Enhancing the Prediction of Lung Cancer Survival Rates using 2D Features from 3D Scans

Tahira Ghani[1] and B. John Oommen[2] ⋆

School of Computer Science, Carleton University, Ottawa, Canada : K1S 5B6.
[1]tahira.ghani@carleton.ca,[2]oommen@scs.carleton.ca

Abstract. The survival rate of cancer patients depends on the type of cancer, the treatments that the patient has undergone, and the severity of the cancer when the treatment was initiated. In this study, we consider adenocarcinoma, a type of lung cancer detected in chest Computed Tomography (CT) scans on the entire lung, and images that are "sliced" versions of the scans as one progresses along the thoracic region. Typically, one extracts 2D features from the "sliced" images to achieve various types of classification. In this paper, we show that the 2D features, in and of themselves, can be used to also yield fairly reasonable predictions of the patients' survival rates if the underlying problem is treated as a *regression* problem. However, the fundamental contribution of this paper is that we have discovered that there is a strong correlation between the *shapes* of the 2D images at successive layers of the scans and these survival rates. One can extract features from these successive images and augment the basic features used in a 2D classification system. These features involve the area at the level, and the mean area along the z-axis. By incorporating additional *shape*-based features, the error involved in the prediction decreases drastically – by almost an order of magnitude. The results we have obtained deal with the cancer treatments done on 60 patients[1] at varying levels of severity, and with a spectrum of survival rates. For patients who survived up to 24 months, the average relative error is as low as 9%, which is very significant.

Keywords : *Medical Image Processing, Lung Cancer Treatment, Prediction of Survival Rates*

1 Introduction

The past few decades have boasted significant progress in diagnostic imaging in the domain of healthcare and medicine. This has, in turn, enhanced the process

⋆ *Chancellor's Professor*; *Life Fellow: IEEE* and *Fellow: IAPR*. This author is also an *Adjunct Professor* with the University of Agder in Grimstad, Norway.

[1] Understandably, it is extremely difficult to obtain training and testing data for this problem domain! Thus, both authors gratefully acknowledge the help given by Drs. Thornhill and Inacio, from the University of Ottawa, in providing us with domain knowledge and expertise for understanding and analyzing the publicly-available dataset

of exposing internal structures hidden by skin and bones through, for example, radiology. The spectrum of medical imaging techniques include, but are not limited to, X-rays, Magnetic Resonance Images (MRIs), ultrasounds, endoscopies, etc., and a wide range of pathological phenomena can now be detected.

The application of these techniques can be refined depending on the subsystem of the body that is under consideration (i.e., cardiovascular, respiratory, abdominal, etc.), *and* the associated imaging technique. Cancer, as a subcategory of these pathologies, constitutes over 100 different types.

Aspects of Lung Cancer: Apart from folklore, the statistics about cancer are disheartening. The American Cancer Society (ACS) estimated their annual statistics for 2018 based on collected historical data [1]. Lung cancer is now the second leading type of cancer for newly diagnosed patients, behind breast cancer, and it has the highest mortality rate out of all cancer sites. The ACS projected 234,030 new cases of lung cancer. They also forecasted that 154,050 deaths would be caused by lung cancer. However, cancers diagnosed at an early phase, such as Stage 1, can be treated with surgery and radiation therapy with an 80% success rate. Late diagnosis, typically, implies a lower survival rate.

Radiomics is the field of study that extracts quantitative measures of tumour phenotypes from medical images using data-characterization algorithms. These features are explored to uncover disease characteristics that are not visible to the naked eye, but which can then be used for prognosis and treatment plans. Many researchers have worked on engineering the feature sets through a radiomic analysis [2] [3]. Our goal, however, is that of predicting the survival rate of lung cancer patients *once they have been diagnosed*, and the result of this study is to demonstrate that a lot of this information resides in the *3D shape* of the tumour. We hope that our study can provide insight into the cancer's severity, and also aid in formulating the treatment plans so as to increase the chances of survival.

1.1 Contributions of this Paper

The contributions of this paper can be summarized as follows:

- Although the diagnosis of lung cancer has been extensively studied, the correlation of the survival times to the tumour's size/shape is relatively unexplored. Our first major contribution is to show that by a regression analysis, we can predict the survival times based on various features of the tumour. Predicting the survival times can essentially aid medical professionals to judge the severity of the cancer, and to design treatment plans accordingly.
- Our features are 2D features obtained from various scan slices. We show that these features, by themselves, yield impressive estimates of survival times.
- The most significant contribution of this paper is the discovery of a *distribution* for the shapes of the scans as they are processed sequentially. From these images, we can obtain relatively simple indices that relate to some geometric features of the sequence that can be used for classification and regression. By augmenting the original 2D feature set with these, we have been able to obtain significantly improved estimations for the survival rates.

– While these results have been proven to be relevant for our lung cancer scenario, we believe that these phenomena are also valid for other tumour-based cancers, and hope that other researchers can investigate the relevance of the same hypothesis for *their* application domains.

After acknowledging the source of the data and briefly surveying the field, we shall discuss the feature extraction process, and analyze the individual features against the "TumourDepth", for any correlation of the measures at successive layers. Section 5 summarizes all the regression results for the various feature sets, and Section 6 concludes the paper.

1.2 Data Source

We have used the publicly available data from The Cancer Imaging Archive[2] (TCIA), a service which hosts an archive of data for de-identified medical images of cancer. The dataset used for this work is the "LungCT-Diagnosis" data [15] on TCIA, uploaded in 2014. The set consists of CT scans for 61 patients that have been diagnosed with adenocarcinoma, a type of lung cancer, with the number of images totalling up to about $4,600$ over all the scans. However, considering only images that have the presence of a cancer nodule, the count reduces to approximately 450 images. With healthcare data, we are, of course, constrained to work with what we have. As we will see, it suffices for the purpose of regression analyses. Throughout the experiments and results explained in this paper, the condensed dataset of 450 images has been consistently split into training and testing data with a 70% and 30% split respectively, with the guarantee that there was no overlap between the two subsets. The dataset also includes the clinical metadata, where the survival time of the patient associated with each scan, is listed.

1.3 Literature Review

Feature extraction schemes in biomedical applications have been found to be specific to the medical context of the goal at hand, i.e., the features that are used vary based on the type of image being processed, as well as the pathological focus (lesions and nodules, texture variance, organ size and wall thickness, etc.). However, it may be beneficial to derive a wide variety of features, and to then reduce the set to those which prove to be most relevant [16]. This can be done through the application of feature elimination or feature selection techniques.

The goal of forming a descriptor vector in the context of nodules and texture, is a task of local feature extraction. Chabat *et al.* [17] and Kim *et al.* [18] aimed to classify obstructive lung diseases based on texture patches as Regions of Interests (ROIs). For each ROI, a statistical descriptor was calculated to describe the CT attenuation characteristics.

However, Kim *et al.* [18] also included other features such as the co-occurrence matrix and the top-hat transform. Additionally, they incorporated measurements

[2] More information can be found at `https://www.cancerimagingarchive.net/`.

to depict shapes in the ROIs, such as circularity and aspect ratios, which were not well represented by statistical measures extracted for the textures. Similarly, the authors of [19] expanded the aforementioned list with morphological features to describe the shapes of the nodules in the ROIs.

The aforementioned features, however, were used in the context goal of classification of a medical pathology. We adapt the research goal of Grove *et al.* [15] by computing and analyzing features that are indicative of the *severity* of the cancerous nodule, whereby severity can be considered as being synonymous with the survival time of a patient after the diagnosis of the cancer. Whereas their study was, for the most part, a hypothesis-based testing methodology, ours is more explorative in nature with the goal of finding descriptive quantitative measurements. It is important to note that our focus is heavily on the construction of the feature set, rather than the customization of regression models that have been used as testing thresholds.

2 Background

2.1 Computed Tomography (CT) Scans

The most common radiological imaging technique incorporates CT scans where X-ray beams are used to take measurements or images (i.e., "slices") from different angles, as shown in Figure 1, as the patient's body moves through the scanner. Depending on the section thickness and the associated reconstruction parameters, a scan can range anywhere from 100 to over 500 sections or images [5]. The scan records different levels of density and tissues which can then be reconstructed to, non-invasively, create a 3D of the human body.

(a) Axial Plane **(b)** Coronal Plane **(c)** Sagittal Plane

Fig. 1: Planes captured in a Computed Tomography (CT) scan.

High-resolution Computed Tomography (HRCT) is specifically used in detecting and diagnosing diffuse lung diseases [4] and cancerous nodules, due to its sensitivity and specificity. It enables the detection and analysis of feature aspects such as morphological lesion characterization, nodule size measurement and growth, as well as attenuation characteristics.

Hounsfield Units: CT numbers are captured and represented as Hounsfield Units (HUs), which serve as a universal standardized dimensionless metric as:

$$HU = 1000 \times \frac{\mu - \mu_{water}}{\mu_{water} - \mu_{air}}, \quad \text{where,} \tag{1}$$

- μ_{water} is the linear attenuation coefficient of water, and
- μ_{air}: is the linear attenuation coefficient of air.

Although measured in HUs, CT scans and other medical imaging reports are saved in the standard Digital Imaging and Communications in Medicine (DICOM) format. There are many DICOM viewer applications for observing and analyzing medical scans specifically.

2.2 The Preprocessing Stage

Working with the images of Chest CT scans means processing data in the DICOM format. This format groups information in the datasets, enabling the attachment of patient and pixel data, as well as technical data, such as the corresponding encoding schemes and window measurements, through attribute tags.

The DICOM images are expressed as 16-bit integer values, where the stored attribute tags specify the default, and include:

- Slice thickness;
- Number of rows and columns (i.e., image dimensions);
- Window centre and window width;
- Rescale intercept, m, and rescale slope[3], b.

The latter are used in the linear conversion of the stored value, SV, to the appropriate Hounsfield Unit (HU), U, where $U = mSV + b$.

As part of the preprocessing stage, the images, originally displayed as shown in Figure 2a, are first converted to be representative of the HUs using the above equation (i.e., $U = mSV + b$) for easier processing and visibility. The images are then scaled to two different window specifications for the Lung Window view, also known as the Pulmonary view, and the Mediastinal view, respectively. The Lung Window view, shown in Figure 2b, displays the texture in the lung and is attained by adjusting the window centre and window width, $[C, W]$, parameters to $[-500, 1400]$. The Mediastinal view, shown in Figure 2c, is attained by adjusting the parameters to $[40, 380]$.

Nodule Segmentation: Rather than segmenting the entire lung region as our Region of Interest (ROI), in this research, we segmented only the cancerous nodule in the "slices" where the presence of the tumour was observed. Similar to the topic of lung segmentation, there is an abundance of published work discussing the automation of so-called nodule segmentation and extraction [9], [10] and [11].

[3] It is important to note that more often than not, the rescale slope was valued at 1. Indeed, we have not encountered a dataset which has a different rescale slope value.

(a) Default View **(b)** Lung Window View **(c)** Mediastinal View

Fig. 2: Different Chest CT views based on window parameter adjustments.

We made masks of the tumours using the ImageJ software[4]. This was done by manually tracing a contour around the nodule on the images where it was present, filling the shape as "white", and clearing the background to "black". The images that did not contain the nodule were cleared to a "black" background.The CT scans were reviewed, and the segmentation of cancer tumours were validated by a clinical doctor from the Ottawa Heart Institute.

3 Nodule-Based Feature Extraction

When computing and compiling a feature set, we considered the scans in a 2D aspect, where each image or "slice" was treated as a single observation. For instance, if a nodule in a scan ranged over 10 images, each image was treated as a single observation or record, and amounted to 10 records in the dataset. Thus, over 60 scans, we extracted approximately 475 images that contained a nodule.

3.1 Texture and Shape-based Features

To adequately analyze the work done for this research, we first created a benchmark of prediction results based on the feature set used in the texture analysis phase of the prior algorithm. This feature set, now referred to as the "Benchmark" feature set, consisted of the Haralick values [13] computed from the Mahotas Python library, which constituted a 12-dimensional vector.

The hypothesis that we worked with in this research was that *irregularities* in a tumour's *shape* are indicative of it being cancerous. Thus, to further investigate the characterizing aspects of the tumour (when it concerns the survival rates), we modified the benchmark feature set by computing additional features relevant to the *shape* of the tumour, with the goal of being able to measure the shape's "regularity". We shall reference these as the "2D Shape" feature set. We appended the benchmark feature set by calculating the following:

[4] The ImageJ Software is a Java-based image processing program developed at the National Institutes of Health and the Laboratory for Optical and Computational Instrumentation.

- The tumour's area, measured in pixels from its mask (see Figure 3a);
- Width and height of the tumour's smallest bounding rectangle;
- Mean Squared Error of the boundary pixel from the center (with respect to the radius of the tumour's minimum enclosing circle (see Figure 3b));
- Moment values of the vector, formed by calculating the distance of the boundary of the tumour from the centre in 10° increments for a full 360° (as shown in Figure 3c), where k defines the k^{th} central moment as in Eq. (2):

$$m_k = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^k. \tag{2}$$

In the above, these reduce to the following for specific values of k:

- Variance (σ^2), where for $k = 2$, $\sigma^2 = m_2$.
- Skewness (S), where for $k = 3$ and s, the standard deviation, $S = \frac{m_3}{s^3}$.
- Kurtosis (K), where for $k = 4$ and s, the standard deviation, $K = \frac{m_4}{s^4}$.

In our study, we also included σ^2, S and K as additional features in the vector.

(a) **(b)** **(c)**

Fig. 3: Figures explaining the process of calculating the tumour's shape features.

4 Feature Set Analysis

Analyzing the Shape Features: We are now in a position to explain how we further analyzed *these* shape features, and to consider the values of these quantities against the tumour's depth. A tumour's 3D shape will, generally, start small and increase in size as it approaches the centre of the nodule and then decrease in size again. To observe the trend of the tumour's progression in size, we placed the shape features, beginning with "Area", against the "TumourDepth", and determined whether the average area and depth of the nodule displayed a correlation with the target variable. We then replaced the "Area" feature with the computed "MeanArea" given by Eq. (3):

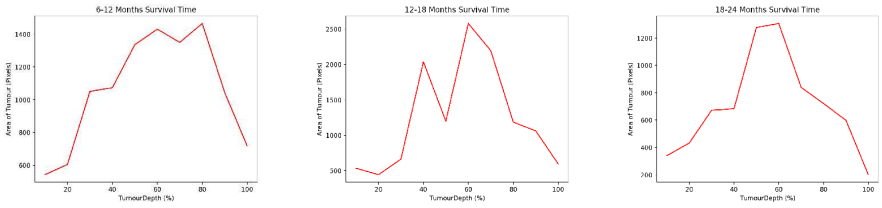
$$MA = \sum_{i=1}^{i=n}(td_i * \frac{a_i}{\sum_{i=1}^{i=n} a_i}), \quad \text{where,} \tag{3}$$

- MA: "MeanArea";
- n: Number of slices which contain the tumour;

– td_i: Tumour's depth at slice i;
– a_i: Tumour's area at slice i.

This process was repeated for all the shape features and appended as new features after removing the original shape feature. This feature set will be referred to as the "Averaged 2D Shape" feature set.

Correlating the Shape Features against Survival Times: To better understand how the shape features correlate against the survival times, we analyzed the averaged "Area" feature further by plotting curves of the tumour's area in relation to the depth of the image in the tumour, in 6-month bins of "SurvivalTime". Figures 4a, 4b and 4c display plots of the curve for the $[6, 12)$, $[12, 18)$ and $[18, 24)$ months bounds respectively, where in the first case, the average area occurred at 59.1% tumour depth[5]. The overall results are summarized in Table 1, where the area is plotted against the percentage of the tumour depth for each 6-month bin present in the data. A notable observation is the steady trend of the decreasing "Area" mean over the first four bins (i.e., those that have a survival time within 2 years of diagnosis). Given this trend, we hypothesised that we could also test the regression models for this survival time frame.



(a) 6-12 Months **(b)** 12-18 Months **(c)** 18-24 Months

Fig. 4: Area vs. TumourDepth Plots for tumours for patients with a survival rate of 6-12 months, 12-18 months and 18-24 months.

"SurvivalTime" Bin	Average Area
(0, 6]	0.685
(6, 12]	0.591
(12, 18]	0.579
(18, 24]	0.555
(24, 30]	0.589
(30, 36]	0.566
(36, 42]	0.679
(42, 48]	0.671
(48, 54]	0.585
(54, 60]	0.546

Table 1: Averages of the tumour's "Area" over bins of 6-Month durations.

[5] The term "Tumour depth" refers to the level (or z-axis) along the tumour as it progresses down the thoracic region.

To verify this hypothesis, we performed a regression analysis on a subset of the data which has a "SurvivalTime" of 24 months or less. Subsets were taken of all of the aforementioned feature sets, and processed (with these augmented features) to recalculate the errors by testing the corresponding regression models.

5 Results and Discussions

5.1 Model Evaluation

To evaluate the performance of the tested regression models, we utilized two measures, namely the Mean Absolute Error (MAE), measured in months, and the Mean Relative Error (MRE), both of which are defined in Eq. (4) and (5) respectively:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - z_i|, \qquad \text{and} \tag{4}$$

$$MRE = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - z_i|}{z_i}, \qquad \text{where:} \tag{5}$$

- n is the number of test-set data points,
- y_i is the predicted value (i.e., the expected survival time in months), and
- z_i is the true value (i.e., the survival time in months).

The MAE is the average difference between the true values and the predicted values. It provides an overall measure of the distance between the two values, but it does not indicate the direction of the data (i.e., whether the result is an under or over-prediction). Furthermore, this is also seen to be a scale-dependent measure, as the computed values are heavily dependent on the scale of the data, and can be influenced by outliers present in the data [14]. In order to circumvent the scale-dependency, we also computed the MRE which introduces a relativity factor by normalizing the absolute error by the magnitude of the true value. This means that the MRE should, generally, consist of values in the range $[0, 1]$.

As mentioned earlier, all the regression tests were done on the data with a 70% to 30% split of the data for training and testing, respectively.

5.2 Regression Results

We used the `scikit-learn` machine learning library to implement the basic models for regression on the 2D Benchmark feature set. When analyzing the recorded metrics, we emphasize that we were attempting to minimize the error and maximize the accuracy of all the tested models. Also, in the interest of

uniformity, we consistently report the results obtained by invoking the three best schemes, namely Linear Regression, kNN Regression and Gradient Boosting[6][7].

Model	MAE (months)	MRE (%)
Linear Regression	12.30	0.76
kNN Regression	12.13	0.79
Gradient Boosting	12.49	0.77

Table 2: Performance of the Regression Models on the "Benchmark" Feature Set.

As can be seen from the MRE indices recorded in Table 2, the error seems to be relatively high when compared to the results of the binary classification attained for this feature set (discussed in [12]). While such a performance may be undesirable, it is certainly not surprising when we consider the following:

1. **Difference of context**: This feature set was initially compiled with the goal of a binary classification task between "Normal" and "Abnormal" lung texture. Processing the entire region belonging to the lung parenchyma enables the visibility of stark differences in texture which can correspond to one of the two aforementioned classes.
2. **Patch size**: The prior algorithm processed a fixed patch size of 37 x 37 pixels, whereas, in this research, we are processing varying patch sizes depending on the size of the bounding box of the tumour in a 2D image.
3. **Tumour isolation**: In this research, we considered only the tumours and not the texture around them in the 3D matrix compilations. Thus, any pixels that were included in the bounding box but not seen to be a part of the tumour were reduced to zero (i.e., black) and ignored from the feature computations. This could surely affect the variability seen in different tumours.
4. **Correlation of depth and survival time**: Although the depth of the image with regards to the tumour would yield different feature measurements, all slices throughout the tumour would correspond to the same value for the target value since it belonged to a single scan. This would, apparently, reduce the correlation between the features and the target variable.

Model	MAE months)	MRE (%)
Linear Regression	12.13	0.70
kNN Regression	14.48	0.89
Gradient Boosting	8.82	0.54

Table 3: Performance of Regression Models with Shape Features

[6] In the interest of conciseness, we do not discuss the details of the Machine Learning methods invoked. We assume that the reader is aware of them. Additional results, which also detail the results of other methods, are included in the thesis of the first author [12], and not given here in the interest of space.

[7] The thesis publication can be found on the Carleton University website, linked as follows: `https://doi.org/10.22215/etd/2019-13731`.

To further support our initial hypothesis, we ran the same regression models (as we earlier did for the baseline results) with the benchmark feature set and the new shape features, i.e., the 2D Shape feature set. The results that we obtained and the respective error values, are listed in Table 3. As can be seen, the Linear Regression models improved in both the MAE and MRE measures. While the kNN Regression digressed with a larger error, Gradient Boosting performed significantly better – with an improvement of 23%.

Model	**MAE (months)**	**MRE (%)**
Linear Regression	11.95	0.70
kNN Regression	14.52	0.86
Gradient Boosting	7.40	0.45

Table 4: Performance of Regression Models with Averaged 2D Shape Feature Set

For the "Averaged 2D Shape" feature set, the Linear Regression displayed a minor improvement in the MAE, while the MRE remained the same. On the other hand, the kNN regression digressed for the MAE measure but improved for the MRE by about 3%. The Gradient Boosting scheme, however, had the best improvement with a decrease of approximately 15% in both the MAE and MRE, confirming the advantage of considering the nodule in its 3-dimensional entirety, compared to the previous 2D feature sets, and to methods that incorporated minimal changes in the Linear and kNN regression schemes. Again, even superior results were obtained on a subset of the data which had a "SurvivalTime" of 24 months or less. Table 5 displays the results of the regression models with respect to each feature set. The results of the MAE improved by at least 60% for all feature sets and regression models. The most notable improvement was with the Gradient Boosting scheme for these modified features, where we obtained a total improvement of almost 70% in the MRE, bringing the absolute error down to 1.29%. This is, by all metrics, remarkable.

Feature Set	**Evaluation Metric**	**Regression Model**		
		Linear	**kNN**	**Gradient Boosting**
Benchmark	MAE	4.27	5.22	4.56
	MRE	0.35	0.40	0.37
2D Shape	MAE	3.53	4.60	1.79
	MRE	0.27	0.35	0.12
Average 2D Shape	MAE	3.23	5.33	1.24
	MRE	0.26	0.41	0.09

Table 5: Regression Results with Data Subset: 'SurvivalTime' Less Than 24 Months

6 Conclusion

In this paper, we discussed the domain of healthcare imaging for diagnostics and the implementation of radiomics on CT scans, in particular, to predict the survival rates of lung cancer patients. We first tested the feature set that was used in a prior algorithm, where the goal was to examine regression models.

By modifying the existing features by including shape descriptors, which were focused on the cancer nodule itself, we were able to obtain improvements in the best regression models. These new features were then analyzed against the "TumourDepth"', from which a strong correlation between the images at successive layers of the scan was discovered. Further investigating the shape aspects of the tumour, we observed the progression of the "Area"-based feature values *versus* tumour progressions in bins of "SurvivalTime", and were able to observe a notable trend for survival rates up to 24 months. By performing a regression testing for the data within this subset yielded an MRE of as low as 9%, and we even obtained a total improvement of almost 70% in the MRE.

With regard to future work, the feature sets that have been explored, should also be tested on other collections of CT scans to ensure consistency. We also believe that similar methods can be used for other tumour-based cancers.

References

1. Siegel, R. L. and Miller, K. D. and Jemal, A. Cancer Statistics. 2018. *CA: A Cancer Journal for Clinicians*, 68:7-30.
2. Paul, R. and Hawkins, S. H. and Schabath, M. B. and Gillies, R. J. and Hall, L. O. and Goldgof, D. B. Predicting malignant nodules by fusing deep features with classical radiomics features. 2018. *Journal of Medical Imaging*, 5.
3. Fan, L. and Fang, M. and Li, Z. and Tu, W. and Wang, S. and Chen, W. and Tian, J. and Dong, D. and Liu, S. Radiomics Signature: A biomarker for the preoperative discrimination of lung invasive adenocarcinoma manifesting as a ground glass nodule 2019 *European Radiology*, 29:889-897.
4. Elicker, B. M. and Webb, W. R. Fundamentals of High-Resolution Lung CT. 2013. Wolters Kluwer.
5. Al Mohammad, B. and Brennan, P. C. and Mello-Thoms, C. A Review of Lung Cancer Screening and the Role of Computer-Aided Detection. 2017. *Clinical Radiology*, 72:433-442.
6. Armato III, S. G. and Sensakovic, W. F. Automated Lung Segmentation for Thoracic CT: Impact on Computer-Aided Diagnosis. 2004. *Academic Radiology*, 11:1011-1021.
7. Zhou, S. and Cheng, Y. and Tamura, S. Automated lung segmentation and smoothing techniques for inclusion of juxtapleural nodules and pulmonary vessels on chest CT images. 2014.]newblock *Biomedical Signal Processing and Control*, 13:62-70.
8. Singadkar, G. and Mahajan, A. and Thakur, M. and Talbar, S. Automatic lung segmentation for the inclusion of juxtapleural nodules and pulmonary vessels using curvature based border correction. 2018. *Journal of King Saud University - Computer and Information Sciences*.

9. Zhao, B. and Gamsu, G. and Ginsberg, M. S. and Jiang, L. and Schwartz, L. H. Automatic Detection of Small Lung Nodules on CT Utilizing a Local Density Maximum Algorithm. 2003. *Journal of Applied Clinical Medical Physics*, 4:248-260.
10. , Armato III, S. G. and Giger, M. L. and MacMahon, H. Automated Detection of Lung Nodules in CT Scans: Preliminary Results. 2001. *Medical Physics*, 28:1552-1561.
11. , Messay, T. and Hardie, R. C. and Tuinstra, T. R. Segmentation of Pulmonary Nodules in Computed Tomography using a Regression Neural Network Approach and its Application to the Lung Image Database Consortium and Image Database Resource Initiative Dataset. 2015. *Medical Image Analysis*, 22:48-62.
12. Ghani, T. Feature Engineering with Radiomics for Optimal Prediction of Survival Rates of Adenocarcinoma Patients. MCS Thesis, Carleton University, Ottawa. 2019.
13. Haralick, R. M. and Shanmugam, K. and Dinstein, H. Textural Features for Image Classification. 1973. *IEEE Transactions on systems, man, and cybernetics*, 6:610-621.
14. Chen, C. and Twycross, J. and Garibaldi, J. M. A New Accuracy Measure Based on Bounded Relative Error for Time Series Forecasting. 2017. *PloS one*, 12.
15. Grove, O. and Berglund, A. E. and Schabath, M. B. and Aerts, H. JWL. and Dekker, A. and Wang, H. and Velazquez, E. R. and Lambin, P. and Gu, Y. and Balagurunathan, Y. Quantitative computed tomographic descriptors associate tumour shape complexity and intratumor heterogeneity with prognosis in lung adenocarcinoma. 2015. *PloS one*, 10.
16. Hall, E. L. and Kruger, R. P. and Dwyer, S. J. and Hall, D. L. and McLaren, R. W. and Lodwick, G. S. A Survey of Preprocessing and Feature Extraction Techniques for Radiographic Images. 1971. *IEEE Transactions on Computers*, 100.
17. Chabat, F. and Yang, G. Z. and Hansell, D. M. Obstructive Lung Diseases: Texture Classification for Differentiation at CT. 2003. *Radiology* 228.
18. Kim, N. and Seo, J. B. and Lee, Y. and Lee, J. G. and Kim, S. S. and Kang, S. H. Development of an Automatic Classification System for Differentiation of Obstructive Lung Disease using HRCT. 2009. *Journal of Digital Imaging*, 22.
19. Demir, O. and Camurcu, A. Y. Computer-Aided Detection of Lung Nodules using Outer Surface Features. 2015. *Bio-Medical Materials and Engineering*, 26.