2022

Using Radiomics to improve the 2-year survival of Non-Small Cell Lung Cancer Patients

Alanna Heather Therese Vial

UNIVERSITY
OF WOLLONGONG
AUSTRALIA

Using Radiomics to improve the 2-year survival of Non-Small Cell Lung Cancer Patients

Alanna Heather Therese Vial

This thesis is presented as part of the requirements for the conferral of the degree:

Master of Philosophy

Supervisor:
Dr. D. Stirling

Co-supervisors:
Dr. C. Ritz & Dr. M. Ros & Dr. M. Field & Dr. L. Holloway & Dr. M. Carolan & Dr. A. Miller

The University of Wollongong
School of Electrical, Computer and Telecommunications Engineering

July, 2022

This research has been conducted with the support of an Australian Government Research Training Program Scholarship.

Abstract

This thesis both exploits and further contributes enhancements to the utilization of radiomics (extracted quantitative features of radiological imaging data) for improving cancer survival prediction. Several machine learning methods were compared in this analysis, including but not limited to support vector machines, convolutional neural networks and logistic regression.A technique for analysing prognostic image characteristics, for non-small cell lung cancer based on the edge regions, as well as tissues immediately surrounding visible tumours is developed. Regions external to and neighbouring a tumour were shown to also have prognostic value. By using the additional texture features an increase in accuracy, of 3%, is shown over previous approaches for predicting two-year survival, which has been determined by examining the outside rind tissue including the tumour compared to the volume without the rind. This indicates that while the centre of the tumour is currently the main clinical target for radiotherapy treatment, the tissue immediately around the tumour is also clinically important for survival analysis. Further, it was found that improved prediction resulted up to some 6 pixels outside the tumour volume, a distance of approximately 5mm outside the original gross tumour volume (GTV), when applying a support vector machine, which achieved the highest accuracy of 71.18%. This research indicates the periphery of the tumour is highly predictive of survival. To our knowledge this is the first study that has concentrically expanded and analysed the NSCLC rind for radiomic analysis.

Acknowledgement

I would like to thank my family and friends who stood by me when I was struggling and supported me to keep going. Specifically, I would like to thank Josh for supporting me through the hard times. I would like to thank my Dad, Peter for encouraging me to always do my best. And I would like to thank my Gran, Marie and Grandfather, Brian, for always believing in me and being proud of me no matter what.

Thank you to my supervisors who spent hours with me thinking through ideas and editing my manuscripts.

Declaration

I, *Alanna Heather Therese Vial*, declare that this thesis is submitted in partial fulfilment of the requirements for the conferral of the degree *Master of Philosophy*, from the University of Wollongong, is wholly my own work unless otherwise referenced or acknowledged. This document has not been submitted for qualifications at any other academic institution.

Alanna Heather Therese Vial

July 1, 2022

Abbreviations

2D **2**-**D**imensional
3D **3**-**D**imensional
3D c-GANs **3**-**D**imensional **C**onditional **G**enerative **A**dversarial **N**etworks
ANN **A**rtificial **N**eural **N**etwork
AUC **A**rea **U**nder the **C**urve
BoV **B**ag of **V**isual words
BN **B**ayesian **N**etworks
CBCT **C**one **B**eam **C**omputed **T**omography
CNN **C**onvolutional **N**eural **N**etwork
CTV **C**linical **T**arget **V**olume
CT **C**omputed **T**omography
CTTA **C**omputed **T**omography **T**exture **A**nalysis
COA **C**ontrast **O**rientated **A**lgorithm
CDBN **C**onvolutional **D**eep **B**elief **N**etwork
DSS **D**ecision **S**upport **S**ystem
DBN **D**eep **B**elief **N**etworks
DM **D**istant **M**etastases
DICOM-RT The **D**igital **I**maging and **CO**ommunications in **M**edicine-**R**adiation **T**herapy
FV **F**isher **V**ector
FDG-PET **F**lu**D**eoxy**G**lucose-**P**ositron **E**mission **T**omography
GMM **G**aussian **M**ixture **M**odelling
Gy **G**rays
GLCM **G**ray-**L**evel **C**o-occurence **M**atrix
GTV **G**ross **T**umour **V**olume
H&N **H**ead **&** **N**eck cancer
IBk or kNN k-**N**earest **N**eighbour classification algorithm
ML **M**achine **L**earning
MRI **M**agnetic **R**esonance **I**maging
MML **M**inimum **M**essage **L**ength
NSCLC **N**on-**S**mall **C**ell **L**ung **C**ancer
PACS **P**icture **A**rchiving **C**ommunications **S**ystems

PET **P**ositron **E**mission **T**omography
PCA **P**rincipal **C**omponent **A**nalysis
RF **R**andom **F**orests
ROI **R**egion **O**f **I**nterest
RLN **R**un **L**ength **N**on-uniformity
SOM **S**elf **O**rganising **M**aps
STS **S**oft-**T**issue **S**arcoma lung cancer
SUVmax maximum **S**tandardised **U**ptake **V**alue
SVM **S**upport **V**ector **M**achines
VB-GMM **V**ariational **B**ayesian-**G**aussian **M**ixture **M**odelling

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Introduction

Cancer continues to contribute to many death's in modern society and while methods to improve cancer outcomes have been improved in recent years this is still a major issue in modern medicine. Personalised medicine is a new cancer research area which hopes to provide individualised cancer treatment based on the outcomes of patients with similar biology and cancer types. To achieve this goal, it has been proposed that a new technique known as Radiomics should be used to find common features in CT imaging data that may indicate the likely outcome for the patient and thereby determine the recommended treatment options effectively. By outcome we are referring to the ability to predict the likelihood that a patient will survive 2-years after the image is taken.

Depending on the type of cancer, clinicians are interested in different end point predictions. In the case of non-small cell lung cancer accurate 2-year and 5-year prediction models would be preferred as this is an extremely aggressive form of cancer and has a poor 2-year survival rate. In the alternative case of breast cancer, the preferred prediction models would be for at least 10-year survival as this cancer now has improved survival rates. Clinicians are also interested in toxicity levels as they do not wish to over or under prescribe radiotherapy treatment.

Radiomics is a term coined by Lambin et al. [1] to describe the process of extracting useful imaging features from radiological data (including nuclear medicine data). The ultimate goal of radiomics is to link these features with outcomes so as to enhance precision medicine [2]. For oncology, this process involves collecting and processing the medical images, delineating the lesion of interest, extracting radiomic features and then using this insight in tandem with conventional semantic insights to predict outcomes or response to therapy.

1.2 Key Research Questions

This thesis aimed to answer the following research questions:

- Is there any information in the periphery of the tumor that can predict 2-year survival likelihood of NSCLC patients?
- What radiomic features can be used to predict 2 year survival likelihood of NSCLC patients?

For the first research question, it was hypothesised early in this work that more accurate tumour delineations would lead to better treatment outcomes. It is for this reason that we decided to explore the information in the periphery of the tumour to determine if this hypothesis showed any merit.

The second research questions is an obvious requirement of the radiomics field as the main purpose of this research area is to determine which features need to be searched for to locate tumours. Once this research is fully applied it may be useful in completing non-invasive screenings and early cancer detection.

1.3 Chapter Summary

Chapter 1, introduces Radiomics, it provides the background motivations for this work, summary of the achievements and goals of the research.

Chapter 2, reviews objective methods for modelling cancer tumours located within CT images using machine and deep learning applied to features derived from radiomics in order to improve patient diagnosis, treatment options and outcomes. The purpose of this chapter is to provide a clear overview of how machine and deep learning can be applied to radiomics. Radiomics is a novel feature transformation method for detecting textural features that are difficult for the human eye to perceive. The ultimate goal of radiomics is to enable radiation therapists and patients to make more informed decisions relating to treatment options.

Chapter 3, examines a technique for developing prognostic image characteristics, also termed radiomics, for non-small cell lung cancer based on a tumour edge region-based analysis. Texture features were extracted from the rind of the tumour in a publicly available 3D CT data set to predict two-year survival. The derived models were compared against the previous methods of training radiomic signatures that are descriptive of the whole tumour volume. Radiomic features derived solely from regions external, but neighbouring, the tumour were shown to also have prognostic value. By using additional texture features an increase in accuracy, of 3%, is shown over previous approaches for predicting two-year survival, upon examining the outside rind including the volume compared to the volume without the rind. This indicates that while the centre of the tumour is currently the

main clinical target for radiotherapy treatment, the tissue immediately around the tumour is also clinically important.

Chapter 4, explores the prediction of two year survival for non-small cell lung cancer patients is a key issue in cancer research using deep learning. This chapter aimed to attain what could be achieved without segmenting the tumour region using 2-Dimensional Convolutional Neural Networks. Recent studies have focused on using radiomics to analyse cancer images, to find imaging features that can aid in two year survival prediction. This study aims to achieve the same goal with the application of CNNs that can find imaging features automatically based on the raw data. This study compares the accuracy of 2D CNNs on a large unseen test data set achieving an accuracy of 56% compared to 59% found by radiomic studies. This analysis did well using 2D data compared to radiomic studies which consider tumours as a whole 3D volume. These results highlight the need to normalise slices on a patient basis by centering on the tumour volume.

Chapter 5, aims to characterise the 2-year survival rate of non-small cell lung cancer patients, based on a tumour edge region-based analysis. This study also compares nine machine learning (ML) techniques to determine the ideal method for predicting cancer survival. This chapter explores the region outside and including the gross tumour volume (GTV) from 0 to 10 pixels. This chapter also describes improved prediction results at 6 pixels outside the tumour volume, a distance of approximately 5mm outside the original GTV, when applying a support vector machine and achieving the highest accuracy of 71.18%. This research indicates the periphery of the tumour is highly predictive of survival. To our knowledge this is the first study that has concentrically expanded and analysed the NSCLC rind for radiomic analysis [3].

Chapter 6, concludes this work and proposes promising directions for future work.

1.4 Key Research Contributions

This thesis contributed the following research outcomes:

- The first research explored the rind of the tumour volume to find links between textural features and 2-year survival outcomes.
  - A. Vial et al., "Assessing the Prognostic Impact of 3D CT Image Tumour Rind Texture Features on Lung Cancer Survival Modelling," in 5th IEEE Global Conference on Signal and Information Processing, 2017. [Citations 2, Nov. 2021]
- The second research contribution reviewed the current research in the field of radiomics and deep learning.

  - A. Vial et al., "The role of deep learning and radiomic feature extraction in cancer-specific predictive modelling: a review," in Translational Cancer Research, pp.803-816, Vol. 7, No. 4, June 2018. [Citations 39, Nov. 2021]

- The third research contribution explored a variety of machine learning techniques to determine which technique provided the best predictive ability for determining 2-year survival.

  - A. Vial et al., "A comparative study of machine learning techniques for the improved prediction of NSCLC survival analysis," 2018 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2018. [Citations 1, Nov. 2021]

Chapter 2

Related Work

Machine learning techniques can be employed to explore the relationships between MRI or CT images processed using radiomics, outcome and radiation dosage data to improve and assist radio-therapists in the treatment of patients with cancer.

Depending on the type of cancer, clinicians are interested in different end point predictions. In the case of non-small cell lung cancer accurate 2-year and 5-year prediction models would be preferred as this is an extremely aggressive form of cancer and has a poor 2-year survival rate [4]. In the alternative case of breast cancer the preferred prediction models would be for at least 10-year survival as this has improved survival rates. Clinicians are also interested in toxicity levels as they do not wish to over or under prescribe radiotherapy treatment.

Section 2.1 introduces the field of radiomics. Section 2.2 provides a brief overview of what radiomics is and why it is important to cancer research. Section 2.3 contains a brief overview of machine learning techniques that have been applied to analyse 3-Dimensional images and how these techniques have been applied to radiation oncology. Section 2.4 discusses in depth how deep learning, namely CNN may be applied to the field of radiomics. Section 6 concludes this review.

2.1 Radiomics

Radiomics is a term coined by Lambin et al. [1] to describe the process of extracting useful imaging features from radiological data (including nuclear medicine data). The ultimate goal of radiomics is to link these features with outcomes so as to enhance precision medicine [2]. For oncology, this process involves collecting and processing the medical images, delineating the lesion of interest, extracting radiomic features and then using this insight in tandem with conventional semantic insights to predict outcomes or response to therapy. The four main processes of radiomics: imaging, segmentation, feature extraction and analysis, are described in detail in the subsequent sections of this literature review as

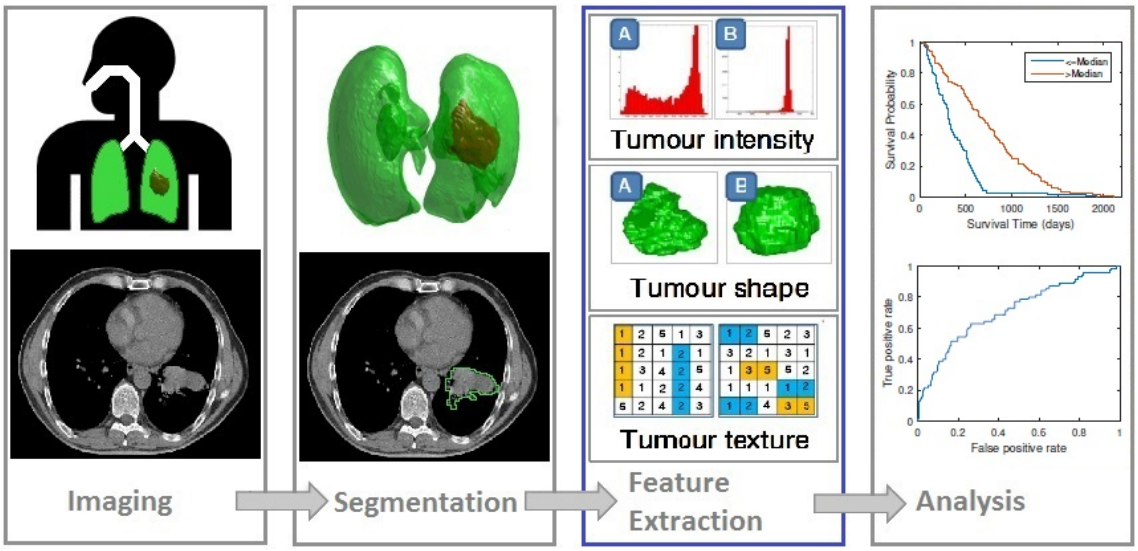

Figure 2.1: This figure shows the process of radiomics from imaging, to segmentation to feature extraction and analysis.

shown in Figure 2.1.

Radiomics can be employed with machine learning and pattern recognition to find and interpret new information within medical images. In this case pattern recognition is a field that can be employed to detect texture features in images, while machine learning can be used to relate these texture features obtained through pattern recognition techniques to correlate these with outcomes. Hence both of these fields are important and complimentary so they will be discussed in detail in this review. Deep learning is a form of machine learning that can be trained to perform pattern recognition, that can be used to detect important imaging features while also relating these features to outcome labels for images, this is discussed in detail in Section 2.4.

This review will focus mainly on the last two aspects of radiomics, namely feature extraction and analysis. The ultimate goal of radiomics is to provide a tool or decision support system that can aid in the clinical decision making process and to provide accurate diagnosis for patients to enable personalised radiotherapy, that should lead to improved treatment outcomes.

2.1.1 Imaging

Imaging involves determining which data needs to be acquired for further analysis. This data can consist of computed tomography (CT), positron emission tomography (PET), magnetic resonance imaging (MRI), fluorodeoxyglucose-PET (FDG-PET), cone beam (CBCT) or PET-CT images. PET scans are useful for observing metabolic processes in the human body, these can often be used to highlight the active area of the tumour. The main disadvantage of PET is that the image resolution is significantly reduced.

CT images have become the primary method of diagnosing cancer; as a result there are

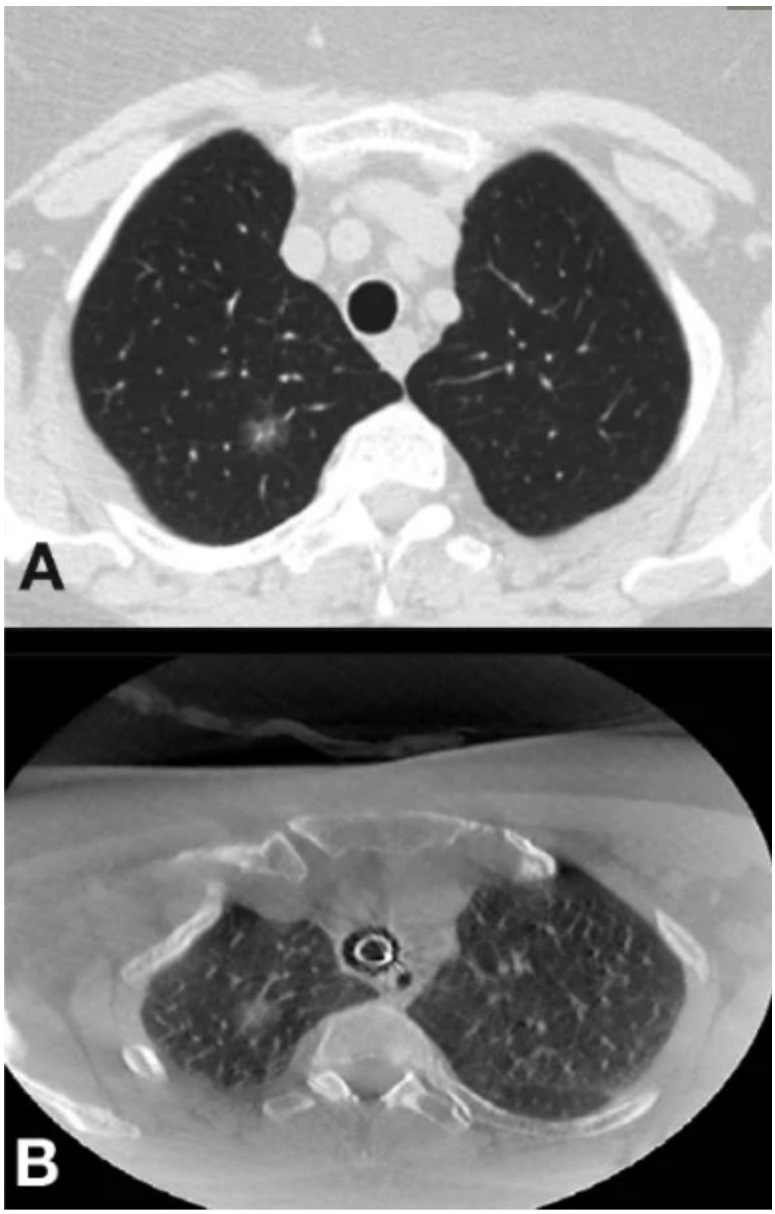

Figure 2.2: This figure shows a comparison of cone beam (CBCT) and a standard CT image of the same patient. Image A is a CT image and image B is a CBCT image [6].

large volumes of data available to learn from [5]. However, there are numerous technical compatibility issues with collating all of this data into a comparable form, as different CT scanners have different resolutions and settings. Hence the imaging data needs to be pre-processed into a usable format and de-identified to allow CT scans from different equipment to be compared accurately as well as to ensure patient confidentiality.

There are various imaging modalities that are used to diagnose cancer and plan treatment such as CT, PET-CT, MRI, CBCT and FDG-PET scans. When considering imaging modalities, various factors need to be considered for further analysis such as resolution, number of scans available and data quality. Here a CT scan used for precise radiation planning to optimise 3D radiation dose and to segment tumour and normal tissues that

are at risk. Whereas CBCT scans are taken quickly for daily treatment target positioning during a course radiation therapy. Therefore, there are many more CBCT scans than CT scans taken in the radiotherapy setting but the images are of poorer quality in resolution and clarity.

PET scans tend to vary significantly as there is no standard type used for all patients, hence this is not ideal for applying machine learning techniques as one can only compare pet scans of the same type, such as FDG-PET with ideally the same CT imaging machine [7]. MRI scans can also vary substantially between different MRI machines. CT imaging is the ideal modality for machine learning as there are a large number of scans available for analysis and the resolution of these images is high. In the case of non-small lung cancer analysis there are a large number of high quality CT scans available for study, which is why this modality was chosen for this study.

This review will focus on CT images, the reason for this is summarised through the comparison of imaging modalities shown in Table 2.1.

Table 2.1: Comparison of the of different medical imaging modalities. The feasibility for radiomic analysis is determined based on the image resolution and number of scans available.

Imaging Modality	Resolution	No. of Scans available	Feasibility for Radiomic Analysis
CT	High	Medium	High
PET	Medium	Low	Low
MRI	Very High	Very Low	Medium
CBCT	Low	Very High	Medium

2.1.2 Segmentation

Segmentation involves selecting the part of the image correctly that contains the tumour so it can be separated from the rest of the image for further analysis.

The current method for delineating a tumour volume is visually by an experienced radiation therapist, who may use assistive contouring tools in treatment planning software to semi-automatically delineate the tumour with some interaction from the oncologist to check and adjust this. Alternatively automatic contouring tools are commonly used to delineate organs at risk. For instance, the atlas segmentation method is used to identify where the major organs are segment them out and then to find irregular tumour tissue within in the body. An accurate method of automatically delineating the tumour volume with no interaction from a radiotherapy oncologist, have not yet been used routinely in clinics.

Velazquez et al. [5], [8] and Schaefer et al. [7] used semi-automatic segmentation methods and compared the delineations to the macroscopic measurement of the surgical specimen (tumour) after removal, which is considered to be the "gold standard" for non-small cell lung cancer (NSCLC). Velazquez et al. [8] uses the freely available 3D-Slicer tool to make the delineations, however Schaefer et al. [7] uses a contrast-orientated algorithm (COA) to determine the region of interest (ROI) or GTV. Both of these papers indicate promising results with the main strength being the delineations were compared against the surgical specimen, however they both had very small sample sizes with Velazquez et al. [8] having 20 samples and Schaefer et al. [7] including 15 samples.

2.1.3 Feature Extraction

Feature extraction comprises the utilising of a variety of image processing techniques and mathematical analyses to find common patterns within the images.

Kumar et al. [9] describes how many radiomic features can be extracted compared to the amount of patients analysed, hence the quantity of features analysed needs to be reduced to avoid overfitting the data. This paper suggests that of the 327 quantitative features only 39 features were found to be reproducible, informative and not redundant. The extraction of features in this paper required applying a statistical formula or imaging filter to determine the radiomic features. Being able to relate a specific feature to a specific cancer type or outcome is very difficult to achieve accurately.

The relationship between radiomic features and the prediction of cancer survival rates is examined in Aerts et al. [10]. The radiomic features can be broken into four main categories (I) texture, (II) Wavelet or image transformations, (III) shape and (IV) tumour intensity features. In this paper, 440 features were extracted and four of these features were selected for analysis. It was found that when these four radiomic features were compared against 2-year survival data the presence or absence of these features could be used to determine a patient's likelihood of survival with moderately high accuracy [10].

Many authors have focused on the extraction of different radiomics features through applying mathematical formulas to images to determine if there is a correlation between these features and survival. It is important to discuss these papers as many have found recurring metrics, such as the maximum Standardised uptake value (SUVmax) to be strongly correlated with survival. Machine learning can be used to determine which metrics are strongly correlated for all cancer types but more importantly it can be used to determine relationships that have not even been considered by letting the data speak for it self and allowing deep learning techniques such as neural networks to determine the key radiomic features.

Vallieres et al. [11] introduced a novel approach where the authors fused FDG-PET scans with MRI imaging to allow the prediction of lung metastasis using radiomics with

improved accuracy in soft-tissue sarcoma (STS) lung cancer. Despite the data losses due to the resolution of FDG-PET scans being far less than MRI images, the fusion of these two modalities allowed the texture and shape of the tumour to be clearly highlighted enabling the authors to more clearly differentiate between patients who developed lung metastasis and those who did not, due to the tumour heterogeneity being easier to predict. It is important to note than if a patient develops lung metastasis they have a very low 3 year survival rate. It was also found that the maximumum standardized uptake value (SUVmax) was significantly associated with lung metastasis risk in STS lung cancer, as SUVmax is strongly correlated with tumour heterogeneity, which has been independently determined in other radiomic studies [12], [13]

2.1.4 Analysis and Prediction

Analysis includes modelling with these features using classifiers such as decision trees, random forests or Gaussian mixture modelling to predict important patient information such as 2-year survival or whether a cancer is likely to spread or not. As an example, decision trees are a machine learning technique that separate outcomes based on the statistical significance, displaying multiple hypotheses as a probability tree. Decision trees are good for improving data interpret ability making them ideal for classification, however it has been demonstrated in literature that they are sensitive to training data, and even small changes can result in different trees. Decision trees can provide extra information to an oncologist that can improve their decision making ability when deciding on the ideal dose to give to a patient. It can also allow experts to provide further input that can help explain the results obtained from the classifier based on their clinical expertise. Locally weighted learning (LWL), is another memory-based learning algorithm, that uses weighted linear regression to determine the relationship between an input and an output.

The type of data that may be useful in determining radiomic features include imaging data, outcome or survival data, GTV, dosage data and histology data. For the purpose of radiomics, outcome and imaging data and gross tumour volume have been used to the greatest extent [2]. This is due to machine learning techniques (specifically supervised learning), requiring a baseline or outcome to predict against effectively. Imaging data is commonly used as this is highly dense data that is believed to be full of important ground truth information so it makes sense to use this along with the gross tumour volume which is the expert opinion of the tumour size and location. However, this does not mean other information should be ignored, histology data could be used to find links that are not immediately obvious provided there is enough information available in the first place. Dosage data could also be used to determine how increased or decreased toxicity may link to outcomes.

Dekker et al. [14] developed a promising lung cancer NSCLC decision support system

(DSS) based on routine care data to progress the field of personalised radiotherapy. The decision support system was employed in predicting 2-year survival in NSCLC patients who had received radical and non-radical radiotherapy doses. This study included 322 patients from the Netherlands in the training dataset and 159 patients from Australia in the clinical dataset. Clinical data such as tumour volume, radiotherapy dosage, lung function, age and gender were analysed using the DSS to predict survival. The DSS separated the patients into 2 groups, poor/medium prognosis and good prognosis with high accuracy ($p < 0.001$). It is important to note, that when the DSS predicted a radical dose of radiotherapy for good prognosis patients, this led to almost a 40% improved chance of survival. Hence the ability to separate patients into good and bad prognosis groups can lead to more informed radiotherapy dosages so that the correct dosage can be applied to improve the patient's quality and length of life. This study also found that Bayesian networks, used to predict survival, produced superior results, as they can be used to impute missing data by using existing data to infer the missing data values, this is helpful in clinical data analysis as these databases often lack a significant amount of data.

2.1.5 Future Motivations for Radiomic Analysis

A potential outcome of radiomics could be the development of a tool that can automatically recommend the optimal radio-therapy dosage based on a single CT scan. The MRI-LINAC is a state-of-the-art device which allows a tumour to be scanned and irradiated at the same time. Combining machine learning and radiomics can allow devices such as the MRI-LINAC to be able to detect and irradiate cancer in a matter of seconds saving a great deal of time and unnecessary exposure to radiation, through the implementation of real-time image processing, analysis and identification of cancerous regions.

2.2 Machine Learning and Radiation Oncology

Machine learning can be separated into three main categories: supervised, unsupervised learning and semi-supervised learning. The simplest way to describe the difference between these two is through an example. Imagine you have asked a trainee radiotherapist to segment out a tumour; they know all the theory behind it but need some practice to become an expert, each time they practice, they are told whether they are correct or incorrect, and they go back and repeatedly practice, improving over time as they learn, this is an example of supervised learning. Hence given the relevant information and enough data supervised learning could be used to determine a tumour GTV, similar to how a trainee radiotherapist would do this.

Basically, supervised learning is where you have an image and a label that you are using to find a representation of the image (e.g. features) that most closely relates to

the class or outcome, in this case, the label is the ground truth for the class or outcome. These could still be features that are not visually detectable by the naked eye. Conversely, unsupervised learning has no labeled data. So you are simply looking for some features or representations of the images that best describe them as a whole, for example finding a set of x classes based on a range of features used to describe each image.

Another alternative is, semi-supervised learning that can be used on data which may be missing some information, such as outcome data, but could use dosage data and imaging data to determine relationships which even an expert radiotherapist may not be able to recognize easily, such as rougher textures in an image being related to higher dosages and hence more active tumour regions.

Supervised and unsupervised learning techniques are often combined to achieve superior results. Table 2.2 lists various machine learning techniques that have already been applied to the field of radiomics, summarising the advantages and disadvantages of each technique.

In terms of feature extraction, supervised feature extraction involves starting with a known feature or formula, to then determine if it is significant for predicting ground truth data such as survival outcome. Supervised feature extraction can involve using deep learning or clustering to extract the relevant textural images from the dataset in the case of a CNN, conversely unsupervised learning can occur if you are using an auto-encoder. Unsupervised learning is ideal when you are trying to characterise a dataset but do not know which features will be helpful so one can use machine learning to identify the features automatically.

Table 2.2: This table provides an overview of the advantages of several state-of-the-art machine learning techniques which have been applied in the field of radiomics.

Data Mining Technique	**Learning Paradigm**	**Algorithm Type**	**Advantages**	**Article Ref. that has used this ML technique in radiomics**
Artificial Neural Network (ANN)	Supervised	Deep Learning	Generative Model	[15]–[17]
Bayesian Network (BN) and Naïve Bayes Classifier	Supervised	Bayesian	Can impute Missing data	[4], [14]–[16]
Convolutional Neural Networks [18]	Supervised	Deep Learning	Data-driven solutions	[19]
Decision Trees	Supervised	Classifier	Comparing outcomes to expertise is easy	[4], [15], [16]
k-Nearest Neighbour (kNN)	Supervised	Instance-based	Location based analysis	[15], [16]
Logistic Regression	Supervised	Regression	Linear classifier, fast, doesn't require huge datasets	[10], [11]
Random Forests	Supervised	Ensemble	Fast	[15]–[17]
Support Vector Machines (SVM)	Supervised	Classifier	Linear classifier, Fast	[4], [15]–[17], [20]–[22]
Rule-Based or Ripple down rules	Supervised	Classifier	Comparing outcomes to expertise is easy	[4], [17]
Bootstrap aggregating or Bagging	Supervised	Ensemble		[15]–[17]
Convolutional Deep Belief Networks [23]–[25]	Supervised	Deep Learning	Data-driven solutions. Highly accurate.	[26]–[29]
Fisher Vector [30]	Supervised	Clustering	Linear classifier ideal for distributed learning.	[31], [32]
Gaussian Mixture Modelling (GMM), k-means and k-Nearest Neighbour (kNN)	Unsupervised	Clustering	Requires Gaussian data. Data driven with number of clusters determined by data.	[33]

2.2.1 Machine Learning Techniques Already Applied in Radiomics

Machine learning techniques have been used extensively in radiomics are shown in Table 2.2. While it can be seen that machine learning is progressing quickly in this field it is also important to note that these techniques focus mainly on analysing the radiomic features determined by Aerts et al. [10]. They do not focus on using deep learning to create these features in the first place, which is the focus of this chapter.

Machine learning techniques such as Bayesian Networks, Neural Networks, Decision Trees and Support Vector Machines (SVM) have been employed in numerous studies to predict 2-year survival [4], [14]–[16]. The majority of these studies have chosen to use SVM, SVMs are often used to distinguish between two different classification options, it is important to note that there are also multi-class SVM schemes that are used to overcome this weakness. Dual classification is adequate if one simply wants to predict a patients likelihood of survival after 2-years, however if one needs to produce a more in-depth analysis other ML techniques should be considered when analysing medical images.

A promising early study related to radiomics by Deasy et al. [22] utilised principal component analysis (PCA) to uncover linear behaviour then combined this with non-linear kernel SVM method to achieve superior results in predicting disease endpoints, this opened the validated the need for statistical learning in radiotherapy to improve personalised treatment outcomes.

Parmar et al. [15] investigated different machine learning techniques and feature selection techniques for NSCLC lung cancer patients. This paper compared twelve ML techniques, including but not limited to, neural networks, decision trees, SVM, random forests, nearest neighbour and discriminant analysis. In this study random forests were found to have the highest overall accuracy, for determining 2-year patient survival. This study including 310 patients in the training cohort and 154 patients in the validation data set. When conducting a study in machine learning at least 300 patients are needed for most techniques [16], however techniques such as neural networks can have far greater performance when a much larger data set is being analysed [34]. To improve upon this study a far larger cohort of patients should be included to determine the best method for determining 2-year survival in NSCLC cancer patients.

Hawkins et al. [4] compared four ML techniques including decision trees, rule based classification, naive bayes and SVM with various feature selection techniques such as test-retest, relief-F and correlation based feature selection. This study included 81 adenocarcinoma NSCLC lung cancer patients in the training cohort. This study concluded that the decision tree once validated with a leave-one-out-cross validation approach obtained the best accuracy for predicting 2-year survival. Two key draw backs of this study are studying a small dataset and not employing a separate cohort to validate the classifier as this can mean the classifer is not as robust and is unlikely to perform as accurately on a

different cohort.

2.2.2 Feature Selection compared to Feature Extraction

Computer Vision is a general term used to describe the broad range of techniques for processing images by computers to understand or derive some useful information from digital images, these image processing techniques include pattern recognition, analysis, classification and segmentation, to name a few.

In radiomics, computer vision may be used to derive image features to describe automatic image segmentation, texture analysis and object detection in order to detect landmarks in an area in order to infer more information about the object or tumour.

It is important to note the difference between feature extraction and feature selection. The goal of feature extraction is to find as many features as possible to describe the data. Feature selection is where the set of such features are reduced to as few features as possible that can represent the data as robustly as possible, while at the same time avoiding the issue of over-fitting the data. Over-fitting the data is a serious problem in machine learning as it this often means that the analysis may have excellent results when applied to the training data, but when new data is presented for analysis it is likely that the model will have very poor results in comparison.

Feature selection is currently the main method of determining radiomic signatures in the literature as used by the authors [1], [10], [11], [35], [36]. Conversely, Ypsilantis et al. [19] has employed convolutional neural networks (CNN) for feature extraction in the field of radiomics.

There are two main approaches for feature extraction being deterministic and non-deterministic extraction. Deterministic feature extraction is the most common method where a mathematical formula is employed to extract features relating to imaging features such as texture, intensity or shape, this method is employed in the following papers [10], [15], [16], [35]–[40]. The previous proven method of deterministic feature extraction were based on Haralick textural feature analysis [41].

There are numerous dimensionality reduction methods for selecting and minimising the number of features once found. These include PCA, linear discriminant analysis, local linear embedding (LLE), mixture of factor analysers (MFA) and t- stochastic neighbourhood embedding (t-SNE).

2.3 Machine Learning in medical image processing

In this section the latest state-of-the-art computer vision and machine learning techniques are explored in relation to solving the two key issues in radiomics namely feature extraction and analysis. Some of the techniques to be discussed include 3D SIFT, the Fisher Vec-

tor (FV), GMM, Decision Trees and SVM. The machine learning techniques discussed in this section have been selected based on their ability to detect textures in images, as it has been found that texture detection is an important aspect of feature extraction in radiomics. These machine learning techniques are summarised in Table 2.2.

2.3.1 Fisher Vector

This section describes how the Fisher vector combined with 3D SIFT can be applied to radiomics. The Fisher vector and 3D SIFT have been combined with CNN to perform traditional machine learning based classification in [31], [32].

Sanchez et al. [30] introduce the Fisher vector (FV) as an alternative supervised learning classification technique based on GMM that has been found to offer state-of-the-art competitiveness in classifying images. The Fisher vector is based on the Fisher Kernel patch encoding technique which uses GMM to describe an image as a finite number of clusters. This is an alternative to the comparable popular representation, the Bag-of-Visual words (BoV). The Fisher vector is a high dimensional vector which has been found to have state-of-the-art accuracy and speed for classifying various image datasets with over 8 million images [30]. This method employs product quantization to improve efficiency and accuracy by reducing the data with negligible losses.

To compute the fisher vector the image is firstly separated into windowed voxels of the tumour, this ensures that the location of the clusters can be retrieved easily. The size of the window may be varied experimentally to determine the window size that achieves optimal results. Local features can be calculated using more traditional radiomics analysis such as Gray-Level Co-occurrence Matrix (GLCM) or LoG based features which have been found to be optimal for predicting distant metastasis in lung adenocarcinomas by Coroller et al. [36].

2.3.2 Gaussian Mixture Modelling

GMM may be used to cluster these features into a finite vocabulary. This is a key difference between the BoV words approach and the Fisher vector as the BoV method employs K-means clustering instead of GMM. The main difference between K-means and GMM is that in K-means clustering you pre-determine the number of clusters that will be formed, however GMM generates as many clusters as is needed to suit the data. In the case where you know the exact number of clusters that are required K-means is superior, however in the case where the total number of clusters is unknown and more clusters are required for a more in depth analysis, GMM is preferable. The feature vector generated from this calculates two very important GMM variables being the mean and variance, for each windowed region, the regions are then clustered based on closest means and furthest variance. This ensures that the clusters similarities and differences are accentuated so that the

clusters are as precise as possible.

After the clusters are computed into a finite vocabulary a histogram can be generated to form words (areas) inside the tumour region. Ideally this histogram should have a large vocabulary of words to analyse so that more information in the form of recurring patterns or motifs can be extracted for classification purposes. At this juncture, 3D SIFT presented by Scovanner et al. [42] as a feature descriptor for 3D images can be used to interpret the words (areas) in the 3D tumour image into signatures which can be utilised for improved efficient linear classification. Paganelli et al. [43] have employed SIFT features for tracking tumours in 4DCT scans and for image registration of CBCT scans.

2.3.3 Supervised Classification

Supervised classification may be employed to predict the overall survival likelihood model from the resulting fisher vector, through comparison with outcome data. For supervised classification there are two main methods including, Variational Bayesian GMM (VB-GMM) and Minimum Message length (MML). VB-GMM is currently the favoured method based on strong theory and is currently more widely accepted in the machine learning community than MML [44]. The key disadvantage of this method is that it produces a high dimensional vector, however despite this it has been proven to perform better for linear classification that BoV making it far more suited to distributed learning.

Distributed learning in this case describes data coming from different sources that is be integrated together and normalised for consistency. Being suited to distributed learning is an essential criteria for this type of research, since to improve models in radiomics, data will need to be collected from multiple hospitals from multiple nationalities in order to achieve a generic model which can be applied in all situations and yet still provide personalised cancer therapy. Lambin et al. [45] stresses the need for a distributed decision support system where instead of sharing data institutions may share a model to describe the data such as the result from a supervised machine learning technique such as SVM, logistic regression or bayesian networks. The models from each individual institution can then be combined to form a consensus with higher accuracy as discovered in [46], [47]. Conversely, Skripcak et al. [48] discusses the implementation of a federated international database containing thousands of de-identified patient data sets from multiple institutions around the world. This type of international database would be ideal for unsupervised learning techniques such as CNN.

There are many alternative classification methods in machine learning such as decision trees, SVM and logistic regression. Their ability to perform well in radiomics depends more on the type of output required. SVMs and logistic regression are useful in separating cohorts into two different groups, for example good and bad prognosis, but they fail to provide much more information than this. Decision trees can be used to separate cohorts

into multiple groups, with the bonus being that the results are completely interpretable, which can allow an expert to understand the result and can allow them to treat patients based on fully informed decisions.

2.3.4 Unsupervised Clustering

Clustering is an application of unsupervised learning used in machine learning to categorise and analyse data. Clustering techniques such as k-means, GMM or fisher vectors can be used to window the tumour into regions that can be clustered to show areas of similarity and the inverse. These clustered areas can then be classified into different tissue, such as necrotic tissue, normal tissue and proliferating/aggressive tumour. By using clustering to locate a tumour, normal tissue toxicity can be reduced by identifying the healthy tissue regions to be avoided when radiotherapy is applied.

Once these clusters have been determined by experimentally varying the window size they can be compared to outcome data to determine if these clusters are representative of bad or poor prognosis. Similarly the clusters can be compared to outcome and dosage data to determine the optimal Grays (Gy) dosage level for a specific type of cluster in Head and Neck (H&N) cancer.

Prise et al. [49] reviews the state-of-the-art research in radiotherapy by analysing the effects of radiotherapy on healthy tissue known as the bystander effect where cells that are not irradiated receive signals from irradiated cells that leads to unforeseen tissue damage. The reviewer suggests that future studies should focus on using multiple small beams to irradiate a tumour rather than using a uniform large beam of the target area, in order to minimise damage to healthy tissue.

In addition, more traditional radiomics techniques can be applied to these clustered regions to find textural features and radiomic signatures that may be related to likelihood of survival. Another potential outcome of this study is that by comparing the clustered regions to genomic data, it may be found that certain clusters have poor or good prognostic depending on the genes of the patient, thereby further personalising the patients treatment in the hope of achieving improved patient treatment.

2.4 Deep Learning in Medical Image Processing

In this section deep learning techniques are explored in detail. Some of the techniques to be discussed include CNN, deep belief networks (DBN) and deep autoencoders. These deep learning techniques discussed in this section have been selected based on their ability to detect textures in images, as textural detection is an important aspect of feature extraction in radiomics.

Deep learning has exponentially increased in recent years in relation to applications in

medical imaging, for two main reasons. Firstly, the amount of available medical imaging data has increased, secondly computers have increased drastically in processing power. However, there is still one main hurdle in applying deep learning to medical images, being that deep learning, which for the majority of cases is also supervised learning, requires extensive amounts of labelled medical data which can be difficult and time consuming to obtain in the medical field [50].

There are six different deep learning architectures namely, deep neural networks, deep autoencoders, deep belief networks, deep boltzmann machines, recurrent neural networks and CNN. The majority of these architectures perform better with a stream of data as opposed to imaging data, which makes them unsuitable for application to medical imaging. However CNN are ideal for 2D imaging data, recently CNN have been expanded to 3D imaging data. It is important to note that tumour detection is a 3-Dimensional imaging problem and thus it is essential to utilise a machine learning technique that can handle this type of data well [50].

2.4.1 Convolutional Neural Networks

An important paper by LeCun et al. [18] gives a detailed review of how deep learning is revolutionizing computer vision through the use of CNN. CNN is a biologically inspired technique based on how the human brain translates visual inputs; it was originally developed in the late 90's but was considered too slow for efficient pattern recognition. However, with the advancement of modern computers this technique has now been found to be superior to other more traditional supervised machine learning techniques, in interpreting and classifying raw data. The conventional supervised machine learning techniques struggle to process raw data in one step, the data needs to be firstly pre-processed, then features are extracted and selected using mathematical models and then finally prediction algorithms can be applied. In addition to this expert knowledge of the data is often required to interpret the features correctly, which is very time consuming.

CNN are a feed-forward network that have been applied to image processing with state-of-the-art results and can be employed to process information for 3D video or volumetric images, which makes it promising for feature selection for volumetric tumour imaging data. The main limitation of this technique is that it requires a large amount of processing and is highly dependent on the information that is fed into the training algorithm, so this data needs to be a good representation of all the different types of outcomes.

The process for creating CNNs can be described in four main steps. Firstly, the convolutional layer is formed where the input data or image is convolved into several small kernels or filters. Secondly the non-linear activation layer is applied, this allows the CNN to model relationships that are nonlinear. The activation usually after each convolutional layer is critically important to the learning process. Thirdly, the pooling layer is formed

where the resulting kernels from the previous step are down-sampled, often via max pooling, which involves separating the image into regions with the highest kernel value from each region being outputted to the next step. These first three steps combined will produce a one-layer CNN. The fourth step involves repeating steps one to three, with the only difference being using the output from the second step as the input. Higher layer CNNs can often produce higher level features similar to the features extracted in the field of radiomics. In this way a CNN can be used to convert 3-Dimensional images into 1-Dimensional vectors to allow formal classification through conventional machine learning methods [50].

A common question in this area is how much data is required to produce useful results in deep learning. While most deep learning applications such as CNN require very large data sets in the order of thousands of examples, this does not necessarily imply thousands of patients are required to produce useful results. In the case of radiomics, this involves very large radiology imaging data sets which may only have a few hundred individual patients. In these cases CNN can be employed to break up the radiology images into smaller patches that can be used to train and classify the data, thereby supplying hundreds of thousands of images from only a few hundred patients. Recent deep learning studies have involved performing CNN on radiology, histopathologic and cell images, which included less than 100 patients, including the validation data set, have produced superior classification results in this area, despite the small number of patients in this studies [29], [31], [32].

2.4.2 Convolutional Deep Belief Networks

DBN are often combined with CNN to model the data effectively. Hinton et al. [25] developed an advanced learning algorithm for DBN which is often combined with the CNN algorithm. DBN develops a generative model with multiple tiers that finds statistical similarities between tiers to increase the speed of training data while maintaining data integrity.

Lee et al. [23] and Wu et al. [24] combine CNN and DBN with promising results. Lee et al. [23] introduces a technique known as probabilistic max-pooling that reduces the data to be analysed and thereby increases the efficiency, however Wu et al. [24] avoids max-pooling as this would increase the uncertainty for shape reconstruction for pattern recognition of 3D point cloud images. Lee et al. [23] demonstrates high accuracy for multiple pattern recognition tasks for large image sets. Wu et al. [24] employed a convolutional deep belief network (CDBN) to recognise 3D shapes on a 3D voxel grid. This technique was found to recognise 3D shapes with high precision and efficiency with low quality input data with state-of-the-art performance in numerous tasks. CDBNs have been applied successfully to medical imaging tasks in [51], [52].

Deep Autoencoders have also been applied to radiomics for feature extraction through data driven learning. This method employs unsupervised learning, hence labelled data is not required for training. There are many variants of Autoencoders, with the key one to radiomics being the convolutional Autoencoder. The convolutional autoencoder, this encoder preserves spatial locality and can be applied to 2-Dimensional images [50]. Deep Autoencoders have been applied to medical imaging data in [53], [54].

2.4.3 Artificial Neural Networks

Conventionally, radiomic features determine a single scalar value to define a complete 3D tumour volume, however, more recent research employs pixel-based features generating many values per feature for a 3D tumour volume [79].These features can then be fed into a classifier to determine the features which have the strongest correlation with outcomes. An example of a useful classifier for this would be a decision tree. The classifier starts with the feature which has the highest correlation with outcomes and then progresses down the tree using the next best feature. In this way machine learning can be used to determine the most significant features and the combination of features that provides the greatest predictive capability; eliminating redundancy. Redundancy refers to individual features that have a strong correlation with other features or feature clusters. Removing redundancy in mathematical modelling reduces conflation of errors.

ANNs are data driven and produce results that are limited by the quality of the input data. In radiology and nuclear medicine, an image or set of images may be fed into a CNN or extracted radiomic features can be used as the input for an ANN. The CNN, however, is a powerful tool in identifying and extracting its own radiomic features from the input image and link these to the outcome for improved results.

In oncology, since tumours are 3D volumes the question arises whether it is better to simplify this 3D information by generating scalar mathematical radiomics features or whether it is better to treat the images as data and enter the raw 3D data into a CNN. Data driven approaches like CNNs, however, have a higher risk of over-fitting to the original training data and not being generalisable to unseen data. Recent developments utilising 3-Dimensional conditional generative adversarial networks (3D c-GANs) contain a generator and a discriminator network in opposition. These networks produce positive and negative input data from the original data, training the network for greater generalisability.

2.4.4 Limitations of Deep Learning in medical imaging

The availability of open source deep learning packages [55]–[63] has led to the rapid adoption of deep learning in the field of medical imaging however in order to apply these packages effectively profound expertise in biology and computer science is required [50].

Deep learning can often suffer from the issue of over-fitting, where it performs with minimal errors with the original training data, however when new data is introduced it can sometimes struggle to perform well. This problem can be solved by introducing a larger training set that can model the population more accurately.

The results from deep learning can also be difficult to interpret using clinical expertise which is non-ideal when they are being used to aid in decision making. Decision trees may be preferable in this case as they can show how the results were generated and which factors are most important in influencing the final result. Also cytological notes referring to tumour diagnosis regarding stage and spread can be included in medical data to aid in interpreting the results obtained through deep learning [50].

2.5 Conclusion

Radiomics is a field that is advancing rapidly with a vast potential for increasing the ability to diagnose and treat cancer effectively. In order to be successful in this field, significant expertise is required in both biology and computer science. This will enable computer vision and machine learning techniques to be applied effectively which has huge potential in solving the issue of cancer analysis in medical imaging. The limitations for 3D analysis with Deep learning make it insufficient for feature extraction and instead it should be considered as a better tool for the classification of radiomic features that are extracted using mathematical means across a 3D tumour volume.

Chapter 3

Deep Learning Classification For Lung Cancer Prognosis

3.1 Introduction

Personalised medicine aims to optimise outcomes by individualising cancer treatments. Research applications in medical imaging using deep learning have intensified in recent years, due to increased availability of data, in addition to the parallel processing by computers which now have increased throughput.

In this chapter CNNs are applied to this problem to extract new features that may be useful in predicting clinical outcomes such as patient two year survival likelihood.

In this research we explore the features identified by a CNN for non-small cell lung cancer (NSCLC) tumours. The gross tumour volume (GTV), is used as the ground truth in this study as it is the oncologist's best estimate of the whole tumour volume. This is the volume that is delineated as the tumour by the oncologist, and this will be used as the mask to interpret the cancer cells from non-cancer cells. For non-small cell lung cancer, accurate two year prediction models are preferred as this cancer is aggressive and has a poor two year survival rate.

3.2 Related Work

CNNs have been applied to medical imaging data with great success [29], [31], [64]–[67], Shin et al. [64] used a CNN to compare hand crafted SVM features to improve the classification and automatic detection of polyps. Wang et al. [65] employed a MV-CNN to segment lung nodules with improved results. In a similar study, Huang et al. [29] employed a 3D-CNN to detect lung nodules in CT images.

Al-masni et al. [66] also used a CNN to detect lesions in breast cancer. Stanitsas et al. [31] employed CNNs to locate benign and malignant tissue in breast and prostate

cancer. Similarly, Wang et al. [67] examined histopathological images using a bilinear CNN to classify the type of stain composition within an image. The above recent deep learning studies have involved the use of CNNs in radiology, histopathologic and cell images for less than 100 patients, including the validation data set, and have produced superior classification results [29], [31], [64]–[67].

Many authors have undertaken the extraction of various radiomic features from images to determine if there is a correlation between these mathematically determined features and survival, and many have found recurring metrics, also strongly correlated with survival [10], [11], [14]. Deep learning can be used to determine which metrics are strongly correlated with outcomes for all cancer types but more importantly, it can be used to determine relationships that have not even been considered by allowing neural networks to determine the key radiomic features.

There are two main approaches for feature extraction; deterministic and non-deterministic extraction. Deterministic feature extraction is the most common method where a mathematical formula is employed to extract features relating to imaging features such as texture, intensity or shape, this method is employed in the following papers [1], [10], [11], [14] in the form of radiomics. Conversely, Ypsilantis et al. [19] has employed Convolutional Neural Networks (CNN) for non-deterministic feature extraction.

3.3 Proposed Methodology

This analysis consists of four main stages; converting the DICOM-RT files into matrices within MATLAB, separating the patients into categories of passed or failed in the two year survival test, applying the GTV mask to the images and training the images to train a suitably structured CNN.

422 NSCLC patient data sets were publicly available at the time of publication, 312 were usable in this study, the files were obtained from [10], [68], [69]. 110 patients were excluded from this study as 2 year survival data was not available for these patients.

Firstly, the patients were separated into two separate folders based on two year survival. One third of the patients perished within 2 years, the other 2 thirds survived 2 more than two years after being scanned. For training the same number of each survival type were used, to remove training bias. The CT images were then normalised, to ensure that each voxel width had the equivalent pixel size namely, $0.0977 * 0.977 * 0.300mm$, while also ensuring that the image sizes were exactly the same, $471 * 525$ pixels.

The GTV mask was subsequently applied to every CT image to eliminate unnecessary data from the analysis. These were then saved as JPEG images. To ensure the network trained fairly, the training data was separated to include all training patient slices in the training data, so that the neural network would not learn off slices from the same patient. A near equal number of pass and fail cases were included, for a total of 1993 images in

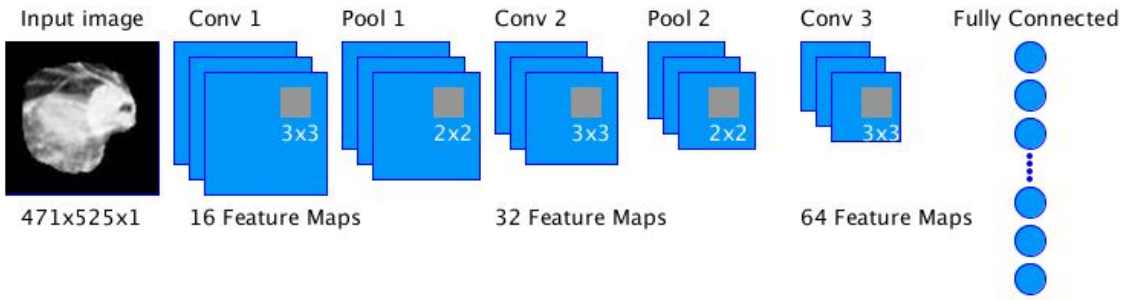

Figure 3.1: This shows an overview of the 3 Layer CNN produced in this study

the training set out of the 4992 images in the overall dataset across 312 patients. The remaining images were used to validate and test the CNN model.

The basic structure when forming a CNN has three main stages, various of these are also often repeated a number of times in the final form. Firstly, the convolutional layer is formed where the input data or image is convolved into several small kernels or filters. Secondly, the pooling layer is formed where the resulting kernels from the previous step are down-sampled, often via a max pooling process. This involves separating the image into regions with the highest kernel value from each region selected and conveyed to the next step. These first two steps combined will produce a one-layer CNN. The third step involves repeating the first and second steps, with the output from the second step serving as the input to the first step.

The layers of the convolutional neural network were designed for this study with the goal of reducing the training time as much as possible. There were 12 layers included in total, producing a 3-layer CNN, seen in Figure 3.1. A 3-layer CNN was chosen so as to minimise over-fitting. The first layer is the input layer, to define the size of the images to be input into the network. The second layer included the convolutional 2D layer where the image is convolved into small kernels or filtered images. The third layer included the Relu layer which is the activation function that introduces non-linearity in the function that is learned, this is part of the optimiser function which applies gradient descent to find the local minimum. The rectified linear activation (Relu) function transforms the summed weighted input into a positive output. The fourth layer is the max pooling 2D layer which selects the largest values to output to the next layer, to down-sample the kernel images. The second and third were then applied twice more. With the seventh layer as a second max pooling layer. The last three layers are the fully connected layer which contains the CNN features found, followed by the soft max layer and the classification layer, which matches each CNN feature to a classification.

The neural network was trained using stochastic gradient descent with momentum, at a learning rate of 0.001, with a minimum batch size of 10. The epoch was varied from 10 to 70 in steps of 10 to determine the best accuracy, by finding the knee point at which the test data lost accuracy due to over-fitting, see Figure 3.2. The knee point is the point in

the data where the accuracy begins to plateau with minimal improvement from additional training. The best accuracy was achieved at 40 epoch using the test data.

The results obtained from the trained neural network were then compared to the results from a logistic regression classifier developed by Vial et al. [3] on the same data after the radiomic texture features were extracted using gray-level co-occurrence matrices. In the study conducted by Vial et al. [3], a 10-fold cross-validation approach was employed to evaluate the models. Each fold of the cross-validation was stratified by the number of outcomes in the total cohort. The logistic regression output was then employed to produce the two year survival Kaplan-Meier curves. These curves are often used in medical research to predict survival by stratifying survival likelihood into two different patient groups [10], [11], [14]. The data was then separated into a pass and fail prognosis group separated by the median values of the output probability. The key difference between this study and the study by Vial et al. [3] is that radiomic analysis was employed to extract the imaging features, alternatively this approach uses CNN to extract the features. It is important to note that radiomics extracts features across an entire 3D tumour volume whereas CNN only extracts features from a single slice.

3.4 Results

The CNN once trained, achieved an overall accuracy of 55.60% by finding the average of the diagonal for the confusion matrix. 1667 images were correctly predicted and 1332 images were incorrectly predicted from the test data for a total of 2999 images.

To improve these results various 2D CNN approaches were considered with limited improvement. The training set was altered and expanded to include alternative training data, this showed no improvement. The training set was increased to include 3500 images with 1500 remaining for testing, this led to the data being over-fitted. A 5-layer CNN was also implemented which again led to the model being over-trained.

A significant increase in performance was found when the test and training sets were randomised based on slices rather than patients, reaching 83%. However, this allowed the CNN to become over-trained or essentially cheat by learning the location of the tumours rather than training on the textural composition of the tumour image.

Table 3.1: This table shows the performance for the CNN compared to the classifier from Vial et al. [3]

Classifier	Overall Accuracy
CNN	55.6%
Texture Features	66.1% [3]
Three Features	58.9% [3]

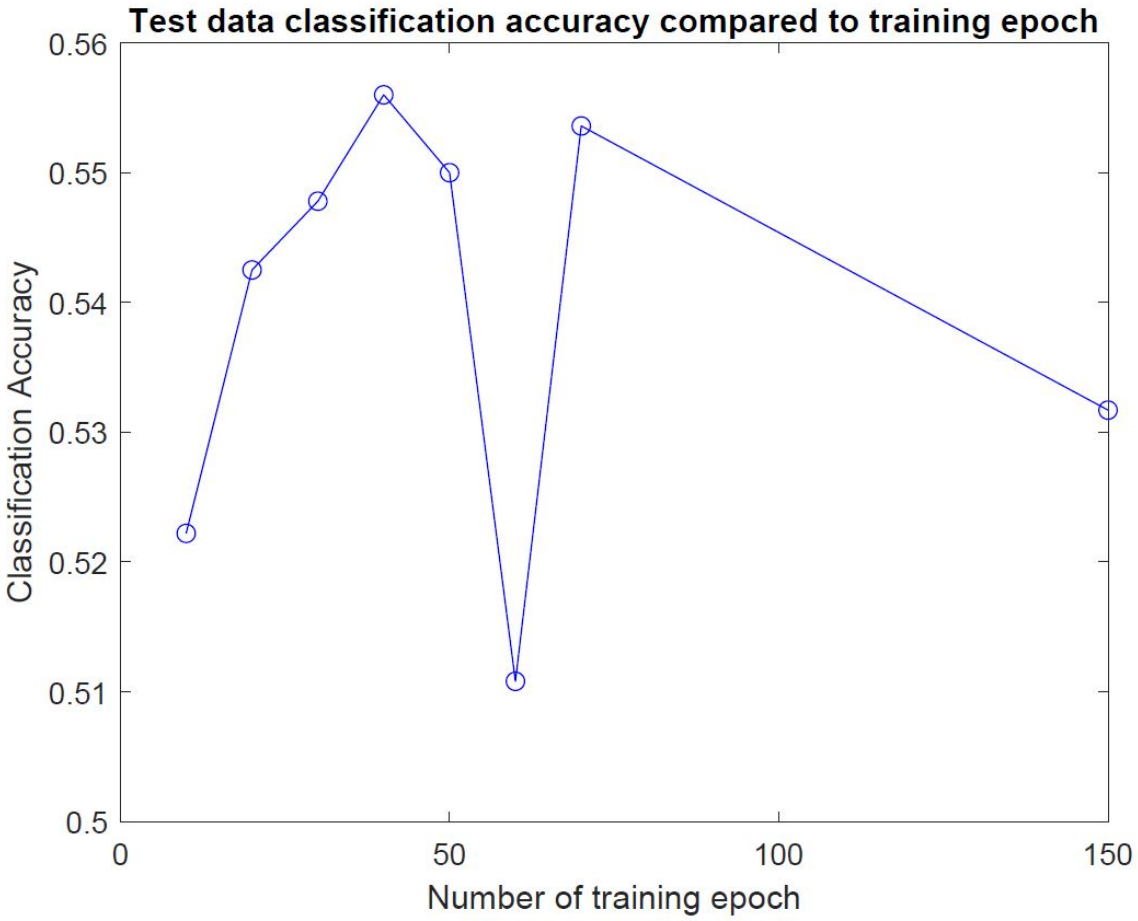

Figure 3.2: This figure shows the classification accuracy of the CNN at varying Epoch values, to determine the knee point where the CNN has been over trained.

Table 3.2: This table shows the confusion matrix generated for the CNN model results at 40 epochs, for survival prediction after 2 years.

Class	Predicted did not survive	Predicted survived
Predicted did not survive	52.22%	47.78%
Predicted survived	41.03%	58.97%

3.5 Discussion

In this study the network was trained using 1993 images out of the 4992 total images, that is one third of the data. The CNN would likely achieve a greater overall accuracy if more of the data was used to train the CNN, however this could also lead to the network being over-fitted. In this study the selection of subset of patients for training were varied several times and the performance did vary a little across the altered subsets.

It can be seen in Figure 3.3 that While the entire database was not examined to test this theory, a number of other examples were checked to confirm that it is often the case that images with larger size, texture and shape variations in images did not survive beyond 2 years. This is in agreement with the literature which suggests that texture, shape and size

Figure 3.3: These exemplars show the example input for the patients of a single CT slice for NSCLC image set that were correctly classified by the CNN.

features are the most important imaging features for the field of radiomics [10], [11].

CNNs are feed-forward networks that have been applied to image processing with impressive results and could be employed to process information for 3D volumetric images. CNNs are limited in that they require a large amount of processing power and are highly dependent on the information that is fed into their training algorithm.

The results obtained in this chapter are un-characteristic of CNNs. There are many possible reasons for the poor CNN results including but not limited to the following: the CNN may have insufficient weights or trainable parameters; the class distribution may be imbalanced, and hence there may be a need for re-balancing or loss function modification, or the image is not centreed on the tumour. It is suggested, there is a need to center the tumour in the image on a per patient basis using the 3D GTV in future studies, so as to increase the ability of the CNN to detect unique features within the image. These results also suggest that CNNs may be promising for feature selection of volumetric tumour imaging data, provided the above improvements are included.

One limitation of other radiomic studies is in the manual selection of the radiomic features, which are determined based on general mathematical formulas associated with analysing a tumour volume. [3], [10], [11]. In this work we use CNNs [18] to determine which features have the highest prognostic significance for determining two year survival, to produce alternative radiomic CNN based features.

A main hurdle in applying deep learning to medical images remains, namely that deep learning requires extensive amounts of labelled medical data and this is difficult and time consuming to obtain within the medical field [50]. CNNs can also often become over-fitted to the data they are trained on making them difficult to apply to new datasets if there is not sufficient data to train them on. A future implementation of this work would involve testing the CNN on a new similar NSCLC dataset to see how the accuracy is affected.

3.6 Conclusion and Future Work

In this exploratory chapter, we considered the ability for CNNs to be applied to radiomics for 2 year survival outcome prediction.

CNNs have great potential to produce improved results in survival prediction, through extracting new radiomic features from whole tumour volumes. However, this study only achieved an accuracy 56% for 2D CNNs on a large unseen test data set.

Despite CNNs having great results in other fields it appears that this is not the best way forward for radiomics. The reason for this is, 3D information is needed in order to provide better predictive ability. This is the key power of radiomics, where one value can be used to describe the whole or a segment of the tumour volume.

Chapter 4

Tumour Rind Analysis For Survival Modelling

4.1 Introduction

Radiomics is an analytic approach to extract quantitative features, such as shape, texture or intensity, within radiological imaging data, such as CT, PET and MRI, by utilising existing image processing techniques. The information from radiomic analyses can be correlated with patient data, such as oncologist-defined ROI, outcome (tumour control, survival), and tumour phenotype represented in histopathology, or biomarkers. Radiomic information could inform the selection of cancer therapy for an individual patient and tumour, and could predict clinical outcomes. The four main stages of radiomic analysis include imaging, segmentation, feature extraction and analysis for classification and prediction of an outcome such as two-year survival, see Figure 2.1.

In this study we explore the textural radiomic features of a NSCLC tumour defined by an oncologist. The marked tumour volume is denoted as the GTV.

The GTV delineated by the oncologist encapsulates the whole 3D volume. Taking this volume, we subsequently define a layer of known thickness of tissue outside the GTV-line by adjoining the GTV (outer rind), and a corresponding layer of identical thickness inside and abutting the GTV-line (inner rind) [70]. These volumes were then used for the analysis of radiomic signatures and two-year survival outcome. Figure 4.1 illustrates the inside and outside rinds of the tumour volume, where green indicates the outside rind and blue the inside rind, furthermore red shows the gross tumour volume (GTV).

This thesis explores the effect on predicting two-year survival by including only imaging textures discussed in Aerts et al. [10]. We used the same public NSCLC dataset that was utilised in [10] to predict the two-year survival. To the best of our knowledge this is the first study exploring the rind of a tumour using radiomic texture analysis. The rationale for observing the rind region is to define a strategy for partitioning and analysing

the texture of tumour regions so as to characterise the relationship between texture and survival in NSCLC patients. The reason we decided to investigate the rind is because different oncologists will delineate a tumour differently, so there may be tumour cells located outside the GTV that may have prognostic value. There is also suggestions in the literature that tumour heterogenity or shape is an important radiomic feature so less circular tumours have worse survival and hence it is important to ensure the outside of the tumour is delineated accurately.

4.2 Related Work

An earlier report demonstrated that four radiomics features could predict the two-year survival of lung cancer patients in a public NSCLC dataset with a high degree of accuracy, in Aerts et al. [10]. We have undertaken a similar analysis using three of these identified highly prognostic features to understand whether the presence of these radiomic features in the inner and outer rinds around the GTV-line can also predict two-year survival. Due to unforeseen technical issues the fourth wavelet feature was not able to be included in this study. These technical issues related to out of bound matrix issues, these were corrected in the subsequent study, hence wavelet features are included in the next chapter. The construction and use of these rinds in radiomics analysis is unique.

Subramaniam et al. [13] extended the previous work in NSCLC by quantifying the heterogeneity of SUV_{max} values from a PET scan. By using Kaplan Meier survival analysis combined with Cox Proportional Hazard regression, they were able to identify 'good' and 'poor' prognosis groups with an improvement in survival prediction.

Tumour heterogeneity for NSCLC patients was further studied by Ganeshan et al. [71], using the technique computer tomography texture analysis (CTTA), that allows coarse and fine image texture features to be filtered, showing a high correlation with survival. Textural analysis has been utilised for classification of medical outcomes in the past [72], [73].

Given that NSCLC commonly develops distant metastases (DM) leading to patient mortality, Coroller et al. [36] sought radiomic signatures which were prognostic for DM in NSCLC patients, again improving survival prediction. They found that Gray-Level Co-occurrence Matrix texture features were useful. Vallières et al. [11] undertook a similar study examining a GTV defined on pretreatment MRI fused with a PET scan in soft tissue sarcomas (STS) to evaluate the risk of DM in the lung. They showed that the fused MRI/PET image set had a superior performance in predicting DM.

These related works reveal the potential for radiomics to improve the accuracy of outcome predictions. Until now, reported analyses have focused on the tumour (GTV) features alone. We have begun to analyse the lung exterior and surrounding the tumour also. This approach is supported by the approach of radiation oncologists who encapsulate the

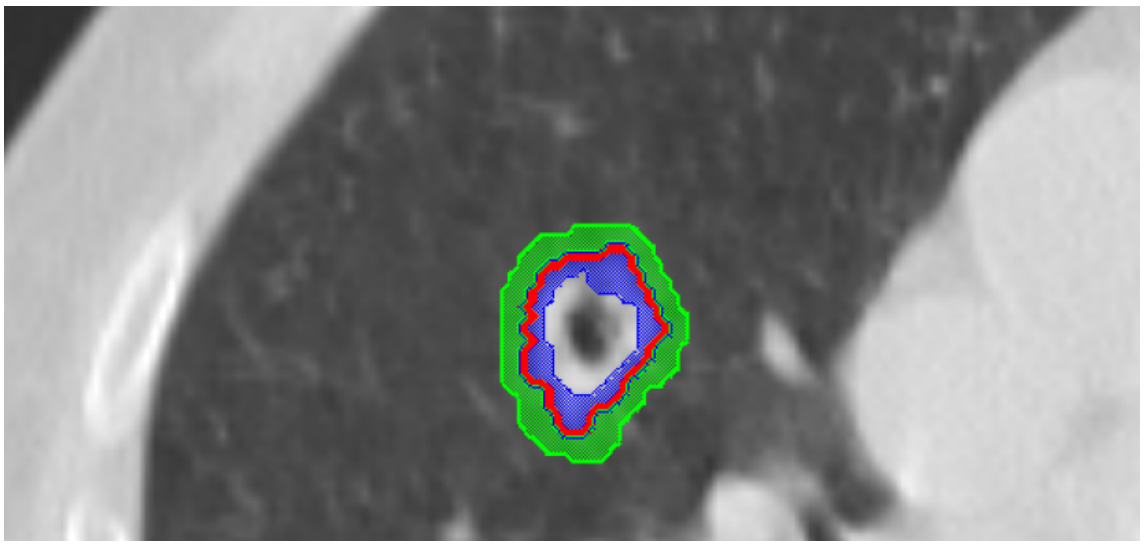

Figure 4.1: This image illustrates the rind concept shown on a single CT slice of a patient NSCLC image set. The inside and outside rinds of the tumour volume, where green indicates the outside rind and blue the inside rind, furthermore red shows the gross tumour volume (GTV).

visible GTV within a CTV which defines the at risk areas which appear to be normal to the naked eye.

4.3 Proposed Methodology

This analysis consists of four main stages; converting the DICOM-RT files into matrices readable in MATLAB, calculating the rind volume for extraction from the CT data, determining the radiomic features and finally producing the two-year survival analysis.

The DICOM-RT format is the standard medical imaging format used to store medical imaging data and is paramount in the efficacious application of the picture archiving and communications systems (PACS) used in most hospitals for radiation therapy [74]. It is essentially a file format that allows multiple CT images to be related using an overarching RTSTRUCT file which describes the oncology delineation data and the dosage data for the patient.

422 NSCLC patient data sets were publicly available at the time of publication, 245 were usable in this study, approximately 14.2GB from over $29,000$ CT image files were examined, with the files obtained from [10], [68], [69]. Only 245 patients had survival periods of at least 2 years at the time of this study. In subsequent studies time progressed and more patients passed the 2 year mark, hence the difference in usable patients between chapter 4 and chapter 5.

To derive the mask, we first define the set of points representing the inside and outside rind, let $\mathbf{x} \in \mathbb{R}^3$ and $\mathbf{z} \in \mathbb{R}^3$ define positions vectors in the 3-Dimensional image set frame. Further, let $\mathscr{V} \subset \mathbb{R}^3$ be the set of all points in the GTV as given by the expert or any other

segmented region of interest (ROI) and $\widehat{\mathscr{V}} \subset \mathscr{V}$ is defined as all points on the boundary of this region.

Suppose there is a distance function that can be constructed for each ROI, $d(\widehat{\mathscr{V}}) : \mathbb{R}^3 \rightarrow \mathbb{R}$, which will be defined by the following

$$d(\mathbf{x}, \widehat{\mathscr{V}}) = \begin{cases} \min \|\mathbf{x}-\mathbf{z}\|, \ \forall \mathbf{z} \in \widehat{\mathscr{V}}, \ \forall \mathbf{x} \notin \mathscr{V} \\ -\min \|\mathbf{x}-\mathbf{z}\|, \ \forall \mathbf{z} \in \widehat{\mathscr{V}}, \ \forall \mathbf{x} \in \mathscr{V} \end{cases}$$

Now the outer boundary region can be expressed by the set

$$S_o = \{\mathbf{x} \mid 0 < d(\mathbf{x}, \widehat{\mathscr{V}}) < L\}, \tag{4.1}$$

where L is the width of a region that contains all line segments perpendicular to the surface. S_o is therefore a region external to the ROI. Conversely, we have

$$S_i = \{\mathbf{x} \mid -L < d(\mathbf{x}, \widehat{\mathscr{V}}) < 0\}, \tag{4.2}$$

defining an inner region such that $S_i \subseteq \mathscr{V}$. Colloquially, we refer to the region, $S = \{S_i \cup S_o\}$, as the rind of a given ROI, obtained by expanding or contracting the ROI uniformly from the surface by L. This can be visualised in Figure 4.1 where the 3D rind is found by first finding the 2D rind for each image.

This rind was then dilated and contracted to ensure there were no pixels missing within the rind, creating a rind exactly $2 \times n$, being 6 pixels deep in this study, all the way around the 3D GTV, this is an expansion of approximately 6mm around the whole GTV volume. This can be seen in Figure 4.2.

We then extracted the three features recognised by the authors in. [10] as having prognostic significance and compared the two-year survival predictions for the whole volume compared to the inside rind and outside rind alone. These features included the statistical energy, shape compactness and the run length non-uniformity of the image, these features are used for comparison in this study.

Let $\mathbf{X}$ signify the 3-Dimensional image matrix, containing N voxels, energy is given by Equation 4.3.

$$energy = \sum_{i}^{N} \mathbf{X}(i)^2 \tag{4.3}$$

Let V represent the volume of the tumour, while A is the surface area of the tumour, shape compactness is given by Equation 4.4.

$$shape = 36\pi \frac{V^2}{A^3} \tag{4.4}$$

Let $p(i, j|\theta)$ be the (i, j)th entry in the given run-length matrix p for a direction θ. Let

N_g the number of discrete intensity values in the image and N_r be the number of different run lengths. The run length non-uniformity (RLN) is given by Equation 4.5.

$$RLN = \frac{\sum_{j=1}^{N_r} \left[\sum_{i=1}^{N_g} p(i,j|\theta)\right]^2}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r} p(i,j|\theta)} \tag{4.5}$$

The second study inspected the 52 texture features described in Aerts et al. [10] and Vallières et al. [11], these features look at the spatial distribution of voxel intensities using gray level run-length and gray level co-occurrence matrices. In Aerts et al. [10] four features were selected from 440, in this study we explore the prognostic significance of the three features described above in equations 4.3, 4.4 and 4.5, excluding the wavelet feature.

Table 4.1: This table provides the AUC results from 20 iterations for all texture features compared to the three feature analysis, in addition to the standard deviation.

Volume	Three Features	Texture Features
	AUC$\pm\sigma$	AUC$\pm\sigma$
Whole volume	0.589$\pm$0.015	0.661$\pm$0.016
Inner rind	0.558$\pm$0.011	0.679$\pm$ 0.024
Outer rind	**0.598**$\pm$0.011	0.689$\pm$ 0.015
Outer rind with volume	0.583$\pm$0.013	**0.699**$\pm$ 0.011
Volume excluding inner rind	0.584$\pm$0.013	0.624$\pm$ 0.015

The data, consisted of DICOM-RT CT image files along with two-year survival outcome for 245 NSCLC patients. A 10-fold cross-validation approach was employed to evaluate the models. Each fold of the cross-validation was stratified based on the propensity of outcomes in the total cohort. Logistic regression was then employed to produce the two-year survival Kaplan-Meier curves, as this is very similar to the multi-variant Cox proportional hazards regression model used by Aerts et al. [10]. These curves are often used in medical research to predict survival by stratifying survival likelihood into two different patient groups [10], [13], [71]. The data was then separated into a good and bad prognosis group separated by the median values.

4.4 Results

The results can be seen in Figure 4.3. As a similar analysis we looked at the 52 texture based features, which showed an improvement in the two-year survival, as shown in Figure 4.4.

It was found that the outside rind could predict the survival better than the whole volume alone. As a result of this discovery, we included an evaluation of the outer rind combined with the whole volume in this study.

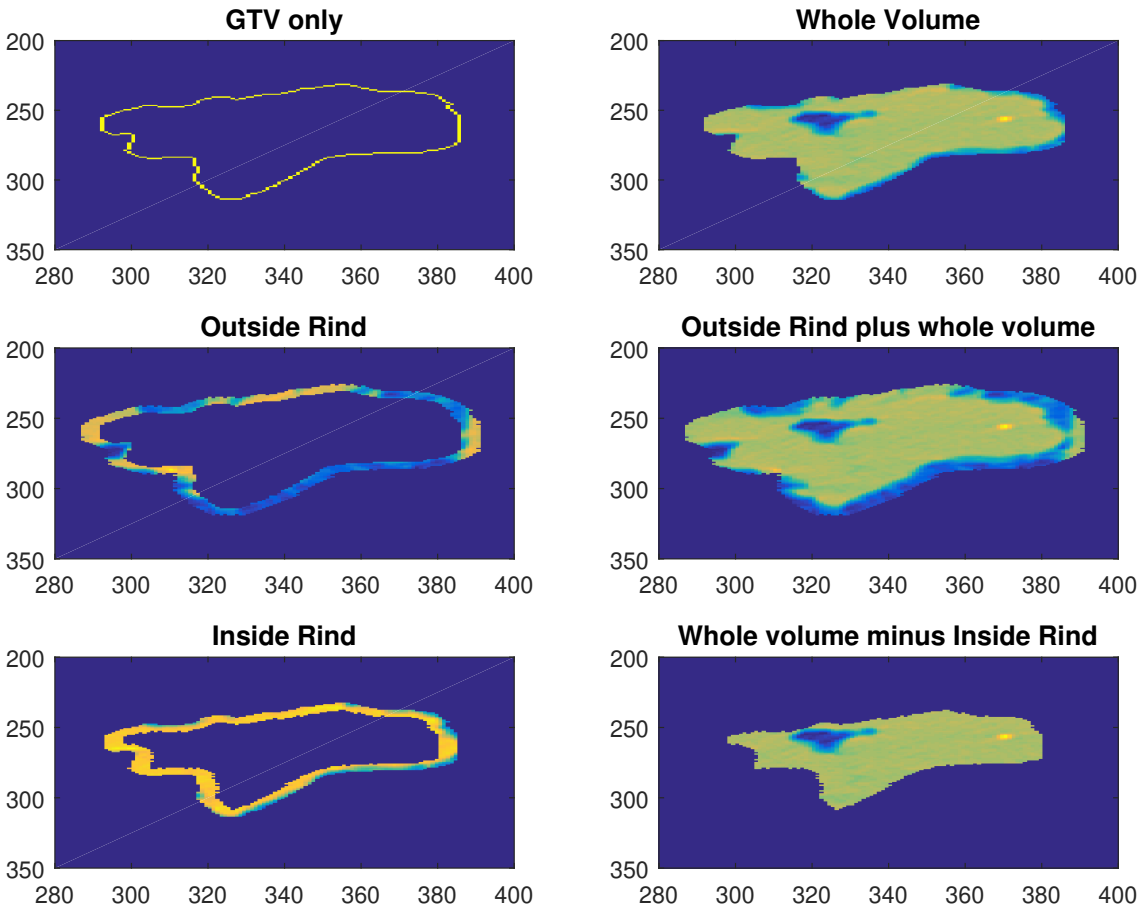

Figure 4.2: Example outside rind compared to inside rind and whole tumour volume for a single CT slice.

Analysing the results for all texture features, we found the outside rind plus the whole volume to produce the best prediction for two-year survival, this was closely followed by the inside rind and the whole volume. The whole tumour volume without the inside rind produced the worst results, when comparing the area under the curve (AUC) values after twenty iterations of the cross-validation, shown in Table 4.1.

Overall results of all textures features was approximately 10% better at predicting survival than the three feature analysis alone. The Kaplan Meier curves in Figure 4.4 also demonstrate that the models based on all texture features improve in discriminating survivors from non-survivors. The bottom blue curve represents the poor prognosis group and the top red curve represents the good prognosis group. Increased separation between these curves indicates greater success in two-year survival classification. This is clearly demonstrated in the case of all texture features being tested however in the case of the three textures only, the difference between these two curves is less significant, see Figure 4.3.

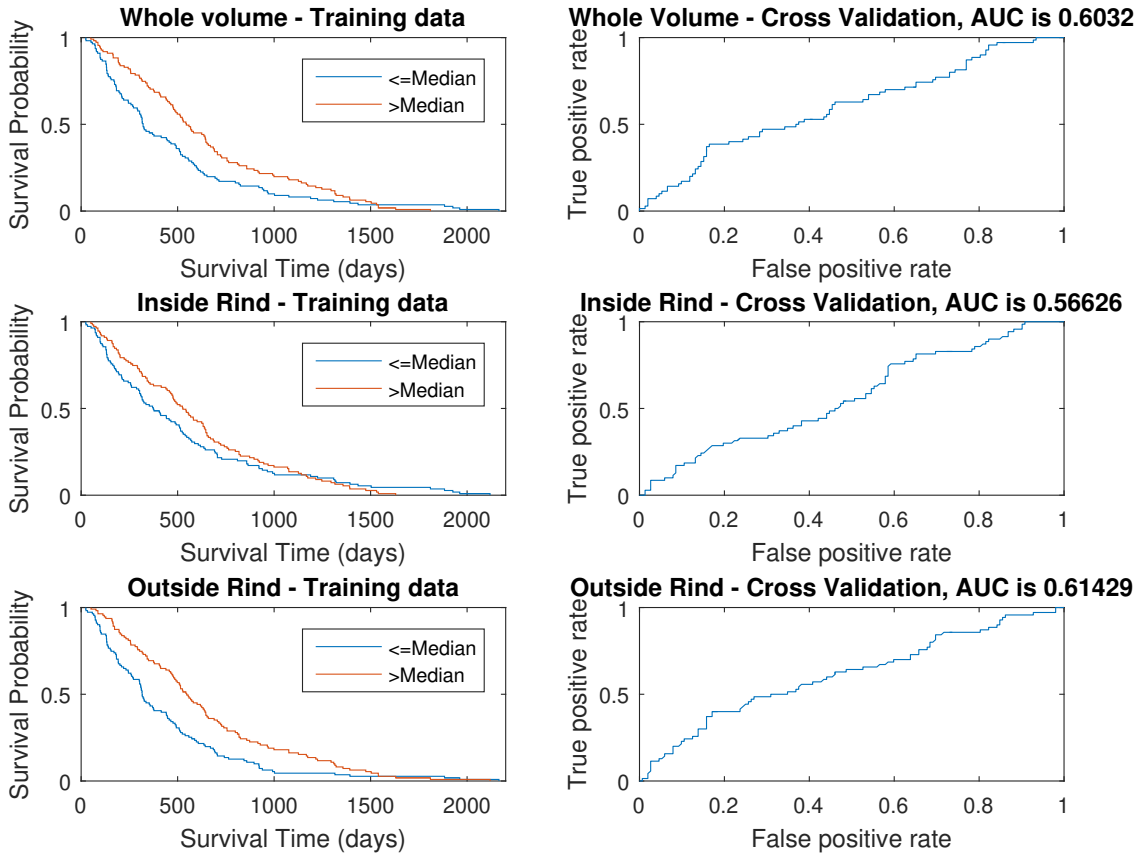

Figure 4.3: The diagrams on the left show the logistic regression of energy, shape and grey level non-uniformity radiomic signatures, while the models on the right show the ROC for classification of the whole volume, inside rind and outside rind for the training data compared to the validation data.

4.5 Discussion

This work presents analytic approaches to delineate a tumour and describes features that produce the best two-year survival prediction. The clinical perspective is that the most important region of the tumour to target is the centre of the tumour, however these results found the whole tumour without the rind (just the centre of the tumour) has the worst performance. In clinical practice, a volume outside the GTV called the CTV is added to try to include cancer infiltrating into surrounding lung that cannot be appreciated by the naked eye.

This work supports this clinical practice by showing that important textural information that relates to two-year survival is not necessarily within the tumour but proximally around it. The likely mechanism for this improvement in predictive performance arises from the ability of radiomic analysis to find pertinent signal information within the image that cannot be appreciated by the eyes of an oncologist, thereby determining the extent of the tumour more accurately. The post hoc analysis of treatment images will identify visually imperceptible instances of 'geographic miss', that defines cases where the radiation did not fully cover the tumour, which obviously will result in worse survival rates.

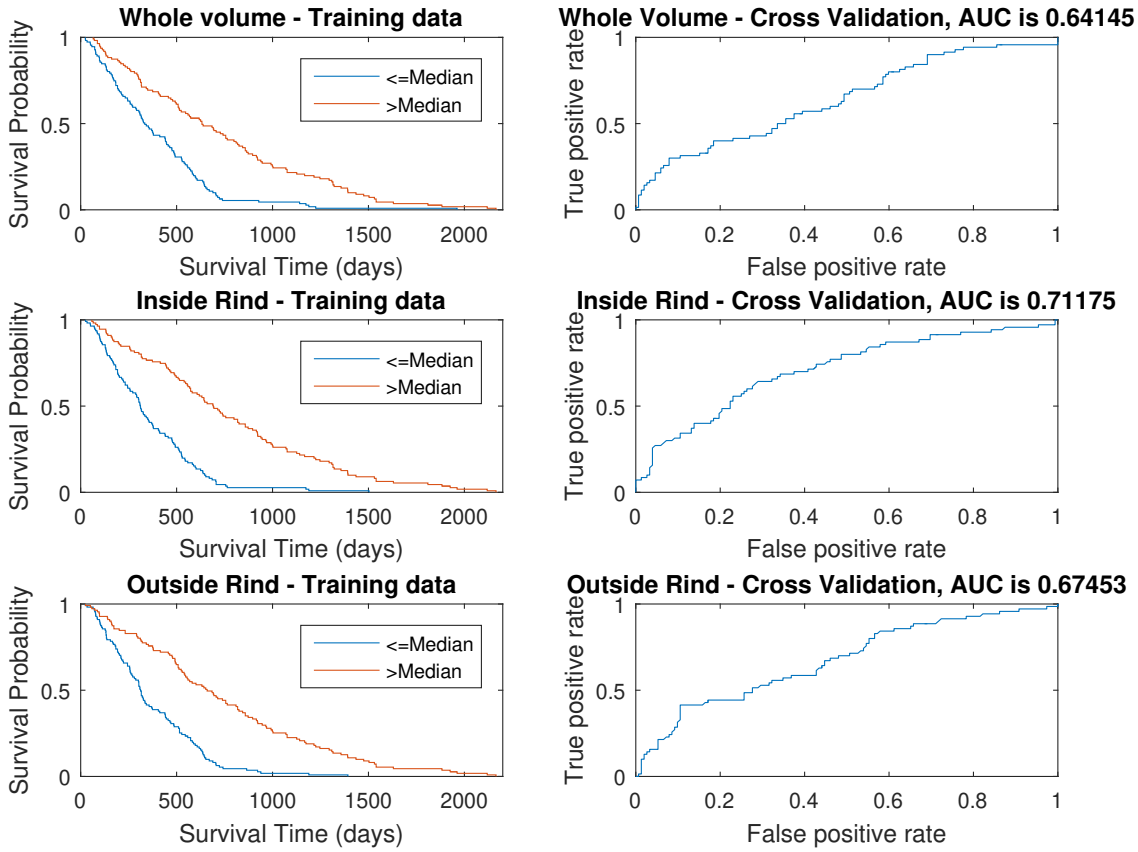

Figure 4.4: The graphs on the left show the Kaplan Meier survival curves as a result of logistic regression on all 52 texture based radiomic features, while the models on the right show the ROC for classification of the whole volume, inside rind and outside rind for the training data compared to the validation data.

4.6 Conclusion and Future Work

In this chapter we explored the textural radiomic features in a whole 3D tumour volume, compared to the inside and outside rind of the tumour only, for CT images of NSCLC. The textural features found within the tumour rinds were very similar to the textures in the entire tumour volume, and this analysis predicted two-year survival with an improved accuracy of 3% for survival classification using textures from the outside rind compared to the whole volume. It is important to note, that there is significant clinical uncertainty in defining the tumour boundaries of the GTV and that this should be considered when viewing these results.

One limitation of this work is only one machine learning technique is used to determine the best radiomic features themselves, in the next chapter we explore additional machine learning techniques to determine the best method for radiomic feature selection and survival modelling.

Chapter 5

Machine Learning Comparison Of Tumour Rind Survival Models

5.1 Introduction

In this chapter, we continue to explore the textural radiomic features found in the rind of NSCLC tumours defined by an oncologist. In this study the rind is gradually expanded pixel by pixel to determine at which point the maximum survival can be determined. The machine learning techniques used to determine this have also been extended in the chapter to determine which techniques provide the best 2 year survival prediction.

This approach was first proposed in [3] to compare the effects of exploring the inside and the outside rind of a tumour volume to verify if additional predictive information could be found here. Preliminary experimental results found that exploring the outside rind along with the whole volume produced the best predictive results. In this study the authors found that the exploration of the inside volume produced worse predictive results. Hence this chapter is only focusing on an exploration of adding the outside rind to the GTV one pixel at a time rather than the inside rind. This approach is further supported by the research conducted by Hao et al. [75] who developed a new shell radiomic feature that was found to be highly predictive of cancer metastasis. They also proposed that future work could explore the region around the tumour to improve cancer metastasis prediction. Cancer metastasis has also been strongly correlated with poor rates of cancer survival [11], [75]. To our knowledge this was the first study to compare multiple machine learning (ML) techniques using radiomics while also exploring the rind of the tumour volume rather than only the GTV.

The effect is explored for predicting a two-year survival by including only imaging textures discussed in Aerts et al. [10]. By utilising the same public NSCLC dataset that was utilised in [10] to predict the two-year survival. The rationale for observing the rind region is to establish a strategy for partitioning and analysing the texture of tumour regions

so as to characterise the relationship between texture and survival in NSCLC patients. Clinically the area around the tumour is irradiated with a curative level of radiation at the centre of the tumour and with lower intensity residual radiation around the outside of the tumour. This work explores the activity of the tumour at its periphery compared to the original volume.

This research is presented as follows. Section 5.2 outlines the relevant related research. Section 5.3 will provide the methodology adopted in this study. Section 4.4 will describe the results of this study. Section 5.5 will present the discussion of this work and Section 5.6 summarises the conclusions and opportunities for future work.

5.2 Existing Radiomics Approaches

Unsupervised learning techniques, such as clustering or self organising maps (SOM), have better performance when very large datasets are provided. Larger cohorts of patients will likely improve the credibility of this applied research to reliably predict 2-year survival in NSCLC patients. Many studies have chosen to use support vector machines and logistic regression [10], [15], [16], which are limited to providing only two classification options. This chapter differs from the papers above as it explores the region outside the GTV in addition to inside the GTV.

Hao et al. [75] introduces a new radiomics descriptor known as the shell feature which is useful in predicting distance metastasis. They also indicate the need to explore the area around the tumour volume as this area may contain useful information that can aid in predicting distance metastasis which can in turn predict survival as these two outcomes appear to be linked in the literature. [11], [36], [75].

An earlier report demonstrated that four radiomics features could predict the two-year survival of lung cancer patients in a public NSCLC dataset with a high degree of accuracy, in Aerts et al. [10]. In the previous chapter, we conducted a similar analysis using three of these identified highly prognostic features to understand whether the presence of these radiomic features in the inner and outer rinds around the GTV-line can also predict two-year survival [3]. The previous study did not include the wavelet features which were included in this study for a full comparison.

These related works reveal the potential for radiomics to improve the accuracy of outcome predictions. Until now, reported analyses have focused on the tumour (GTV) features alone. We have begun to analyse the lung exterior and surrounding the tumour also. This approach is supported by the approach of radiation oncologists who encapsulate the visible GTV within a CTV which defines the at risk areas which appear to be normal to the naked eye.

5.3 Methodology

This chapter compares support vector machines, logistic regression, naive bayes, J48 decision trees, lazy IBk nearest neighbour, bayesian networks, random forest, locally weighted learning and AdaBoost for rinds 0-10 pixels outside the GTV. These nine machine learning techniques were chosen as they have achieved prior success in this field [15], [16].

At the time of publication, 422 NSCLC patient data sets were publicly available, 312 were usable in this study, approximately 19.8GB from over $37,000$ CT image files were examined. Only 312 were usable in this study as tumour delineations were not available for the other patients, the files were obtained from [10], [68], [69]. The initial investigation included 250 NSCLC patients classified using logistic regression [3], this study has since been expanded to 312 NSCLC patients classified using multiple machine learning techniques.

In addition to this, the original study only considered 76 radiomic features, including the shape, texture and statistical radiomic features. This study has since been expanded to include the 536 wavelet features, allowing for 612 radiomic features in total.

The initial results of this study have introduced some new research questions, the main question is "How far can the tumour rind be expanded before the prediction becomes inaccurate?" To address this the entire original dataset was altered to ensure that all the voxel resolutions and widths were equivalent via interpolation, namely, $0.977 * 0.977 * 0.300mm$, while also ensuring that the image sizes were exactly the same, $471 * 525 * 99pixels$, to be precise.

The rinds were then created at an expansion of 0 to 10 pixels in increments of 1 pixel. Since the initial results suggested that the rind plus the original tumour volume generally yielded the best results, the rind mask was added to the original tumour mask to create the new mask for each dataset.

Upon combining this new mask with the original CT Data the radiomic features were subsequently computed for each patient. The results were then analysed using the above machine learning techniques to compare the results and determine the best rind width and technique. For ease of comparison WEKA [76], [77] was employed to test these machine learning algorithms using 10 fold cross validation.

Subsequently, nine machine learning techniques were employed to determine the radiomic features which provided the best prognostic value and these features were consequently utilised to predict the 2-year survival for comparison.

This analysis consists of four main stages; converting the DICOM-RT files into matrices for input to MATLAB, calculating the rind volume for extraction from the CT data, determining the radiomic features and finally producing the two-year survival analysis.

As previously mentioned, the DICOM-RT format is the standard medical imaging for-

mat used to store medical imaging data and is paramount in the efficacious application of the picture archiving and communications systems (PACS) used in most hospitals for radiation therapy [74]. It is essentially a file format that allows multiple CT images to be related using an overarching RTSTRUCT file which describes the oncology delineation data.

The method for deriving the mask is found in the previous chapter defined by equations 4.1, 4.2 and 4.3.

This rind was then dilated and contracted to ensure there were no pixels missing within the rind, creating a rind exactly $2 \times n$, being 6 pixels deep in this study, all the way around the 3D GTV, this is an expansion of approximately 6mm around the whole GTV volume. This can be seen in Figure 4.2.

Subsequently, we extracted the four features recognised by the authors in [10] as having prognostic significance and compared the two-year survival predictions for the whole volume compared to the inside rind and outside rind alone. Three of these features included the statistical energy, shape compactness and the run length non-uniformity of the image, these features are used for comparison in this study and were described mathematically in the previous chapter respectively by equations 4.3, 4.4, 4.5.

The fourth feature, wavelet Grey Level Non-uniformity HLH feature is the high-pass sub band of the Grey Level Non-uniformity feature above with a high pass filter in the x-direction, a low pass filter along the y-direction and a high pass filter along the z direction, described as follows.

$$RN_{HLH}(i,j,k) = \sum_{p=1}^{N_H} \sum_{q=1}^{N_L} \sum_{r=1}^{N_H} H(p)L(q)H(r)RN(i+p, j+q, k+r) \tag{5.1}$$

This feature is useful as it is applying different low and high pass filters to better observe the textural differences in the images. The second study inspected the 52 texture features described in Aerts et al. [10] and Vallières et al. [11], these features look at the spatial distribution of voxel intensities using gray level run-length and gray level co-occurrence matrices. In Aerts et al. [10] four features were selected from 440, in this study we explore the prognostic significance of the four features described above in equations 4.3,4.4, 4.5 and 5.1.

The data, consisted of DICOM-RT CT image files along with two-year survival outcome for 312 NSCLC patients. A 10-fold cross-validation approach was employed to evaluate the models. Each fold of the cross-validation was stratified based on the propensity of outcomes in the total cohort. Logistic regression was then employed to produce the two-year survival Kaplan-Meier curves, as this is very similar to the multivariant Cox proportional hazards regression model used by Aerts et al. [10]. These curves are often used in medical research to predict survival by stratifying survival likelihood into two different patient groups [10], [13], [71]. The data was then separated into a good and bad

prognosis group separated by the median values.

Once the rinds were determined, these were combined with the whole GTV tumour volume to extract the radiomic texture features. 53 radiomic texture features were included in this analysis, the texture features were selected as they were found to have the best prognostic value as found in the previous study by Vial et al. [3].

The data was analysed in WEKA [76], [77] to determine the accuracy of each machine learning technique. A 10-fold cross validation was applied to the dataset with each experiment conducted 10 times to ensure repeat ability and accuracy of the results.

5.4 Results

The results indicate that Support Vector Machine (SVM) achieved the highest accuracy for all rind widths, and the highest overall result at a distance of 6 pixels. Each pixel has been interpolated to the same scale of $0.977mm$ hence 6 pixels is a distance of $5.862mm$ outside the original GTV. It is important to note that although the best performance is from the SVM at 6 pixels, shown in Table 5.2 , the standard deviations for all the methods and pixel depths are quite large, so there is still some uncertainty in the results and further research is needed in future studies.

Table 5.1: This table shows the top 10 Significant features discovered by the Decision Tree and Support Vector Machine algorithms using voting to check for frequency and significance of the features.

Gray Level Co-occurrence Matrix (GLCM) Cluster Shade (cshad)
Gray Level Co-occurrence Matrix (GLCM) Correlation Matrix (corrm)
Low Low Low (LLL) Mean
Low Low Low (LLL) Maximum
High High High (HHH) Kurtosis
High Low High (HLH) Gray Level Run Length Matrix (GLRLM) (Large Run High Gray-Level Emphasis) LRHGE
Low Low High (LLH) Gray Level Size Zone Matrix (GLSZM) Large Zone High Gray-Level Emphasis (LZHGE)
High Low Low (HLL) Gray Level Run Length Matrix (GLRLM) Gray Level Non-Uniformity (GLN)
Gray Level Co-occurrence Matrix (GLCM) Energy
Low High Low (LHL) Gray Level Run Length Matrix (GLRLM) Gray Level Variance (GLV)

A Gray Level Co-occurrence Matrix (GLCM) describes the probability density function of a region of neighbouring voxels across 13 angles, this was first detemine by Haralick et al. [41] as is used as the basis for all first order features. A Gray Level Run Length Matrix (GLRM) quantifies the length in number of pixels in a line within the image that contain the same gray level intensity. A Gray Level Size Zone Matrix (GSZM) similarly

to a GRLM quantifies the number of connected grey levels of the same intensity found in the same region.

The cluster shade feature determines the uniformity and skewness for the GLCM. The correlation matrix feature measures the similarity of voxels within the GLCM. The mean feature describes the average grey level value with the region of interest. The maximum feature refers to the maximum grey level value found within the ROI. Kurtosis is a measure of how close the peak of the distribution is to the mean value. Gray Level Non-Uniformity measures the similarity of the voxels within the image and is a good measure of tumour homogeneity.

Wavelet features are distinguished as the features that lead with the description HHH, LLL, HLH, HLL or LLH. These wavelet features are obtained using different combinations of High and Low pass filtering along the three axes x, y and z respectively e.g. HHH, LLL, HLH, HLL or LLH. It was found that many wavelet features showed significance for survival discovery as they make up a large proportion of the significant features, these wavelet features are namely, LLL-mean, LLL-maximum, HHH-Kurtosis, HLH-GLRLM-LRHGE, LLH-GLSZM-LZHGE, HLL-GLRLM-GLN and LHL-GLRLM-GLV.

The top 10 Significant features discovered by the Decision Tree and Support Vector Machine algorithms using voting to check for frequency and significance of the features were GLCM-cshad, GLCM-corm, LLL-mean, LLL-maximum, HHH-Kurtosis, HLH-GLRLM-LRHGE, LLH-GLSZM-LZHGE, HLL-GLRLM-GLN, GLCM-energy and LHL-GLRLM-GLV. It can be seen here that many wavelet features showed significance for survival discovery as they make up a large proportion of the significant features, these wavelet features are namely, LLL-mean, LLL-maximum, HHH-Kurtosis, HLH-GLRLM-LRHGE, LLH-GLSZM-LZHGE, HLL-GLRLM-GLN and LHL-GLRLM-GLV, as shown in Table 5.1

5.5 Discussion

This work presents analytic approaches to delineate a tumour and describes features that produce the best two-year survival prediction. The clinical perspective is that the most important region of the tumour to target is the centre of the tumour, however these results found the whole tumour without the rind (just the centre of the tumour) has the worst performance. In clinical practice, a volume outside the GTV called the CTV is added to try to include cancer infiltrating into surrounding lung that cannot be appreciated by the naked eye.

This work supports clinical practice by showing that important textural information that relates to two-year survival is not necessarily within the tumour but around it. The likely mechanism for this improvement in predictive performance arises from the ability of radiomic analysis to find pertinent signals within the image that cannot be appreciated by

the eyes of an oncologist, thereby determining the extent of the tumour more accurately. The post hoc analysis of treatment images will identify visually imperceptible instances of 'geographic miss' which will result in worse survival.

The improved prediction obtained from 6 pixels outside the original GTV is inline with clinical treatment methods. The GTV, is defined as the region of disease that can be seen, the CTV, is defined as the volume clinicians believe microscopic disease exists, this has a volume to the rind in this study. The edge of the CTV is likely biologically to be the region where cells escape from to generate metastasis. As overall survival is the end point for this study, and increased likelihood of metastasis is linked with poor survival [10], [11], it is intuitive that including the rind on the GTV in the analysis produces improved outcome results. We have established here mathematically, that the CTV is more prognostic than the GTV, for survival prediction.

One of the limitations of this study is the use of just one metric: accuracy. More thorough future investigations will analyse other metrics such as precision, sensitivity, specificity, which are complementary to accuracy as they quantify specific aspects of the performance. Another limitation is the lack of Kaplan-Meier curves and other detailed analysis of the survival results, which will be provided in future investigations.

5.6 Conclusion and Future Work

This chapter aimed to characterise the 2-year survival rate of non-small cell lung cancer patients, through a novel approach that explores the rind around the tumour volume as delineated by an oncologist. This study also compared 9 different machine learning techniques including Random Forests, Support Vector Machines, and Logistic Regression to determine the ideal method for predicting 2-year cancer survival. The rind was also expanded linearly from 1 to 10 pixels and it found improved prediction results at 6 pixels outside the tumour volume, a distance of approximately 5mm outside the original GTV, when applying a support vector machine an accuracy of 71.18% was achieved. This research indicates that the characteristics of the periphery of the tumour are important to include in prognostic tools to stratify patients for survival prediction. Many previous studies only use the GTV region for prognosis [10], [15], [16]. Future work, would involve exploring additional machine learning techniques to see if improved results may be achieved. The rind may also be expanded to up to 20 pixels as GTV delineations by clinicians can vary by up to 2cm [78], [79].

Table 5.2: This table provides the results from 10 iterations of each machine learning technique for all 0-10 pixel rinds with GTV

Volume	SVM	LR	Naive Bayes	J48	kNN	Bayes Net	RF	LWL	AdaBoost
	AUC±σ	AUC±σ	AUC±σ	AUC±σ	AUC±σ	AUC±σ	AUC±σ	AUC±σ	AUC±σ
Whole volume	68.2±7.0	**66.0**±10.2	64.0±11.3	64.0±11.3	66.3±10.1	57.2±11.6	69.4±11.3	66.3±7.9	68.4±10.0
1 pixel	70.9±7.9	64.6±9.8	**66.9**±11.1	66.9±9.1	**64.9**±11.4	**69.4**±10.4	69.1±8.0	**69.4**±10.4	69.9±10.0
2 pixels	69.3±8.1	64.5±10.3	64.9±11.4	67.2±8.8	62.2±11.2	67.8±11.2	69.5±7.6	67.9±9.6	**70.5**±10.3
3 pixels	68.9±7.9)	63.5±9.7	63.3±11.3	69.8±9.1	58.2±11.2	68.6±11.6	69.53±7.73	69.6±10.00	70.0±10.1
4 pixels	70.1±7.8	64.6±9.8	60.3±11.9	68.6±8.7	53.5±11.5	68.7±11.8	70.4±7.7	69.9±10.4	70.4±10.5
5 pixels	69.9±8.1	64.3±10.6	62.2±12.0	68.3±8.7	56.9±11.1	66.6±11.1	70.2±7.9	67.7±10.4	69.2±10.4
6 pixels	**71.9**±7.8	63.7±9.8	59.2±12.2	67.06±9.28	60.7±11.0	66.4±11.3	**71.5**±7.5	65.8±10.2	66.5±10.2
7 pixels	69.6±7.4	64.6±10.1	60.3±11.5	66.6±8.4	60.5±11.7	65.3±11.4	70.8±7.6	66.6±9.9	66.8±9.7
8 pixels	69.8±7.1	63.7±10.4	62.0±11.8	67.5±8.3	58.3±12.0	65.2±11.8	71.0±7.7	64.7±9.4	66.4±9.7
9 pixels	69.0±7.3	62.2±10.3	62.1±11.9	67.1±8.6	62.5±10.5	63.7±11.3	63.7±7.5	65.6±8.7	65.7±9.6
10 pixels	68.7±6.9	62.6±9.9	62.6±12.0	65.8±10.0	63.3±11.5	64.8±10.4	70.0±7.8	66.2±8.0	65.8±9.7

Chapter 6

Conclusion

6.1 Introduction

Radiomics is a field that is advancing rapidly with a vast potential for increasing the ability to diagnose and treat cancer effectively. In order to be successful in this field, significant expertise is required in both biology and computer science. This will enable computer vision and machine learning techniques to be applied effectively which has huge potential in solving the issue of cancer analysis in medical imaging.

6.2 Deep Learning

In Chapter 3, we explored the textural radiomic features in a whole 3D tumour volume, compared to the inside and outside rind of the tumour only, for CT images of NSCLC. The textural features found within the tumour rinds were very similar to the textures in the entire tumour volume, and this analysis predicted two-year survival with an improved accuracy of 3% for survival classification using textures from the outside rind compared to the whole volume. It is important to note, that there is significant clinical uncertainty in defining the tumour boundaries of the GTV and that this should be considered when viewing these results.

CNNs have great potential to produce improved results in survival prediction, through extracting new radiomic features from whole tumour volumes. This research shows that 2D CNNs require the tumour volume to be first normalised by segmenting out and centering the volume using the 3D GTV for each patient in order to be effective. Alternatively, 3D approaches such as radiomics and 3D CNN could also be utilised in the future for improved results. This study achieved an accuracy 56% for 2D CNNs on a large unseen test data set. As more public datasets become made available CNNs are likely to have greater success, in this area but for now there is not enough data available for CNNs to have success in this field, like they have in other areas.

Due to the challenges in adapting CNN for 3D survival prediction, it was suggested that a simpler approach should be investigated so we looked into traditional machine learning techniques for this task.

6.3 Rind Analysis

In Chapter 4, we explored the textural radiomic features in a whole 3D tumour volume, compared to the inside and outside rind of the tumour only, for CT images of NSCLC. It was determined that the textural features found within the tumour rinds were very similar to the textures in the entire tumour volume and hence this analysis could be used to predict two-year survival with an improved accuracy of 3% for survival classification using textures from the outside rind compared to the whole volume. It is important to note, that there is a plethora of clinical uncertainty in defining the tumour boundaries or GTV and that this should be considered when viewing these results.

As these results proved interesting we decided to expand on this study in the Chapter 5, where additional machine learning techniques were explored and the rinds were expanded linearly out to 10 pixels rather than granularly as in Chapter 4.

6.4 Rind Analysis and Machine Learning

In Chapter 5, we explored the textural radiomic features in the whole 3D tumour volume, compared to the ten rinds gradually drawn 1 pixel outside the GTV. This chapter aimed to characterise the 2-year survival rate of non-small cell lung cancer patients, through a novel approach that explored the rind around the tumour volume as delineated by an oncologist. This study also compared 9 different machine learning techniques to determine the ideal method for predicting 2-year cancer survival. SVMs proved to be the best overall machine learning technique of the 9 compared. This chapter found improved prediction results at 6 pixels outside the tumour volume, a distance of approximately 5mm outside the original GTV, when applying a support vector machine an accuracy of 71.18% was achieved. This research indicates that the characteristics of the periphery of the tumour are important to include in prognostic tools to stratify patients for survival prediction. Many previous studies only use the GTV region for prognosis [10], [15], [16]. Future work, would involve exploring additional machine learning techniques to see if improved results may be achieved. The rind may also be expanded to up to 20 pixels as GTV delineations by clinicians can vary by up to 2cm [78], [79].

It was also found that several textural based wavelets features were used by the machine learning algorithms to determine survival. Previous studies have also found these wavelet features to be highly suggestive of survival [10].

This thesis challenges the traditional clinical ideas of radiotherapy where the centre of the tumour is treated with the highest dose, however this research indicates that the characteristics of the periphery of the tumour are important to include in prognostic tools to stratify patients for survival prediction.

6.5 Future Work

In future work, an exploration of creating pixel based Radiomic Features and determining if this shows more information. Pixel based Radiomic Features are where we generate features based on individual sections of the tumour rather than the whole volume. This provided a far more granular view of the turmour volume and made visualising the tumour differently possible, which could be useful for making tumours easier to visualise for oncologists. We began exploring this approach in an additional study [80] and found some promising results. In this study we were able to use the pixel based features to recreate the original images and to show hidden textures, that were not visible to the naked eye. [80] In future studies these features would be compared to survival analysis results to see if they could be used for survival prediction.

6.6 Summary

This work has suggested that the periphery of the tumour volume is more prognostic of cancer survival likelihood than the interior textures. This may be as increased heterogeneity of the tumour exterior may lead to increased likelihood of cancer metastasis. These results could also lead to new clinical thinking on the tumour rind rather than just the GTV. This may also suggest that the clinical approach to focus the majority of the radiation on the centre of the tumour should be reconsidered and the periphery should be targeted instead if possible.

Bibliography

[1] P. Lambin et al., "Radiomics: Extracting more information from medical images using advanced feature analysis," *European Journal of Cancer*, vol. 48, no. 4, pp. 441–446, Mar. 2012.

[2] A. Vial, D. Stirling, M. Field, *et al.*, "The role of deep learning and radiomic feature extraction in cancer-specific predictive modelling: A review," *Translational Cancer Research*, vol. 7, no. 3, 2018.

[3] A. Vial, D. Stirling, M. Field, *et al.*, "Assessing the Prognostic Impact of 3D CT Image Tumour Rind Texture Features on Lung Cancer Survival Modelling," in *5th IEEE Global Conference on Signal and Information Processing*, 2017.

[4] S. H. Hawkins et al., "Predicting Outcomes of Non-small Cell Lung Cancer Using CT Image Features," *IEEE Access*, vol. 2, pp. 1418–1426, 2014.

[5] E. R. Velazquez et al., "A semiautomatic CT-based ensemble segmentation of lung tumors: Comparison with oncologists' delineations and with the surgical specimen," *Radiotherapy and Oncology*, vol. 105, no. 2, pp. 167–173, Nov. 2012.

[6] M. Pritchett, S. Schampaert, C. S. Joris de Groot, and I. van der Bom, "Cone-beam ct with augmented fluoroscopy combined with electromagnetic navigation bronchoscopy for biopsy of pulmonary nodules," *Journal of Bronchology Interventional Pulmonology*, vol. 25, no. 4, pp. 274–282, 2018. DOI: 10.1097/LBR.0000000000000536.

[7] A. Schaefer et al., "PET-based delineation of tumour volumes in lung cancer: Comparison with pathological findings," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 40, no. 8, pp. 1233–1244, 2013.

[8] E. Velazquez et al., "Volumetric CT-based segmentation of NSCLC using 3D-Slicer," *Scientific Reports*, vol. 3, 2013.

[9] V. Kumar et al., "Radiomics: The process and the challenges," *Magnetic Resonance Imaging*, vol. 30, no. 9, pp. 1234–1248, 2012.

[10] H. J. W. L. Aerts et al., "Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach," *Nature Communications*, vol. 5, 2014.

[11] M. Vallières et al., "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, p. 5471, 2015.

[12] B. Ganeshan et al., "Tumour heterogeneity in oesophageal cancer assessed by CT texture analysis: Preliminary evidence of an association with tumour metabolism, stage, and survival," *Clinical Radiology*, vol. 67, no. 2, pp. 157–164, Feb. 2012.

[13] A. R. Subramaniam et al., "A novel metric for quantification of homogeneous and heterogeneous tumors in PET for enhanced clinical outcome prediction," *Physics in Medicine and Biology*, vol. 61, no. 1, p. 227, 2016.

[14] A. Dekker et al., "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, Oct. 2014.

[15] C. Parmar, P. Grossmann, J. Bussink, P. Lambin, and H. J. Aerts, "Machine Learning methods for Quantitative Radiomic Biomarkers," *Scientific Reports*, vol. 5, p. 13 087, Aug. 2015.

[16] C. Parmar et al., "Radiomic Machine-Learning Classifiers for Prognostic Biomarkers of Head and Neck Cancer," *Frontiers in Oncology*, vol. 5, 2015.

[17] A. O. de Carvalho Filho et al., "Lung-Nodule Classification Based on Computed Tomography Using Taxonomic Diversity Indexes and an SVM," *Journal of Signal Processing Systems*, pp. 1–18, 2016.

[18] Y. LeCun et al., "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.

[19] P. Ypsilantis et al., "Predicting Response to Neoadjuvant Chemotherapy with PET Imaging Using Convolutional Neural Networks," *PLOS ONE*, vol. 10, no. 9, Sep. 2015.

[20] Y. Sun, H. Reynolds, D. Wraith, *et al.*, "Predicting prostate tumour location from multiparametric MRI using Gaussian kernel support vector machines: a preliminary study," *Australasian Physical & Engineering Sciences in Medicine*, vol. 40, no. 1, pp. 39–49, 2017.

[21] C. Dehing-Oberije et al., "Development and External Validation of Prognostic Model for 2-Year Survival of Non-Small Cell Lung Cancer Patients Treated With Chemoradiotherapy," *International Journal of Radiation Oncology*Biology*Physics*, vol. 74, no. 2, pp. 355–362, Jun. 2009.

[22] I. E. N. Deasy et al., "Predicting radiotherapy outcomes using statistical learning techniques," *Physics in Medicine & Biology*, vol. 54, no. 18, S9, 2009.

[23] H. Lee et al., "Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations," in *Proceedings of the 26th Annual International Conference on Machine Learning SE - ICML '09*, ACM, 2009, pp. 609–616.

[24] Z. Wu et al., "3D ShapeNets: A Deep Representation for Volumetric Shapes," Jun. 2014. eprint: 1406.5670.

[25] G. E. Hinton, S. Osindero, and Y. Teh, "A fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, May 2006.

[26] S. Ji, W. Xu, M. Yang, and K. Yu, "3D Convolutional Neural Networks for Human Action Recognition," vol. 35, no. 1, pp. 221–231, 2013.

[27] J. Donahue, L. A. Hendricks, M. Rohrbach, *et al.*, "Long-term Recurrent Convolutional Networks for Visual Recognition and Description," no. 99, p. 1, 2016. DOI: 10.1109/TPAMI.2016.2599174.

[28] W. Ouyang, X. Zeng, X. Wang, *et al.*, "DeepID-Net: Deformable Deep Convolutional Neural Networks for Object Detection," no. 99, p. 1, 2016. DOI: 10.1109/TPAMI.2016.2587642.

[29] X. Huang, J. Shan, and V. Vaidya, *Lung nodule detection in CT using 3D convolutional neural networks*, 2017.

[30] J. Sánchez et al., "Image classification with the fisher vector: Theory and practice," *International Journal of Computer Vision*, vol. 105, no. 3, pp. 222–245, 2013.

[31] P. Stanitsas, A. Cherian, X. Li, A. Truskinovsky, V. Morellas, and N. Papanikolopoulos, "Evaluation of feature descriptors for cancerous tissue recognition," in *2016 23rd International Conference on Pattern Recognition (ICPR)*, 2016, pp. 1490–1495.

[32] Z. Gao, L. Wang, L. Zhou, and J. Zhang, "HEp-2 Cell Image Classification With Deep Convolutional Neural Networks," vol. 21, no. 2, pp. 416–428, 2017.

[33] M. Muja and D. G. Lowe, "Scalable Nearest Neighbor Algorithms for High Dimensional Data," vol. 36, no. 11, pp. 2227–2240, 2014.

[34] H. T. H. Phan, A. Kumar, D. Feng, M. Fulham, and J. Kim, "Unsupervised two-path neural network for cell event detection and classification using spatiotemporal patterns," *IEEE Transactions on Medical Imaging*, vol. 38, no. 6, pp. 1477–1487, 2019. DOI: 10.1109/TMI.2018.2885572.

[35] R. T. H. Leijenaar et al., "The effect of SUV discretization in quantitative FDG-PET Radiomics: the need for standardized methodology in tumor texture analysis.," *Scientific reports*, vol. 5, no. August, 2015.

[36] T. P. Coroller et al., "CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma," *Radiotherapy and Oncology*, vol. 114, no. 3, pp. 345–350, 2015.

[37] Y. Balagurunathan et al., "Reproducibility and Prognosis of Quantitative Features Extracted from CT Images," *Translational Oncology*, vol. 7, no. 1, pp. 72–87, 2014.

[38] A. Cunliffe et al., "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, 2015.

[39] R. T. H. Leijenaar et al., "External validation of a prognostic CT-based radiomic signature in oropharyngeal squamous cell carcinoma," *Acta Oncologica*, vol. 54, no. 9, pp. 1423–1429, Oct. 2015.

[40] Y. T. Zhu et al., "Deciphering Genomic Underpinnings of Quantitative MRI-based Radiomic Phenotypes of Invasive Breast Carcinoma," *Scientific Reports*, vol. 5, 2015.

[41] 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.

[42] P. Scovanner, S. Ali, and M. Shah, "A 3-dimensional sift descriptor and its application to action recognition," in *Proceedings of the 15th ACM International Conference on Multimedia*, ser. MM '07, ACM, 2007, pp. 357–360.

[43] C. Paganelli, M. Peroni, F. Pennati, *et al.*, "Scale Invariant Feature Transform as feature tracking method in 4D imaging: A feasibility study," in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2012, pp. 6543–6546.

[44] F. Valente et al., "Variational bayesian gmm for speech recognition.," in *INTERSPEECH*, 2003.

[45] P. Lambin et al., "Decision support systems for personalized and participative radiation oncology," *Advanced Drug Delivery Reviews*, vol. 109, pp. 131–153, Jan. 2017.

[46] A. Jochems et al., "Developing and validating a survival prediction model for NSCLC patients through distributed learning across three countries," *International Journal of Radiation Oncology*Biology*Physics*, Jun. 2017. DOI: 10.1016/j.ijrobp.2017.04.021.

[47] T. M. Deist et al., "Infrastructure and distributed learning methodology for privacy-preserving multi-centric rapid learning health care: euroCAT," *Clinical and Translational Radiation Oncology*, vol. 4, pp. 24–31, Jun. 2017.

[48] T. Skripcak et al., "Creating a data exchange strategy for radiotherapy research: Towards federated databases and anonymised public datasets," *Radiotherapy and Oncology*, vol. 113, no. 3, pp. 303–309, 2014.

[49] K. M. Prise and J. M. O'Sullivan, "Radiation-induced bystander signalling in cancer therapy," *Nature reviews. Cancer*, vol. 9, no. 5, pp. 351–360, May 2009.

[50] D. Ravì, C. Wong, F. Deligianni, *et al.*, "Deep Learning for Health Informatics," vol. 21, no. 1, pp. 4–21, 2017.

[51] X. Zhen, Z. Wang, A. Islam, M. Bhaduri, I. Chan, and S. Li, "Multi-scale deep networks and regression forests for direct bi-ventricular volume estimation," *Medical Image Analysis*, vol. 30, pp. 120–129, 2016.

[52] T. Brosch and R. Tam, "Manifold Learning of Brain MRIs by Deep Learning BT - Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013: 16th International Conference, Nagoya, Japan, September 22-26, 2013, Proceedings, Part II," Springer Berlin Heidelberg, 2013, pp. 633–640.

[53] M. R. Avendi, A. Kheradvar, and H. Jafarkhani, "A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac MRI," *Medical Image Analysis*, vol. 30, pp. 108–119, 2016.

[54] J.-Z. Cheng, D. Ni, Y.-H. Chou, *et al.*, "Computer-Aided Diagnosis with Deep Learning Architecture: Applications to Breast Lesions in US Images and Pulmonary Nodules in CT Scans," vol. 6, p. 24 454, Apr. 2016.

[55] C. Berkeley. "Caffe." [Online]. Available: http://caffe.berkeley. (2017).

[56] Microsoft. "Cntk." [Online]. Available: https://github.com/Microsoft/CNTKy. (2017).

[57] Skymind. "Deeplearning4j." [Online]. Available: http://deeplearning4j.org/. (2017).

[58] Google. "Tensorflow." [Online]. Available: https://www.tensorflow.org. (2017).

[59] W. Research. "Wolfram math." [Online]. Available: https://www.wolfram.com/mathematica/. (2017).

[60] U. de Montreal. "Theano." [Online]. Available: http://deeplearning.net/software/theano//. (2017).

[61] K. K. R. Collobert and C. Farabet. "Torch." [Online]. Available: http://http://torch.ch/. (2017).

[62] F. Chollet. "Keras." [Online]. Available: https://keras.io/. (2017).

[63] N. Systems. "Neon." [Online]. Available: https://github.com/NervanaSystems/neon. (2017).

[64] Y. Shin and I. Balasingham, "Comparison of hand-craft feature based SVM and CNN based deep learning framework for automatic polyp classification," in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2017, pp. 3277–3280. DOI: 10.1109/EMBC.2017.8037556.

[65] S. Wang, M. Zhou, O. Gevaert, *et al.*, "A multi-view deep convolutional neural networks for lung nodule segmentation," in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2017, pp. 1752–1755. DOI: 10.1109/EMBC.2017.8037182.

[66] M. A. Al-masni, M. A. Al-antari, J. M. Park, *et al.*, "Detection and classification of the breast abnormalities in digital mammograms via regional Convolutional Neural Network," in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2017, pp. 1230–1233. DOI: 10.1109/EMBC.2017.8037053.

[67] C. Wang, J. Shi, Q. Zhang, and S. Ying, "Histopathological image classification with bilinear convolutional neural networks," in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2017, pp. 4050–4053. DOI: 10.1109/EMBC.2017.8037745.

[68] K. Clark et al., "The cancer imaging archive (TCIA): Maintaining and operating a public information repository," *Journal of Digital Imaging*, vol. 26, no. 6, 2013.

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

[70] C. McGibney et al., "Analysis of dose distribution in the 'Rind'-a volume outside the PTV-in 3-dimensional conformal radiation therapy of non-small cell lung cancer," *Radiotherapy and Oncology*, vol. 66, no. 1, pp. 87–93, May 2003.

[71] B. Ganeshan et al., "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.

[72] Rahmadwati and G. Naghdy et al., "Classification cervical cancer using histology images," in *2010 2nd International Conference on Computer Engineering and Applications, ICCEA 2010*, vol. 1, 2010, pp. 515–519.

[73] T. A. Alshawi, F. E. A. El-Samie, and S. A. Alshebeili, "Magnetic resonance and computed tomography image fusion using bidimensional empirical mode decomposition," in *2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Dec. 2015, pp. 413–417.

[74] M. Y. Y. Law and B. Liu, "DICOM-RT and Its Utilization in Radiation Therapy," *RadioGraphics*, vol. 29, no. 3, pp. 655–667, May 2009.

[75] H. Hao, Z. Zhou, S. Li, *et al.*, "Shell feature: a new radiomics descriptor for predicting distant failure after radiotherapy in non-small cell lung cancer and cervix cancer," Sep. 2017. [Online]. Available: `http://arxiv.org/abs/1709.09600`.

[76] M. A. H. Eibe Frank and I. H. Witten, "Data mining: Practical machine learning tools and techniques," 2016.

[77] T. C. Smith and E. Frank, "Statistical genomics: Methods and protocols," in Springer, 2016, ch. Introducing Machine Learning Concepts with WEKA, pp. 353–378. [Online]. Available: `http://dx.doi.org/10.1007/978-1-4939-3578-9_17`.

[78] C. Wee, W. Sung, H.-C. Kang, *et al.*, "Evaluation of variability in target volume delineation for newly diagnosed glioblastoma: A multi-institutional study from the Korean Radiation Oncology Group," *Radiation Oncology*, 2015.

[79] F. Zhao, M. Li, L. Kong, G. Zhang, and J. Yu, "Delineation of radiation therapy target volumes for patients with postoperative glioblastoma: A review," *OncoTargets and Therapy*, 2016.

[80] H. Clifton, A. Vial, D. Stirling, and et al., "Using machine learning applied to radiomic image features for segmenting tumour structures," *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, 2019.