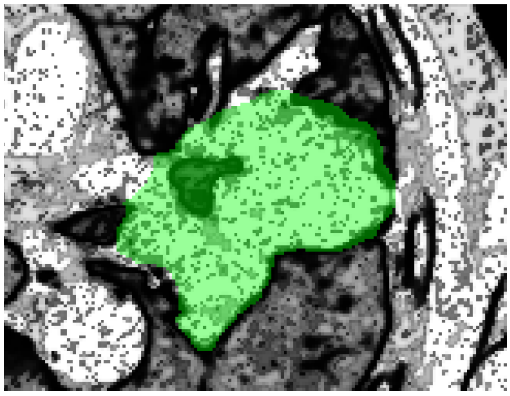


Fig. 10: Cropped GLCM-32GL prediction surrounding the GTV

Taking these interpretations into account, it is clear that the determination of precision using public datasets could be artificially low. Such findings should elicit a reassessment of the drawn GTV using additional images such as PET scan and operative findings. This process mirrors the intended use of the "tumour" delineation which is to act as a decision support system for oncologists.

Given that sensitivity reflects the accuracy of the "tumour" model predictions, one can appreciate that the accuracy of the drawn GTV will impact on an automated process that extracts features from the GTV. Pure signals are more likely to yield better sensitivity, and so appreciating that there is an air pocket within the GTV which is mislabelled as GTV, means that sensitivity will be compromised, as the model does not identify air as GTV anywhere else. While it is assumed that the GTV drawn by the oncologist is 'ground truth', when it not drawn to exclude obvious non-tumour, all subsequent analyses will be compromised, even if only by artificially lowering the sensitivity of the measure.

It is not shown in these images, but the "tumour" model also found "tumour" pixels in the contralateral hilum. The public dataset does not provide PET scans to assess this difference.

Although ML training applies heuristic methods to create an optimised model, insight from oncologists such as confirmed HU ranges for tumour tissue can add value to the model outcomes. Investigating manual pruning in conjunction with C4.5 DT will eliminate FP predictions in bone or large HU values.

Figure 10 shows the cropped GLCM-32GL. The GTV surface is peppered with darker pixels that portray similarities with the FN predictions in Figure 9. Aerts et al. [6] uncovered a link between texture and tumour histological types, so the inclusion of varying histologies may also contribute to an improved model.

VI. Conclusion

This study investigates an uncommon path in radiomic literature, which utilises pixel based texture features on CT scans to develop a model assessing tumour segmentation.

The need for independent tumour site verification to correctly establish statistical measures requires that further study be undertaken in a dataset which includes PET scans to provide additional functional data to verify accuracy for the clinician. Once established, this model has the potential to be used to classify unlabelled clinical datasets before the intervention of an oncologist.

Acknowledgements

This research has been conducted with the support of the Australian Government Research Training Program Scholarship, along with the Radiation Oncology Staff Specialist Trust Funds at Illawarra Cancer Care Centre, Wollongong NSW and the Liverpool Cancer Therapy Centre, Liverpool NSW.

References

[1] S. Sebesan and I. Olver, Eds., *Clinical oncology for medical students*, Sydney: Cancer Council Australia. [Online]. Available: https://wiki.cancer.org.au/oncologyformedicalstudents/Clinical_Oncology_for_Medical_Students.

[2] AIHW. (2017). Cancer in australia 2017, Australian Institude of Health and Walfare, [Online]. Available: https://www.aihw.gov.au/reports/cancer/cancer-in-australia-2017/contents/summary.

[3] C. S. Dela Cruz, L. T. Tanoue, and R. A. Matthay, "Lung cancer: epidemiology, etiology, and prevention," eng, *Clinics in chest medicine*, vol. 32, no. 4, pp. 605–644, Dec. 2011, ISSN: 1557-8216. DOI: 10.1016/j.ccm.2011.09.001. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/22054876%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3864624/.

[4] R. Gillies, P. Kinahan, and H. Hricak, "Radiomics: Images are more than pictures, they are data," *Radiology*, vol. 278, no. 2, pp. 563–577, 2016.

[5] P. Lambin, E. Rios-Velazquez, R. Leijenaar, S. Carvalho, R. Van Stiphout, P. Granton, C. Zegers, R. Gillies, R. Boellard, A. Dekker, and H. Aerts, "Radiomics: Extracting more information from medical images using advanced feature analysis," *European Journal of Cancer*, vol. 48, no. 4, pp. 441–446, 2012.

[6] H. Aerts, E. Velazquez, R. Leijenaar, C. Parmar, P. Grossmann, S. Cavalho, J. Bussink, R. Monshouwer, B. Haibe-Kains, D. Rietveld, F. Hoebers, M. Rietbergen, C. Leemans, A. Dekker, J. Quackenbush, R. Gillies, and P. Lambin, "Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach," *Nature Communications*, vol. 5, 2014.

[7] A. Dekker, S. Vinod, L. Holloway, C. Oberije, A. George, G. Goozee, G. P. Delaney, P. Lambin, and D. Thwaites, "Rapid learning in practice: A lung cancer survival decision support system in routine patient care data," *Radiotherapy and Oncology*, vol. 113, no. 1, pp. 47–53, 2014, ISSN: 0167-8140. DOI: 10.1016/j.radonc.2014.08.013.

[8] R. Sicilia, E. Cordelli, S. Ramella, M. Fiore, C. Greco, E. Molfese, M. Miele, E. Vinciguerra, P. Cornacchione, E. Ippolito, R. M. D'Angelillo, G. Iannello, and P. Soda, "Exploratory radiomics for predicting adaptive radiotherapy in non-small cell lung cancer," in *2018 IEEE 31st International Symposium on Computer-Based Medical Systems (CBMS)*, Jun. 2018, pp. 250–255.

[9] Z. Zhang, "Knowledge representation for machine design support systems," in *2009 Second International Conference on Intelligent Networks and Intelligent Systems*, Nov. 2009, pp. 689–692.

[10] A. Vial, D. Stirling, M. Field, M. Ros, C. Ritz, M. Carolan, L. Holloway, and A. A. Miller, "The role of deep learning and radiomic feature extraction in cancer-specific predictive modelling: A review," *Translational Cancer Research*, vol. 7, no. 3, 2018.

[11] B. Ganeshan, E. Panayiotou, K. Burnand, S. Dizdarevic, and K. Miles, "Tumour heterogeneity in non-small cell lung carcinoma assessed by ct texture analysis: A potential marker of survival," *European Radiology*, vol. 22, no. 4, pp. 796–802, 2012.

[12] D. Ravi, C. Wong, F. Deligianni, M. Berthelot, J. Andreu-Perez, B. Lo, and G. Yang, "Deep learning for health informatics," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 1, pp. 4–21, Jan. 2017, ISSN: 2168-2194. DOI: 10.1109/JBHI.2016.2636665.

[13] B. Chen, R. Zhang, Y. Gan, L. Yang, and W. Li, "Development and clinical application of radiomics in lung cancer," *Radiation Oncology*, vol. 12, no. 1, p. 154, Sep. 2017, ISSN: 1748-717X. DOI: 10.1186/s13014-017-0885-x.

[14] H. Suit and W. D. Bois, "The importance of optimal treatment planning in radiation therapy," *International Journal of Radiation Oncology*Biology*Physics*, vol. 21, no. 6, pp. 1471–1478, 1991, ISSN: 0360-3016. DOI: https://doi.org/10.1016/0360-3016(91)90321-T.

[15] E. R. Velazquez, C. Parmar, M. Jermoumi, R. H. Mak, and A. v. Baardwijk, "Volumetric ct-based segmentation of nsclc using 3d-slicer," *Scientific Reports*, vol. 3, no. 3529, Dec. 2013. DOI: 10.1038/srep03529.

[16] "Pet-ct-based auto-contouring in non-small-cell lung cancer correlates with pathology and reduces interobserver variability in the delineation of the primary tumor and involved nodal volumes," *International Journal of Radiation Oncology*Biology*Physics*, vol. 68, no. 3, pp. 771–778, 2007, ISSN: 0360-3016. DOI: https://doi.org/10.1016/j.ijrobp.2006.12.067.

[17] C. B. Caldwell, K. Mah, Y. C. Ung, C. E. Danjoux, J. M. Balogh, S. Ganguli, and L. E. Ehrlich, "Observer variation in contouring gross tumor volume in patients with poorly defined non-small-cell lung tumors on ct: The impact of 18fdg-hybrid pet fusion," *International Journal of Radiation Oncology*Biology*Physics*, vol. 51, no. 4, pp. 923–931, 2001, ISSN: 0360-3016. DOI: https://doi.org/10.1016/S0360-3016(01)01722-9.

[18] R. J. Steenbakkers, J. C. Duppen, I. Fitton, K. E. Deurloo, L. J. Zijp, E. F. Comans, A. L. Uitterhoeve, P. T. Rodrigus, G. W. Kramer, J. Bussink, K. D. Jaeger, J. S. Belderbos, P. J. Nowak, M. van Herk, and C. R. Rasch, "Reduction of observer variation using matched ct-pet for lung cancer delineation: A three-dimensional analysis," *International Journal of Radiation Oncology*Biology*Physics*, vol. 64, no. 2, pp. 435–448, 2006, ISSN: 0360-3016. DOI: https://doi.org/10.1016/j.ijrobp.2005.06.034.

[19] J.-F. Daisne and A. Blumhofer, "Atlas-based automatic segmentation of head and neck organs at risk and nodal target volumes: A clinical validation," *Radiation Oncology*, vol. 8, no. 1, p. 154, Jun. 2013, ISSN: 1748-717X. DOI: 10.1186/1748-717X-8-154.

[20] J. Dehmeshki, H. Amin, M. Valdivieso, and X. Ye, "Segmentation of pulmonary nodules in thoracic ct scans: A region growing approach," *IEEE Transactions on Medical Imaging*, vol. 27, no. 4, pp. 467–480, Apr. 2008, ISSN: 0278-0062. DOI: 10.1109/TMI.2007.907555.

[21] G. Delpon, A. Escande, T. Ruef, J. Darrfffdfffdon, J. Fontaine, C. Noblet, S. Supiot, T. Lacornerie, and D. Pasquier, "Comparison of automated atlas-based segmentation software for postoperative prostate cancer radiotherapy," *Frontiers in Oncology*, vol. 6, p. 178, 2016, ISSN: 2234-943X. DOI: 10.3389/fonc.2016.00178.

[22] C. Parmar, P. Grossmann, D. Rietveld, M. M. Rietbergen, P. Lambin, and H. J. W. L. Aerts, "Radiomic machine-learning classifiers for prognostic biomarkers of head and neck cancer," *Frontiers in Oncology*, vol. 5, p. 272, 2015, ISSN: 2234-943X. DOI: 10.3389/fonc.2015.00272.

[23] R. T. H. Leijenaar, S. Carvalho, F. J. P. Hoebers, H. J. W. L. Aerts, W. J. C. van Elmpt, S. H. Huang, B. Chan, J. N. Waldron, B. O'sullivan, and P. Lambin, "External validation of a prognostic ct-based radiomic signature in oropharyngeal squamous cell carcinoma," *Acta Oncologica*, vol. 54, no. 9, pp. 1423–1429, 2015, PMID: 26264429. DOI: 10.3109/0284186X.2015.1061214.

[24] A. Cunliffe, S. G. I. Armato, R. Castillo, N. Pham, T. Guerrero, and H. A. Al-Hallaq, "Lung texture in serial thoracic computed tomography scans: Correlation of radiomics-based features with radiation therapy dose and radiation pneumonitis development," *International Journal of Radiation Oncology * Biology * Physics*, vol. 91, no. 5, pp. 1048–1056, Apr. 2015, ISSN: 0360-3016.

[25] R. T. H. Leijenaar, G. Nalbantov, S. Carvalho, W. J. C. van Elmpt, E. G. C. Troost, R. Boellaard, H. J. W. L. Aerts, R. J. Gillies, and P. Lambin, "The effect of suv discretization in quantitative fdg-pet radiomics: The need for standardized methodology in tumor texture analysis," *Scientific Reports*, vol. 5, 11075 EP -, Aug. 2015, Article.

[26] T. P. Coroller, P. Grossmann, Y. Hou, E. Rios Velazquez, R. T. H. Leijenaar, G. Hermann, P. Lambin, B. Haibe-Kains, R. H. Mak, and H. J. W. L. Aerts, "Ct-based radiomic signature predicts distant metastasis in lung adenocarcinoma," *Radiotherapy and Oncology*, vol. 114, no. 3, pp. 345–350, Mar. 2015, ISSN: 0167-8140. DOI: 10.1016/j.radonc.2015.02.015.

[27] C. Parmar, P. Grossmann, J. Bussink, P. Lambin, and H. Aerts, "Machine learning methods for quantitative radiomic biomarkers," *Scientific Reports*, vol. 5, 2015.

[28] A. Chu, C. Sehgal, and J. Greenleaf, "Use of gray value distribution of run lengths for texture analysis," *Pattern Recognition Letters*, vol. 11, no. 6, pp. 415–419, 1990.

[29] A. Pratondo, C.-K. Chui, and S. Ong, "Integrating machine learning with region-based active contour models in medical image segmentation," *Journal of Visual Communication and Image Representation*, vol. 43, Dec. 2016. DOI: 10.1016/j.jvcir.2016.11.019.

[30] L. Hou, D. Samaras, T. M. Kurc, Y. Gao, J. E. Davis, and J. H. Saltz, "Patch-based convolutional neural network for whole slide tissue image classification," *Proceedings. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2016, pp. 2424–2433, 2016, PMC5085270[pmcid], ISSN: 1063-6919. DOI: 10.1109/CVPR.2016.266. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/27795661.

[31] E. Maksoud and M. Elmogy, "3d brain tumor segmentation based on hybrid clustering techniques using multi-views of mri," in. Jun. 2016, pp. 81–104, ISBN: 978-3-319-33791-3. DOI: 10.1007/978-3-319-33793-7_4.

[32] P. Lambin, E. Roelofs, B. Reymen, E. Velazquez, J. Buijsen, C. Zegers, S. Carvalho, R. Leijenaar, G. Nalbantov, C. Oberije, M. Scott Marshall, F. Hoebers, E. Troost, R. Van Stiphout, W. Van Elmpt, T. Van Der Weijden, L. Boersma, V. Valentini, and A. Dekker, "'rapid learning health care in oncology' - an approach towards decision support systems enabling customised radiotherapy," *Radiotherapy and Oncology*, vol. 109, no. 1, pp. 159–164, 2013.

[33] M. Vallieres, C. Freeman, S. Skamene, and I. El Naqa, "A radiomics model from joint fdg-pet and mri texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities," *Physics in Medicine and Biology*, vol. 60, no. 14, pp. 5471–5496, 2015.

[34] H. J. W. L. Aerts et al. (2015). Data From NSCLC-Radiomics. The Cancer Imaging Archive. http://doi.org/10.7937/K9/TCIA.2015.PF0M9REI.

[35] *The cancer imaging archive collections*, 2018. [Online]. Available: http://www.cancerimagingarchive.net/.

[36] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-3, no. 6, pp. 610–621, Nov. 1973.

[37] M. M. Galloway, "Texture analysis using gray level run lengths," University of Maryland, Tech. Rep., Jul. 1974.

[38] M. M. Galloway, "Texture analysis using gray level run lengths," *Computer Graphics and Image Processing*, vol. 4, no. 2, pp. 172–179, 1975.

[39] G. Thibault, B. FERTIL, C. Navarro, S. Pereira, N. Lfffdfffdvy, J. SEQUEIRA, and J.-L. MARI, "Texture indexes and gray level size zone matrix application to cell nuclei classification," Nov. 2009.

[40] C. Sun and W. G. Wee, "Neighboring gray level dependence matrix for texture classification," *Computer Vision, Graphics, and Image Processing*, vol. 23, no. 3, pp. 341–352, 1983, ISSN: 0734-189X.

[41] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: An update," *SIGKDD Explorations*, vol. 11, no. 1, pp. 10–18, 2009.

[42] J. R. Quinlan, *C4.5: Programs for Machine Learning*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1993, ISBN: 1-55860-238-0.

[43] D. Palamakumbure, P. Flentje, and D. Stirling, "Consideration of optimal pixel resolution in deriving landslide susceptibility zoning within the sydney basin, new south wales, australia," *Computers & Geosciences*, vol. 82, pp. 13–22, 2015, ISSN: 0098-3004. DOI: https://doi.org/10.1016/j.cageo.2015.05.002. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0098300415001065.

[44] P. Flentje, T. Miner, D. Stirling, D. Palamakumbure, and D. Windle, "Landslide inventory and susceptibility zoning across SE Australia," in *Developments in Engineering Geology*, Geological Society of London, Jan. 2016, ISBN: 9781862399686. DOI: 10.1144/EGSP27.11. [Online]. Available: https://doi.org/10.1144/EGSP27.11.