Radiogenomics for Radiation Treatment Assessment of Advanced Lung Cancers

by

Jingxi Weng

Graduate Program in Medical Physics
Duke University

Date:_____

Approved:

_____
Fang-Fang Yin, Supervisor

_____
Michael Corradetti

_____
Mark Oldham

Thesis submitted in partial fulfillment of
the requirements for the degree of
Master of Science in the Graduate Program in Medical Physics
in the Graduate School
of Duke University

2019

ABSTRACT

Radiogenomics for Radiation Treatment Assessment of Advanced Lung Cancers

by

Jingxi Weng

Graduate Program in Medical Physics
Duke University

Date:_____
Approved:

_____
Fang-Fang Yin, Supervisor

_____
Michael Corradetti

_____
Mark Oldham

An abstract of a thesis submitted in partial
fulfillment of the requirements for the degree
of Master of Science in the Graduate Program in Medical Physics
in the Graduate School of
Duke University

2019

Copyright by
Jingxi Weng
2019

Abstract

Background: Radiomics describes the study of converting medical images into high-dimensional quantitative features and following analysis for further decision making and genomics focuses on the understanding genomes of individual organisms and characterizations of different genomes. Radiogenomics is a new emerging method that combines both radiomics and genomics together in clinical studies as well as researches the relation of genetic characteristics and radiomic features. It has the potential as a tool for medical treatment assessment in the future. In this study, we used machine learning methods to build two models for treatment assessment: 1) the output is p53 mutation, and the inputs are radiomic features; 2) the output is patient overall survival, and the inputs are radiomic features and p53 mutation. The modelling process was divided into feature selection and classification. Machine learning is a popular area of artificial intelligence that can make machines "learn by itself". Machine learning algorithm learns from datasets called "training data", and generates a prediction model from its learning process. The prediction model can then be used to make predictions and decisions from other datasets.

Purpose: 1) To investigate the correlation between p53 mutation and radiomic features in lung cancer, and to detect p53 mutation from radiomic features using different machine learning methods, 2) To investigate the correlation between genomic (p53

mutation), radiomic, radiogenomic features and overall patient survival in lung cancer using machine learning methods.

Material and Methods: The study used 24 patients with advanced lung cancers who had received radiotherapy and chemotherapy. CT was used as medical imaging modality in radiomics study. A radiomics study was then performed which involved three parts: Pre-treatment (Pre-Tx) Radiomics, Post-treatment (Post-Tx) Radiomics, and Delta Radiomics. The pre-Tx radiomic features were calculated from treatment planning CT images, the post-Tx radiomic features were calculated from the follow-up CT images after the radiotherapy, and the delta radiomic features were calculated as the change of radiomic features between cancer treatment. 19 of 24 patients had both pre-Tx and post-Tx CT images. Totally 61 representative radiomic features were extracted from CT images, including Intensity features, Grey Level Co-occurrence Matrix features, Grey Level Run Length Matrix Features, Grey Level Size Zone Matrix features, Neighborhood Grey Level Difference Matrix features, and Morphological features. Feature selection was implemented to avoid feature redundancy. Spearman Correlation analysis and Lasso regression were used for feature selection for p53 mutation detection. Cox regression and lasso regression were used for feature selection for patient survival prediction. Then, several common machine learning based classification methods were used for modelling of p53 mutation detection and patient survival prediction, including linear discriminative analysis, quadratic discriminative analysis, Naïve Bayes, Linear

Support Vector Machine, Kernel Support Vector Machine, Bootstrap Aggregating (Bagging), Logistic Regression, and Lasso generalized linear regression. Radiomic models were used for p53 mutation detection in tumor. Radiogenomic models based on combined radiomic features and p53 mutation were used for patient overall survival prediction. To avoid bias, the leave-one-out cross validation method was used for both feature selection and classification. Receiver Operator Characteristic (ROC) Curves were used as an evaluation method for the model, and Area Under Curve (AUC) values were compared for different classification methods.

Results: For p53 mutation detection, the highest AUC of pre-Tx radiomics (24 patients), pre-Tx radiomics (19 patients), and post-Tx radiomics (19 patients) was 0.6993, 0.5606, and 0.6591. For patient survival prediction, the highest AUC of pre-Tx radiomics (24 patients), pre-Tx radiomics (19 patients), post-Tx radiomics (19 patients), and delta radiomics (19 patients) was 0.7045, 0.7125, 0.6063, and 0.8000, and the highest AUC of pre-Tx radiogenomics (24 patients), pre-Tx radiogenomics (19 patients), post-Tx radiogenomics (19 patients), and delta radiogenomics (19 patients) was 0.7500, 0.7375, 0.5857, and 0.9143.

Conclusion: From limited dataset, it might be feasible to detect p53 mutation by both pre-Tx and post-Tx radiomics. Lasso and LSVM has shown the best performance in classification.

For predicting the overall patient survival, different features were selected. This may be related to the limited data available for the study. It may also be related to the different characteristics of pre-Tx, post-Tx and delta radiomics. Intensity and texture features showed high frequency being selected for pre-Tx and delta features, and morphological features showed high frequency for post-Tx radiomics. However, we also found that the combination of delta radiomics and p53 mutation showed a better patient survival prediction than pre-Tx, post-Tx, delta radiomics and p53 mutation alone. The reason might be related to the difference of tumor reaction to radiation due to p53 mutation. KSVM and Bagging showed highest performance compared with other classification methods.

Keyword: Radiogenomics, radiomics, delta radiomics, genomics, p53, lung cancer, radiotherapy.

Contents

Abstract iv

List of Tables x

List of Figures xi

Acknowledgements xiii

1. Introduction 1

    1.1 Background 1

    1.2 Previous Work 11

    1.3 Research Aims 13

2. Material and Methods 14

    2.1 Materials 14

    2.2 Methods 16

3. Results 35

    3.1 p53 Mutation Detection 35

    3.2 Patient Survival Prediction 39

    3.3 Pre-Tx Radiomic Patient Survival Model （24 Patients） 42

    3.4 Pre-Tx Radiomic Patient Survival Model （19 Patients） 43

    3.5 Post-Tx Radiomic Patient Survival Model （19 Patients） 44

    3.6 Delta Radiomic Patient Survival Models (19 Patients) 45

    3.7 Genomic Patient Survival Model 46

3.8 Pre-Tx Radiogenomic Patient Survival Models (24 Patients) 48
3.9 Pre-Tx Radiogenomic Patient Survival Models (19 Patients) 50
3.10 Post-Tx Radiogenomic Patient Survival Model 51
3.11 Delta Radiogenomic Patient Survival Model 52
3.12 Comparison of Different Models 53
4. Discussion 54
4.1 p53 Mutation Detection 54
4.2 Patient Survival Prediction 57
4.3 Relation Between p53 Mutation and Radiomics 62
4.4 Limitation 62
5. Conclusion and Future Work 63
5.1 Conclusion 63
5.2 Future Work 63
References 65

List of Tables

Table 1: Clinical characteristics of lung cancer patients 15

Table 2 : Extracted Radiomic Features 21

Table 3: Selected Features for p53 Mutation Detection in First Feature Selection 36

Table 4: Selected Features for p53 Mutation Detection in Second Feature Selection 37

Table 5: AUC Comparison of p53 Detection Models 39

Table 6: Selected Features for Patient Survival Prediction in First Feature Selection 40

Table 7: Selected Features for Patient Survival Prediction in Second Feature Selection ..41

Table 8: AUC Values for Patient Survival Prediction (Pre-Tx, 24 Patients)..... 43

Table 9: AUC Values for Patient Survival Prediction (Pre-Tx, 19 Patients)..... 44

Table 10: AUC Values for Patient Survival Prediction (Post-Tx, 19 Patients) 45

Table 11: AUC Values for Patient Survival Prediction (Delta, 19 Patients) 46

Table 12: AUC Values for Patient Survival Prediction (p53 Mutation)..... 48

Table 13: AUC of Pre-Tx Radiogenomic Model (24 Patients)..... 49

Table 14: AUC of Pre-Tx Radiogenomic Model (19 Patients)..... 50

Table 15: AUC of Post-Tx Radiogenomic Models 51

Table 16: AUC of Delta Radiogenomic Models 52

Table 17: Highly Frequent Features in p53 Mutation Detection 54

Table 18: High Frequently Selected Features in Patient Survival Prediction 57

List of Figures

Figure 1: Flow of Radiomics Study 4

Figure 2: Molecular Structure of protein p53 (Cho, Gorina, Jeffrey 1994) 10

Figure 3: Overall Scheme of Study 14

Figure 4: Genetic Mutation of Patients 16

Figure 5: Overall Research Flowchart 16

Figure 6: Workflow of Radiomic Feature Selection 18

Figure 7: Workflow of Cross Validation and Model Evaluation 18

Figure 8: Constraint of L1-norm 26

Figure 9: Maximum Internal between two hyperplanes in LSVM 28

Figure 10: Overfitting 31

Figure 11: Categories of prediction results in ROC curve analysis 32

Figure 12: Generation of ROC Curve 33

Figure 13: ROC Curves of p53 Mutation Detection with Pre-Tx Radiomics (24 Patients) 37

Figure 14: ROC Curves of p53 Mutation Detection with Pre-Tx Radiomics (19 Patients) 38

Figure 15: ROC Curves of p53 Mutation Detection with Pre-Tx Radiomics (19 Patients) 38

Figure 16: ROC Curves for Patient Survival Prediction (Pre-Tx, 24 Patients) 42

Figure 17: ROC Curves for Patient Survival Prediction (Pre-Tx, 19 Patients) 43

Figure 18: ROC Curves for Patient Survival Prediction (Post-Tx, 19 Patients) 44

Figure 19: ROC Curves for Patient Survival Prediction (Delta, 19 Patients) 45

Figure 20: ROC Curves for Patient Survival Prediction from p53 mutation (24 Patients) 47

Figure 21: ROC Curves for Patient Survival Prediction from p53 mutation (19 Patients) 47

Figure 22： ROC Curves of Pre-Tx Radiogenomics (24 Patients) 49

Figure 23: ROC Curves of Pre-Tx Radiogenomic Models (19 Patients) 50

Figure 24: ROC Curves of Post-Tx Radiogenomics 51

Figure 25: ROC Curves of Delta Radiogenomics 52

Figure 26 Comparison of Different Patient Survival Prediction Models (19 Patients) 53

Acknowledgements

I want to thank my advisor, Dr. Fang-Fang Yin for his instructive mentorship and informative advice to my research.

I want to thank Dr. Michael Corradetti for his precious research data, CT contouring, and being my thesis committee.

I want to thank Dr. Mark Oldham for being my thesis committee.

I want to thank Dr. Corbin Jacobs for the CT contouring.

I want to thank Dr. Kyle Lafata, Dr. Lei Zhang and Yushi Chang who helped me with my research.

I want to thank my roommates Haoming Li and Si Cen who helped me a lot in daily life.

Also, I ‘d like to thank my friends in the lab, and other students in the medical physics graduate program, especially Ruiqi Li, Wenchih Tseng, Zhanerke Abisheva, Chuan He, and Ran Wei.

At last, thank my family.

1. Introduction

1.1 Background

1.1.1. Radiomics

Radiomics, composed of the word root "Radio" and "omics". "Radio" stands for radiation, and "omics" comes from ancient Greek (Oxford English Dictionary, 2019), which means "all constituents considered collectively". As its name already shows, radiomics is a process that converts medical images into high-dimensional quantitative features. The features, can then be used as training data for decision making in medicine. Radiomics shows important roles in precision medicine, thanks to its non-invasive characteristics. The new emerging machine learning technique also promotes the development of radiomics.

Radiomics has shown its ability in many areas in medicine, including prediction of patient survival (Aerts, Velazquez Leijenaar et al., 2014), distant metastasis (Coroller, Grossmann, Hou et al., 2015), cancer recurrence (Li, Zhu, Burnside et al., 2016), cancer stages (Liang, Huang, He et al., 2016), cancer risks (Hawkins, Wang Liu, et al., 2016), and genetic features (Li, Qian, Xu et al., 2017). The prediction ability of radiomics makes it a powerful tool for treatment assessment in medicine.

Feature extraction is the key process of radiomics. Common radiomic features can be classified into several groups (Hugo J. W. L. Aerts, 2014; Alex Zwanenburg, 2018): (1)

First Order Features, (2) Morphological Features, (3) Texture Features, (4) Wavelet Features.

(1) First order features describe the voxel intensities distribution within the CT images. Usually it can be divided into three main types: local intensity features, statistical features, and intensity histogram features. (2) Morphological features describe the geometrical aspects of the region, for example shape, sphericity, and volume. (3) Texture features are one of the most important type of radiomic features. Texture features describe the texture of image, in other words, the spatial distribution of the image. (4) Wavelet features are based on wavelet transform of the image. Wavelet transform is an image transform method that transform original into a combination of multiple wavelets.

Delta radiomics is a new emerging area of radiomics which focuses on changes of radiomic features. (Fave, Zhang, Yang, 2017) Usually, delta radiomic features are calculated based on the radiomic features at two different time stages, e.g. before and after the cancer treatment. Because delta radiomics contains the information of time or treatment, it has been used for prognosis of cancer study in many cancers.

1.1.2 Genomics

Genome means the genetic material of an organism. Similar to radiomics, genomics is the biological study focuses on genomes, especially the characterization, structure, evolution, mapping, and editing of genomes. (National Human Genome

Research Institute) Genomics has already been widely used in biology, medicine, bioengineering, and anthropology. One of the most famous project in genomics is Human Genome Project which holds the goal of DNA sequencing of all the genes of human genome. This plan was started at 1984, and declared complete at 2003. Another popular area in genomics is genomic medicine, which is part of precision medicine. Genomic medicine is an emerging medical study that combines genomic information of each patient into their treatment. Therefore, each patient can have their own treatment plan based on their individual genomes. (National Human Genome Research Institute)

1.1.3 Radiogenomics

The definition of radiogenomics is still ambiguous. (Mazurowski, 2014) So far, radiogenomics contains two different aspects of medical research. One common use of radiogenomics refers to the relation between specific genomes of patients and the reaction of patient to radiation. In 2009, the Radiogenomics Consortium was established with the research focus in this area.

Another use of radiogenomics refers to the combination of radiomics and genomics, also known as imaging genomics. It can be used to refer to any relationship between medical imaging characteristics and genomics, gene expression patterns, gene mutations, and other gene-related characteristics. (Pinker, Shitano, Sala et al., 2018) Radiogenomics has been a rapidly developing area because of the non-invasive nature of medical imaging and a deeper understanding of tumor genetics. So far,

radiogenomics has shown important applications in many different potential tumor sites including brain, prostate, liver, breast etc. (Pinker, Shitano, Sala et al., 2018)

1.1.4 Flow of Radiomics Study

A typical radiomics study includes several steps: (1) Image Acquisition, (2) Image Segmentation, (3) Feature Extraction, (4) Feature Selection (5) Modelling, as shown in Figure 1.

Figure 1: Flow of Radiomics Study

1.1.4.1 Image Acquisition

Image acquisition is the process of acquiring medical images of patients. Common medical imaging techniques used in radiomics include computed tomography (CT), magnetic resonance imaging (MRI), cone-beam computed tomography (CBCT), positron emission tomography (PET), and ultrasound imaging. Medical images usually come from the clinical treatment of patients or medical database.

1.1.4.2 Image Segmentation

Image segmentation refers to the process of acquiring regions of interest (ROI) from medical images. This process usually includes the identification of ROI and then the contour of ROI. The identification of ROI is one of the key points in radiomics study

since all the quantitative features are extracted from the ROI volume. Image segmentation is also a challenging part in radiomics study since the boundary between ROI and other parts may be unclear. Image segmentation can be achieved by either manually, semi-automatically, or automatically. Gross Tumor Volume (GTV) or Planned Tumor Volume (PTV) is usually chosen as ROI, but it is also possible that other organs are considered as ROI.

1.1.4.3 Feature Extraction

Feature extraction is the process of extracting radiomic features from ROI. There is no identical criterion for which features are extracted, and extracted features differs study by study. Common extracted features include First Order Features, Morphological Features, Texture Features, Wavelet Features as mentioned above. Feature extraction can be achieved by many programs or codes. Usually the image is first imported as DICOM images, then ROI is mapped from the whole image. The corresponding matrixes are calculated, and the features are calculated based on their definition. Finally, the calculated features are exported into corresponding files.

1.1.4.4 Feature Selection

Feature selection is another crucial step in radiomics study. Feature selection is the process to select useful features from huge amount of features. Extracted features in radiomics are usually in huge amount and probably will cause overfitting if all of them are used in modelling. Overfitting is the phenomenon that the trained model can fits

pretty well for the training samples but fits badly for the validation group. The main reason for overfitting is that generated model contains more parameters than the data has justified. Too many parameters will make the model fit the training data with a high accuracy but shows a poor performance in validation group or in general occasions. Therefore, feature selection is essential to avoid overfitting, especially when the number of features is larger than the degree of freedom of data. Besides avoiding overfitting, feature selection can also decrease the time needed for modelling and provide a simpler model.

1.1.4.5 Decision Making Modeling

Decision Making Modeling is the process of building a model for decision making based on selected useful features. The model will then be applied to generate a prediction, or a classification for diagnosis or prognosis. For example, a classification model can be used to differentiate tumors at different stages, or predict which patients are likely to suffer from cancer recurrence etc.

1.1.5 Machine Learning in Radiomics Study

Machine learning is a study in which statistical models or algorithms are built based on patterns instead of detailed instructions. In other words, machine learning is a process that machine can learn from itself. Machine learning is an important tool in radiomics analysis, especially in feature selection and modelling. Machine learning based methods consider all the features as pure numbers and try to conclude the pattern

of numbers. It can find the relationship between features that are unclear or obscure to humans. Popular machine learning methods used in radiomics include neural network, random forest, support vector machine and other methods.

1.1.6 Lung cancer

Lung cancer is one of the most common cancers in the world. It is the most deadly cancer among both males and females. There were approximately 2 million new cases of lung cancer in 2018. (World Cancer Research Fund, 2019) Long term tobacco smoking is the most common cause of lung cancer and approximately 90% of the lung cancer patients have smoking-related experience. In the United States, only 17.4% of people diagnosed with lung cancer survive for than five years, which is a relatively low survival compared with average survival of cancer patients (The average 5-year survival of cancer patients in the US from 2007 to 2013 is 69%). (National Cancer Institute, 2016; American Cancer Society, 2018)

Based on the form of tumor cells, lung cancer can be divided into two broad types: non-small-cell lung carcinoma (NSCLC) and small-cell lung carcinoma(SCLC). NSCLC accounts for 80% of world lung cancer.

1.1.7 Genetic Mutation of Tumor

1.1.7.1 Carcinogenesis

Carcinogenesis is a biological process which involves the formation of tumor. More specifically, it is a process involving the transformation of a normal cell to a tumor

cell. The process is highly related to the cell division, cell mutation and the genome of cells. Cell mutation can be caused by various reasons including but not limited to radiation, chemicals, mistakes in reproduction etc. Normally, harmed or mutated cells in human body will go through a process called apoptosis, which is like a "suicide" mode for cells. Thanks to the existence of apoptosis, harmed or mutated cells will be cleared from human body. However, it is also possible that mutated cells keep reproducing by itself, and finally lost control. The unlimited reproduction of cells will then develop into malignant tumor, or cancer.

1.1.7.2 Oncogene and Tumor Suppressor Gene

Two types of genes hold important positions in carcinogenesis, which are oncogenes and tumor suppressor genes. Oncogenes are genes that have the potential to cause cancer. Oncogenes are usually mutated or highly expressed in tumor cells. Oncogenes are developed from proto-oncogenes which usually show important roles in many essential cell functions including cell growth, cell proliferation and apoptosis. However, proto-oncogenes may develop into oncogenes due to various mutations. Common proto-oncogenes include EGFR, VEGF, HER2, etc (Weinstein and Joe, 2006).

Tumor suppressor gene, on the other hand, are genes that regulate the production of protein that functions as tumor suppressor. When the tumor suppressor gene is mutated or lose its function, the cell is more likely to progress into cancer. Tumor suppressor genes function in several aspects: It can suppress the cell cycling, repair

damaged DNA, and initiate apoptosis (Sherr, 2004). Some already known tumor suppressor genes include TP53, PTEN, etc.

1.1.7.3 p53

p53 (also known as TP53) is one of the earliest discovered tumor suppressor genes and probably the best known tumor suppressor gene ever found. The structure of p53 protein is shown below in Figure 2. Gene p53 regulates the production of protein p53, which is an important regulator in repairing DNA damage and regulation of apoptosis. In a normal cell, protein is inactivated until it is activated by outside signals like DNA damage or hypoxia. After the activation of p53 protein, it will arrest cell cycle and induce one of the two pathways: DNA repair or cell apoptosis. Whether the cell goes through apoptosis or start DNA repairing depends on the situation of the cell (Harris and Levine, 2005). Protein p53 is well known as the "guardian of genome" as its role in preventing cancer.

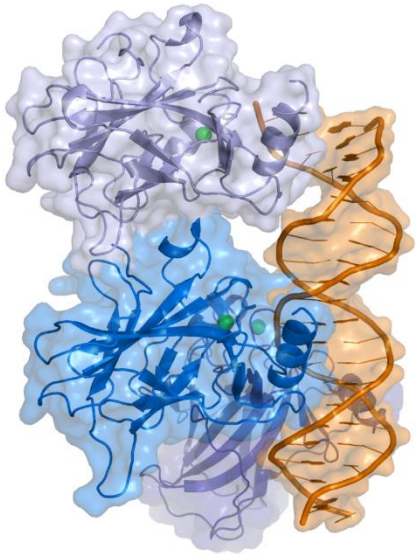

Figure 2: Molecular Structure of protein p53 (Cho, Gorina, Jeffrey 1994)

Gene p53 is especially associated with lung cancer. p53 mutation has been examined in 70% to 90% of small cell lung cancer (SCLC) patients (Huret, 2014) and 50% of non-small cell cancer (NSCLC) patients (NSCLC) (Mogi and Kuwano, 2011). Previous patient statistics have also showed that p53 mutation can be used as a predictor in NSCLC. (Ahrendt, Hu, Buta et al., 2003) In 2017, researchers conducted a research about correlations between radiomic features and several biological pathways, but not direct correlation between p53 mutation and radiomics is found so far. (Grossmann, Stringfield, El-Hachem et al., 2017)

1.1.8 Cell Free DNA (cfDNA)

Cell free DNA, as the name already shows, is DNA fragments outside the cells in human body. One important source of cfDNA is from tumor cells, especially from

circulating tumor cells that are circulating in human blood vessels. Therefore, cfDNA has the potential to be used for tumor examination. Compared with other tumor examination methods like biopsy, the sampling of cfDNA does much less harm to patients. However, there are also important drawbacks of using cfDNA versus tumor examination, since not all cfDNA comes from tumor, and the characteristic of cfDNA depends on various factors besides tumor.

1.2 Previous Work

1.2.1 Previous Work in Machine Learning in Radiomics

Machine learning has been widely used in radiomics because of its ability to build self-learning models from various features. Common outputs of machine-learning based radiomic models include prediction of overall patient survival (Parmar, Grossmann, Bussink et al., 2015; Chang,2018), cancer recurrence (Saha, Harowicz, Wang et al., 2018), prostate lesions (Kwon, Reis, Breto et al., 2018), kidney stone (De Perrot, Hofmeister, Burgermeister et al., 2019), genotype (Wu, Jin, Yu et al., 2019). Common machine learning methods in radiomics study include but not limited to support vector machine (SVM), neural network, random forest, logistic regression, least absolute shrinkage and selection operator (LASSO).

1.2.2 Previous Work in Delta Radiomics

Delta radiomics introduces the time component into radiomics study. Delta radiomics analyses the differences between radiomic features at two different time stages. Research involving delta radiomics is still very limited, but some work has demonstrated the utility of delta radiomics in the analysis of overall survival in lung cancer. (Fave, Zhang, Yang et al., 2017) The inclusion of delta radiomics has significantly improved the prediction of overall survival compared with the original model with clinical factors and pre-Tx radiomics features.

1.2.3 Previous Work in Radiogenomics

Similar to delta radiomics, radiogenomics is also a new area related with radiomics. Although relatively new, there are already many exciting results in radiogenomics. Previous results show that specific biological pathways can be predicted by radiomics and models with combination of genomic, radiomic, and clinical features achieved highest prognostic result in lung cancer. (Grossmann, Stringfield, El-Hachem et al., 2017) Also, there are studies have shown that mutation of p53/PTEN/EGFR and other common oncogenes or tumor suppressor genes can be predicted by radiomic features in brain cancer. (Li, Qian, Xu et al., 2018; Zinn, Singh, Kotrotsou et al., 2017) Besides lung cancer and brain cancer, similar studies are also implemented in liver cancer, breast cancer, breast cancer, gynecological cancer, etc. (Pinker, Shitano, Sala et al., 2018)

Although there is more and more radiogenomics research, so far there is no research focusing on the combination of radiogenomics and delta radiomics.

1.3 Research Aims

Aim 1: To investigate the correlation between p53 mutation and radiomic features in lung cancer, and to detect p53 mutation from radiomic features using machine learning methods.

Aim 2: To investigate the correlation between genomic (p53 mutation), radiomic, radiogenomic features and overall patient survival in lung cancer using machine learning methods.

2. Material and Methods

2.1 Materials

2.1.1 Patient Statistics

The study has followed Duke IRB protocol Pro00076119. In the study, 24 patients with advanced thoracic malignancies were reviewed. (Corradetti, Torok, Hatch et al., 2017) All the patients were treated with Intensity Modulated Radio Therapy (IMRT) with Cisplatin-etoposide chemotherapy. The average dose was 66 Gy, and ranged from 58 to 74 Gy. Blood plasma was sampled pre-Tx, at weeks 2, 5, and 10 weeks post-Tx. One overall workflow was shown in Figure 3. Other clinical factors were shown in Table.1. The survival of patients was documented for the follow-up of the study. For the patients who are alive now, the follow-up of the survival is censored/ended.

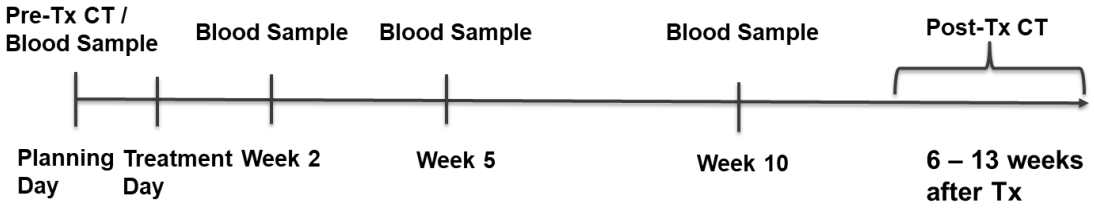

Figure 3: Overall Scheme of Study

Table 1: Clinical characteristics of lung cancer patients

Clinical Factors		Number of Patients (24)
Gender	Male	12
	Female	12
Tumor Histology	NSCLC	21
	SCLC	3
1 Year Survival	≥ 1 year	20
	≤ 1 year	4
2 Year Survival	≥ 2 years	12
	≤ 2 years	12
Smoking Habit	Yes	20
	No	4

2.1.2 Tumor Genetic Mutation

The cell free DNA (cfDNA) of the patients was collected before the treatment, and at 2 weeks, 5 weeks, and 10 weeks after the treatment started. cfDNA was isolated by the Maxwell RSC cfDNA Plasma Kit and quantified using the QuantiFluor ONE dsDNA system. Of 90 cfDNA samples, 58 samples were detected with mutation. In this study, we focused on p53 mutation, which showed the highest occurrence among all the detected mutations, as shown in Figure 4. Other genetic mutations were considered as too few to provide a reliable result.

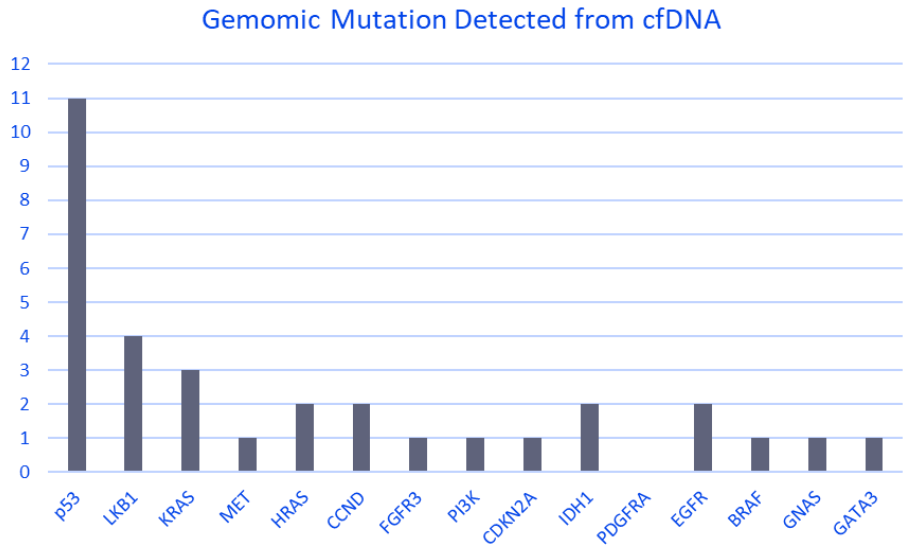

Figure 4: Genetic Mutation of Patients

2.2 Methods

2.2.1 Research Workflow

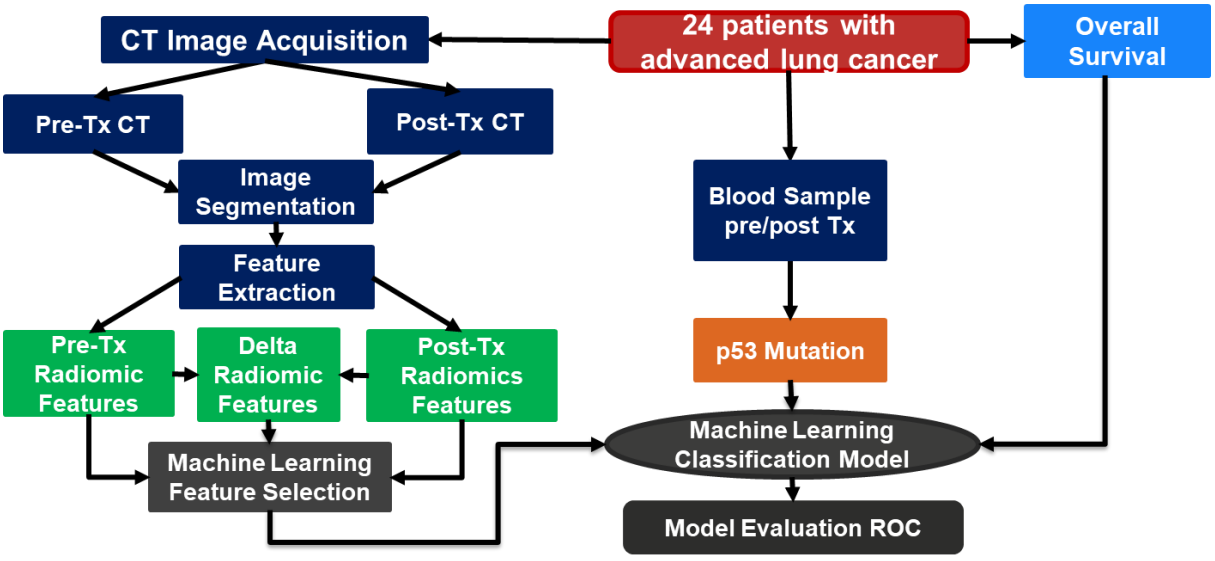

Figure 5: Overall Research Flowchart

The flowchart shown in Figure 5 is the overall workflow of the research. 24 patients with advanced lung cancers were used as the data source. (19 of 24 patients had both pre-Tx and post-Tx CT images) The survival condition of patients was recorded, shown as overall survival in the figure. In this study, 2-year survival was used as indicator of patient survival, in which "1" means the patient survived more than 2 years, and "0" means the patient didn't survive for more than two years. p53 was used as genomic information in this study. p53 mutation information was detected from blood sample of patients. For recording of genetic mutation, "1" was used for mutation group, and "0" was used for wild type group. The radiomic features were acquired based on procedure of radiomics study. First, both pre-Tx and post-Tx CT images were acquired. The contours of the pre-treatment CT images were accomplished by corresponding physicians for their treatment planning, and the post-treatment CT images were contoured by two radiation oncologists. Then, the radiomic feature extraction was accomplished with an in-house feature extraction toolbox developed by Duke Radiation Oncology, and 61 common radiomic features in total were extracted. The radiomic features included intensity features, grey level co-occurrence matrix (GLCOM) features, grey level run length matrix (GLRLM) features, grey level size zone matrix (GLSZM) features, neighborhood grey level difference matrix (NGLDM) features, and morphological features. All the features were normalized using Z score. Delta radiomic features were calculated as the relative change between post-Tx radiomic features and

pre-Tx radiomic features. Then, all the radiomic features were applied for feature selection. The method used for feature selection was different for two different aims. Selected radiomic features were then applied in classification models for p53 mutation detection or patient survival prediction. Leave-one-out cross validation was used to avoid overfitting. The models were evaluated with ROC curve analysis.

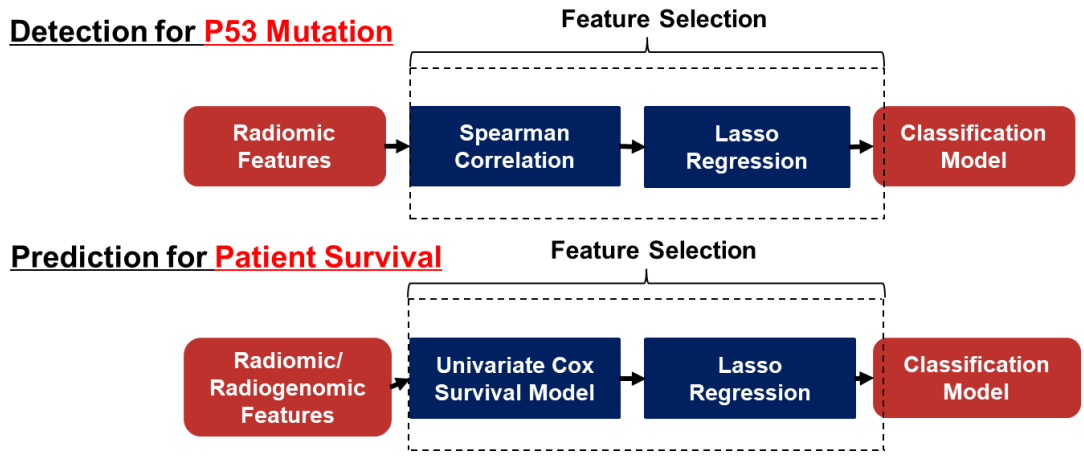

Figure 6: Workflow of Radiomic Feature Selection

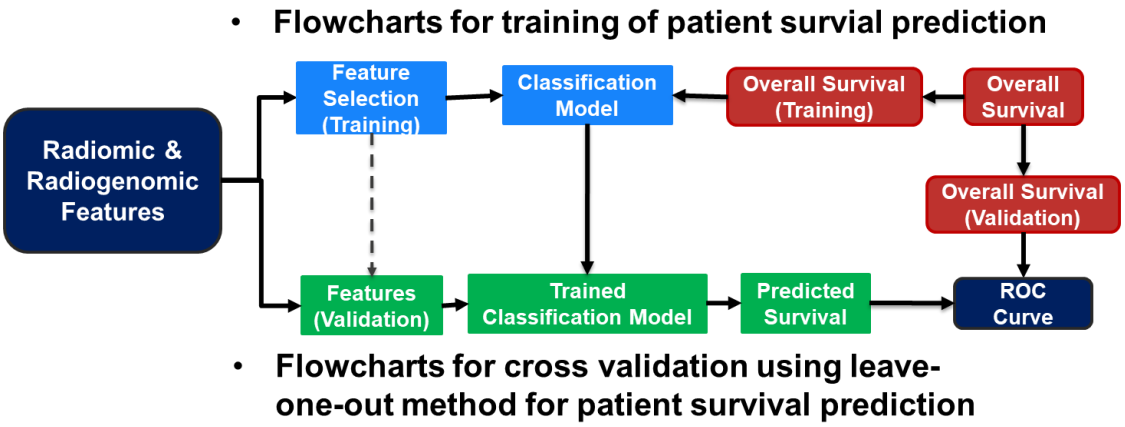

Figure 7: Workflow of Cross Validation and Model Evaluation

Figure 6 shows the workflow of feature selection process in radiomics, and Figure 7 shows the workflow of cross validation and model evaluation. The dash line in Figure 7 means that the features used for validation is same genre with the selected features in the training process.

2.2.2 Image Acquisition

For each patient, the CT images applied in treatment planning of radiotherapy were used for pre-treatment CT images. The first CT images after the radiotherapy were acquired as post-treatment CT images. One patient passed away before any post-Tx CT images was taken. Therefore, only 23 post-Tx CT images were taken in the study.

2.2.3 Image Segmentation

Gross tumor volume (GTV) was contoured as ROI in this study. The image segmentation was achieved in two stages. The pre-Tx contours were finished by corresponding physicians in the radiotherapy for each patient. The post-Tx contours were finished by two radiation oncologists. There was no tumor found in lung after treatment for 5 patients, so only 19 patients were contoured for post-Tx radiomics.

2.2.4 Feature Extraction

The feature extraction is achieved using an in-house feature extraction tool developed based on MATLAB 2017a by Duke Hospital Radiation Oncology. The image was first divided into 64 grey level bins, and then ROI is mapped based on the contouring. In total 61 features were extracted. These 61 features included common

genres of radiomic features in cancer study, and were used in various previous radiomics studies. (Zwanenburg, Leger, Valliere et al, 2018) The 61 features include 4 intensity histogram based features, 22 grey level co-occurrence matrix (GLCOM) texture features, 11 grey level run length matrix (GLRLM) texture features, 13 grey level size zone matrix (GLSZM) features, 5 neighboring grey level difference matrix (NGLDM) features, and 6 morphological features. The mathematical calculation of radiomic features could be seen in the reference (Zwanenburg, Leger, Valliere et al, 2018). All the radiomic features extracted are listed in the Table.2. In this study, all the features are indicated with numbers from 1 to 61.

The delta feature is calculated as the relative change of radiomic features before and after cancer treatment. The formula is listed in Equation 1.

$$Delta\ features = (Features_{post} - Features_{pre})/Features_{pre}$$

Equation 1: Equation for Delta Feature Calculation

Table 2 : Extracted Radiomic Features

Feature Type	Feature Name
Intensity Features	Energy
	Entropy
	Kurtosis
	Skewness
Gray Level Co-occurrence Matrix Features	AutoCorrelation
	Cluster Prominence
	Cluster Shade
	Cluster Tendendy
	Contrast
	Correlation
	Differential Entropy
	Dissimilarity
	Energy2D
	Entropy2D
	Homogeneity1
	Homogeneity2
	Info Measure Correlation1
	Info Measure Correlation2
	Inv Difference Moment Normalized
	Inv Difference Normalized
	Inv Variance
	Maximum Probability
	Sum Average
	Sum Entropy
	Sum Variance
	Variance

Feature Type	Feature Name
Gray Level Run Length Matrix Features	Short Run Emphasis
	Long Run Emphasis
	Gray Level NonUniformity
	Run Length Non Uniformity
	Run Percentage
	Low Gray Level Run
	High Gray Level Run
	Short Run Low Gray Level
	Short Run High Gray Level
	Long Run Low Gray Level
	Long Run High Gray Level
Gray Level Size Zone Matrix Features	SmallZoneEmphasis
	LargeZoneEmphasis
	GrayLevelNonUniformity
	SizeZoneNonUniformity
	SizePercentage
	LowGrayLevelSizeEmphasis
	HighGrayLevelSizeEmphasis
	SmallSizeLowGrayLevelEmp
	SmallSizeHighGrayLevelEmp
	LargeSizeLowGrayLevelEmp
	LargeSizeHighGrayLevelEmp
	VariationOfIntensity
	VariationOfArea
Neighborhood Gray Level Difference Matrix Features	Coarseness
	Contrast
	Busyness
	Complexity
	TextureStrength
Morphological Features	Compactness 1
	Compactness 2
	Sphericity
	SpherDisprop
	SurfaceArea
	Volume

2.2.5 Feature Normalization

Feature normalization was used to avoid possible impacts to feature selection result due to large variations of feature values. All the extracted features were normalized using Z score. The formula of Z score calculation is shown in Equation 2. μ is the mean of dataset, and σ is standard deviation.

$$z = \frac{x - \mu}{\sigma}$$

Equation 2: Calculation of Z Score

2.2.6 Feature Selection

The feature selection process consisted of two steps for both p53 mutation detection and patient survival prediction. To detect p53 mutation, first a Spearman Correlation test was implemented with all radiomic features, and Lasso regression were used for the second step. In patient survival prediction models, the first step was a univariate Cox Regression Model, and the second step was also Lasso Regression. The details are shown below.

2.2.6.1 Spearman Correlation

Spearman correlation coefficient is a common coefficient used to evaluate the correlation between two variables. It is calculated based on the numeral ranking of variables without assumption of Gaussian Distribution unlike Pearson Coefficients. As the first step to select features for p53 mutation detection, features with a p-value smaller than 0.1 were selected. p-value smaller than 0.1 means that we have 90%

confidence that two features were correlated instead of coincidence. 0.1 is a relatively loose criterion for significance. We used this moderate limit on one hand to filter those irrelevant features, on the other hand to reserve features as many as possible for next step of feature selection.

2.2.6.2 Cox Regression Model

Cox regression model, also known as Cox proportional hazard model, is a common model used in patient survival analysis. (Cox, 1972) Cox model is applied for the first step to select those features that have high correlation with patient survival.

$$\mathrm{H}(x_i, \mathrm{t}) = h_0(t)\exp(\sum_j x_{ij} b_j)$$

Equation 3: Cox Hazard function

Equation 3 has shown the hazard function in Cox survival model, $H(x_i, t)$ is the hazard at time t for the patient (In survival analysis, hazard usually refers to the failure, disease recurrence or patient death in patient survival analysis), $h_0(t)$ is the baseline hazard function, $X_i = (x_{i1}, x_{i2}, x_{i3}, \dots, x_{ip})$ is the predictor value, and b is the impact of each feature on the hazard. One of the assumption of this model is that the baseline hazard function depends on time t, but the predictor variables do not depend on t. The hazard ratio, which is the ratio between hazard at time t and baseline hazard, represents the relative risk of failure with the predictor values.

In this study, univariate Cox models are built for each feature among the 61 extracted features. The goal of doing this is to identify which feature has significance

with patient survival. p-value is also calculated in the model. Features with a p-value < 0.1 were selected into next round of feature selection.

The patient survival was indicated with 2-year survival. If the patient had survived more than 2 years, it was recorded as class 1, or else it was recorded as class 0. In Cox regression model, the survival time of those patients who had survived more than 2 years was documented as 24 months as censored data.

2.2.6.3 Lasso Regression

"Lasso" method stands for Least Absolute Shrinkage and Selection Operator. Lasso regression is a machine-learning based process which can be used in both feature selection and classification. It is first introduced in geophysics in 1986 (Santosa and Symes, 1986), then is reintroduced in 1996 by Robert Tibshirani (Tibshirani, 1996). Lasso regression is very effective in reducing the number of features in complicated system. It is also compatible with many models and has shown good feature selection ability in Cox model. (Tibshirani, 1997) Therefore, here we have applied Lasso method for the second step of feature selection in both p53 mutation detection and patient survival prediction.

In lasso regression, the algorithm tries to minimize the cost function shown in Equation 4 below.

$$\text{Cost} = \frac{1}{N}||y - X\beta||_2^2 + \lambda||\beta||_1$$

Equation 4: Cost Function of Lasso Regression

$||\beta||$ means L1 or L2 norm, which is indicated as footprint number. $\lambda, ||\beta||_1$ works as a punishment component in the regression process, and λ value increases in each iteration. As λ increases, the algorithm needs to try to reduce the number of features to decrease the cost. λ is actually the Lagrangian multiplier of the limit function: $||\beta||_1 \leq t$. Since there is a limit for the L1-norm of β, it is possible that some coefficients will drop to zero as iteration continues, as shown in the Figure 8.

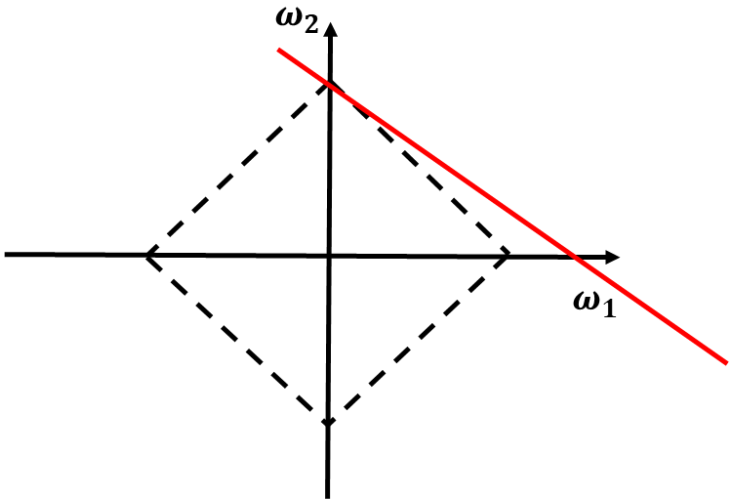

Figure 8: Constraint of L1-norm

2.2.7 Classification

The classification of binary outcome of p53 mutation or 2-year survival are implemented with several machine-learning based methods:

(1) Linear Discriminative Analysis (LDA): Discriminant analysis is a classification method invented by British scientist Sir. R. A. Fisher. In LDA, the model assumes that the features satisfy multivariate Gaussian Distribution, and it assumes that each class

has same covariance matrix but different means. The multivariate Gaussian Distribution formula is shown in Equation 5.

$$\mathrm{P(x|k)} = \frac{1}{(2\pi|\Sigma_k|)^{1/2}} \exp(-\frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1}(x-\mu_k))$$

Equation 5: Multivariate Gaussian Distribution

The classification result is shown as below in Equation 6. $\hat{y}$ is the predicted classification, K is number of classes, C(y|k) is the cost of classifying the class k into class y, P(x|k) is multivariate Gaussian Distribution, $\frac{P(k)}{P(x)}$ is the prior probability.

$$\hat{y} = \underset{y=1,\ldots,K}{\operatorname{argmin}} \sum_{k=1}^{K} \frac{P(x|k)P(k)}{P(x)} C(y|k)$$

Equation 6: Classification Prediction

(2) Quadratic Discriminative Analysis (QDA): The method is almost same with LDA. However, in QDA, both means and covariances are different from each other.

(3) Linear Support Vector Machine (LSVM): LSVM uses hyperplanes to differentiate different groups. Any hyperplane can be written in the form as Equation 7.

$$\vec{w} \cdot \vec{x} - b = 0$$

Equation 7: Hyperplane Function

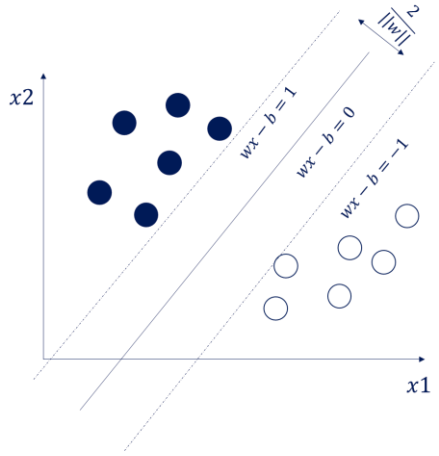

Figure 9: Maximum Internal between two hyperplanes in LSVM

As shown in the Figure 9, LSVM tries to make the interval of two groups of data as large as possible. If the interval is too small, it means that it is more likely that one class which is close the margin mistakenly classified into another class. The classifier is achieved by minimizing the loss function shown in Equation 8.

$$Loss = \left[\frac{1}{n}\sum_{i=1}^{n}\max\left(0,1-y_i(\vec{w}\cdot\vec{x}-b)\right)\right]+\lambda\|\vec{w}\|^2$$

Equation 8: Loss Function of SVM Method

(4) Kernel Support Vector Machine (KSVM): KSVM is similar to LSVM, and the only difference is that every dot product is replaced by nonlinear kernel function.

(5) Naïve Bayes: Naïve Bayes is a classification method based on Bayes theorem shown in Equation 9. The assumption of Naïve Bayes is that all features are independent with each other. Although the independence assumption is hard to accomplish, Naïve

Bayes can also show good performance in other conditions. Equation 10 has shown an example of Bayes classifier.

$$\mathrm{P(A|B)} = \frac{P(A)P(B|A)}{P(B)}$$

Equation 9: Bayes theorem

$$\hat{y} = arg \max_{k \in \{1,\ldots,k\}} p(C_k) \prod_{i=1}^{n} p(x_i|C_k)$$

Equation 10: Bayes Classifier

(6) Bagging: Bagging is the abbreviation of Bootstrap aggregating. Bootstrap aggregating is based on the idea of Bootstrap sample, which is proposed by Leo Breiman in 1994. (Breiman, 1994) To understand the idea of bagging, let's first assume a training set D with size d, everytime we will randomly select one sample and put it back for d times as training data. (The samples can be repeated) If we make this process for infinite times, then the probability of one data that is not chosen even for one time is 0.368, which means that there is a probability of 63.2% that one sample is in training dataset. Bagging is the combination of Bootstrap sample and simple classification methods. For example, Bagging in this study is the combination between Bootstrap samplling and 100 decision trees.

(7) Logistic Regression: Logistic Regression is a regression method mainly used for the modeling of binary variable. A logistic function is a sigmoid function which has a "S" shape. It has the form shown in Equation 11.

$$\mathrm{p(x)} = \frac{1}{1 + e^{-\theta^T X}}$$

Equation 11: Logistic Function

Logit is defined as the inverse of the logistic function, shown in Equation 12.

$$\text{logit p(x)} = \log(\frac{p}{1-p}) = \theta^T X$$

Equation 12: Definition of Logit

The parameter θ of the model is determined by the maximized likelihood function shown in Equation 13.

$$\mathrm{L(\theta|x)} = \mathrm{Pr(Y|X;\theta)} = \prod_i p_\theta x_i^{y_i} (1 - p_\theta(x_i))^{(1-y_i)}$$

Equation 13: Likelihood Function of logistic regression

(8) Lasso Regression: Lasso Regression is a regression analysis that can be used for both feature selection and regularization. Details of lasso have already been discussed in ”2.2.6.3 feature selection”.

2.2.8 Cross Validation

Overfitting is a common phenomenon in classification in which the classification result of training group achieves high accuracy, but achieves very low accuracy in validation group. One important reason of overfitting is that the number of parameters is too large, and the system is too complicated that the model has “over fitted”. Figure 10 is an example of overfitting. In this figure, the boundary with black color shows a good classification but still have some missing points. The green boundary shows a

complicated boundary that can achieve 100% accuracy, but this boundary is too complicated and is prone to validation.

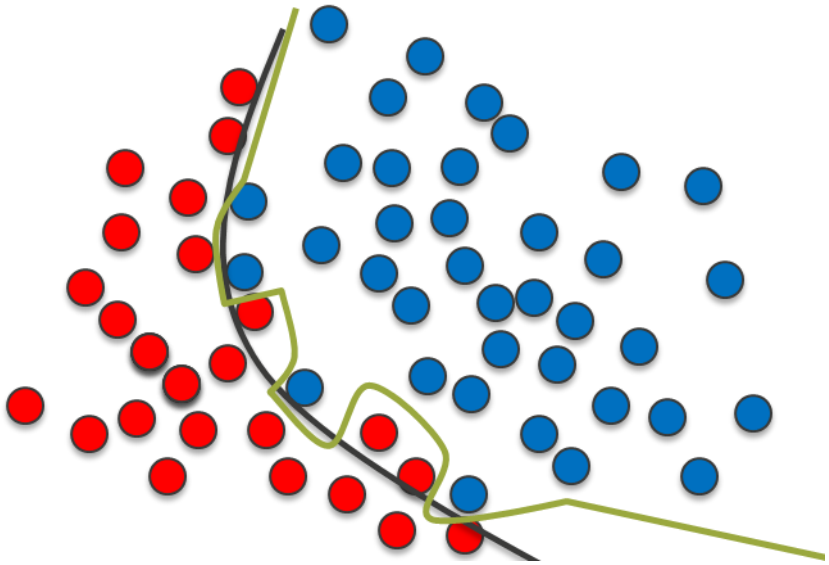

Figure 10: Overfitting

Overfitting is especially common when the number of features are larger than the size of data sets. Leave-one-out cross validation is used in this study to avoid overfitting. Cross validation means that instead of finding a separate validation group, we separate our data into training data and validation data. After the model is built, we tested the model with validation dataset. In this study, cross validation is used for both feature selection and classification. It means that every time we will use one patient as validation, and apply feature selection and classification methods to all the other patients. The model then is tested with the validation patient. The process runs for 24 or 19 times, based on the total number of patients. However, the first step of feature

selection, which is univariate Spearman Coefficient analysis or Cox modelling, is not affected by cross validation since the procedure is univariate.

2.2.9 Model Evaluation

In this study, ROC (Receiver Operating Characteristics) curve is used to evaluate the performance of a classification model. Invented during World War II, it was first used by the US army for the prediction of Japanese aircraft. (Green and Swets, 1966) Since then, ROC curve analysis became a very common method to analyze the performance of classifiers and prediction models and has been widely used in radiomics study. In this study, the outcome of the model is p53 mutation or patient survival, and it is represented as a binary system, in which 1 represents that there is p53 mutation detected or survival is longer than 2 years and 0 represents there is no p53 mutation detected or survival is shorter than 2 years. The outcome of a classifier or a prediction model can be classified into four categories: true positive (TP), false positive (FP), true negative (TN), and false negative (FN), shown in Figure 11. (Positive and Negative means two types of the binary outcome)

	Condition Positive	Condition Negative
Predicted Condition Positive	TP	FP
Predicted Condition Negative	FN	TN

Figure 11: Categories of prediction results in ROC curve analysis

$$\text{True positive fraction (TPF)} = \frac{TP}{TP + FN} = Sensitivity$$

Equation 14: Equation for calculation of TPF

$$\text{False positive fraction (FPF)} = \frac{FP}{FP + TN} = 1 - Specificity$$

Equation 15: Equation for calculation of FPF

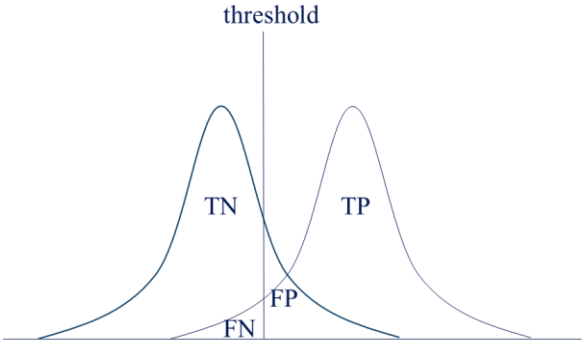

Figure 12: Generation of ROC Curve

True Positive Fraction (TPF) and False Positive Fraction (FPF) can then be calculated based on the results of prediction. The calculation of TPF and FPF are also shown above in Equation 14 and 15. TPF is also known as sensitivity, and (1 – FPF) is known as specificity. ROC curve is a graphical plot that illustrates the prediction of performance with the x axis shows false positive rate (FPF) and y axis shows true positive rate (TPF). The ROC curve is generated as the threshold of classification model changes from minimum value to maximum value, as shown in Figure 12. When the threshold is very low, all the negative and positive outcomes will be classified as positive, and TPF = FPF = 1. When the threshold is very high, all the negative and positive outcomes will be misclassified as negative, and therefore, TPF = FPF = 0. When

the threshold is moving from minimum to maximum, corresponding dots are plotted on the ROC coordinate. On the ROC curve of an optimal classifier, there is only one extra point which is (FPF = 0, TPF = 1). It means that as long as there is an existing threshold, the classifier can classify the outcome perfectly. However, it is almost impossible in reality. Mostly, there are multiple points on ROC coordinate as the threshold changes. If we connect the dots together, then we will get a ROC curve. The area under curve (AUC) of ROC is commonly used as a quantitative measurement for the performance of a classifier. A perfect classifier has AUC area of 1, a complete random classifier has AUC area of 0.5, and the worst classifier has AUC area of 0, but if we switch the classification result, it will be a perfect classifier.

3. Results

We extracted 61 radiomic features in both pre-Tx CT images and post-Tx CT images. 24 patients had pre-Tx contouring, and 19 patients had post-Tx contouring. For p53 mutation detection, we built three radiomic models, which were based on pre-Tx radiomics (24 patients and 19 patients), and post-Tx radiomics (19 patients). We built four radiomic models for patient survival prediction, which were based on pre-Tx radiomics (24 and 19 patients), post-Tx radiomics (19 patients), and delta radiomics (19 patients). We also built four radiogenomic models for patient survival prediction with the combination of corresponding radiomic features and p53 mutation as genomic features. Leave-one-out validation is used to avoid overfitting, and ROC curve analysis is used for model evaluation.

3.1 p53 Mutation Detection

After the first step of feature selection, 1 feature was selected in pre-Tx radiomics for both 24 patients and 19 patients, 6 features were selected in post-Tx radiomics, and no features were selected in delta radiomic features. There is no model for delta radiomics, since there is no delta radiomics selected for p53 mutation detection. Selected features and their corresponding correlation coefficients and p values are shown in Table 3. (Spearman Coefficient is indicated by r)

Table 3: Selected Features for p53 Mutation Detection in First Feature Selection

Pre-Tx (24 patients)	Pre-Tx (19 patients)	Post-Tx (19 patients)	Delta (19 patients)
Kurtosis (r = 0.44, p = 0.03)	**Variation of Area** (r = 0.39, p = 0.1)	**Skewness** (r = 0.56, p = 0.01), **Inverse Variance** (r = -0.43, p = 0.07), **Long Run Emphasis** (r = 0.49, p = 0.03), **Run Length Non Uniformity** (r = -0.39, p = 0.1), **Run Percentage** (r = -0.45, p = 0.05), **Large Size Low Grey Level Emphasis** (r = 0.43, p = 0.07)	**None**

Selected features from Spearman Correlation were used for second step feature selection and then classification. Since feature selection and classification were all cross validation, the selected features were also different based on specific condition. Selected features with their frequency being selected is shown in Table 4. The ROC curve performance of p53 mutation prediction is shown in Figure 13, Figure 14 and Figure 15. Table 5 shows corresponding AUC values, and the highest value is indicated with underline.

Table 4: Selected Features for p53 Mutation Detection in Second Feature Selection

Pre-Tx (24 patients)	Pre-Tx (19 patients)	Post-Tx (19 patients)	Delta (19 patients)
Kurtosis (24 times)	**Variation of Area** (17 times)	**Skewness** (19 times), **Inverse Variance** (7 times), **Long Run Emphasis** (17 times), **Large Size Low Grey Level Emphasis** (2 times)	**None**

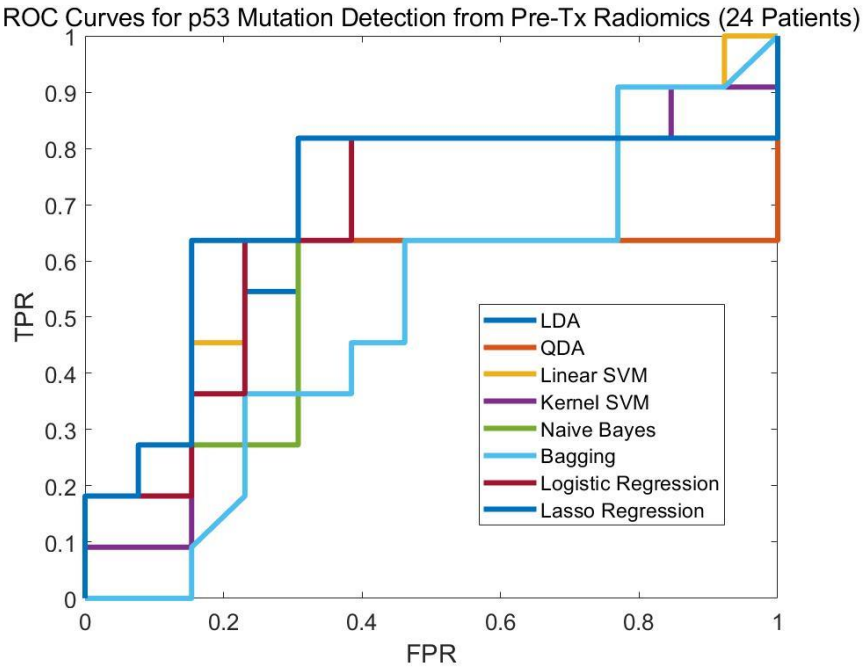

Figure 13: ROC Curves of p53 Mutation Detection with Pre-Tx Radiomics (24 Patients)

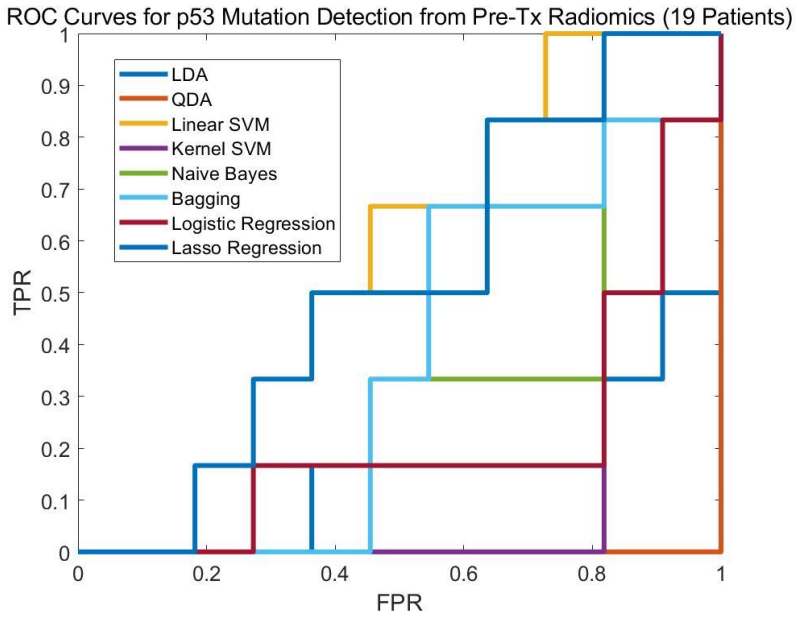

Figure 14: ROC Curves of p53 Mutation Detection with Pre-Tx Radiomics (19 Patients)

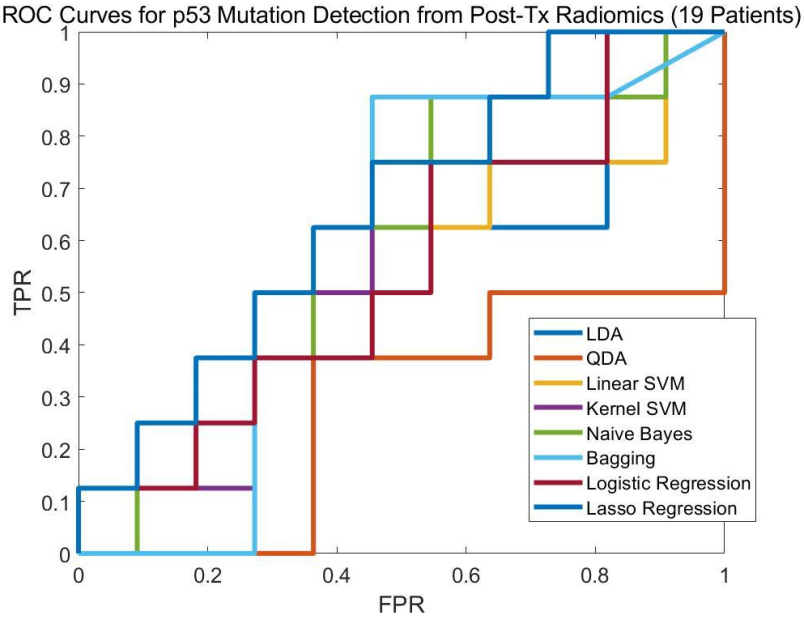

Figure 15: ROC Curves of p53 Mutation Detection with Pre-Tx Radiomics (19 Patients)

Table 5: AUC Comparison of p53 Detection Models

Classification Model	Pre-Tx (24)	Pre-Tx (19)	Post-Tx (19)
LDA	0.5385	0.1515	0.4886
QDA	0.5524	0	0.2841
LSVM	0.6853	0.5606	0.5114
KSVM	0.6503	0.1818	0.5682
NB	0.6364	0.3030	0.6023
Bagging	0.5105	0.3788	0.5739
Logistic Regression	0.6573	0.2121	0.5455
Lasso Regulation	0.6993	0.5152	0.6591

3.2 Patient Survival Prediction

For patient survival prediction, four radiomic models (pre-Tx with 24 or 19 patients, post-Tx, and delta) and four radiogenomic models are built. Radiomic features selected from Cox regression are shown in Table 6. Regression Coefficient is indicated by b. These selected features will then go through the second selection process and classification. 15 pre-Tx features were selected from the dataset of 24 patients, 13 pre-Tx features were selected from the dataset of 19 patients, 2 features were selected from post-Tx radiomics, and 3 features were selected from delta radiomics. Radiomic features selected from the second feature selection with their frequency being selected are shown in Table 7.

Table 6: Selected Features for Patient Survival Prediction in First Feature Selection

Pre-Tx (24 patients)	Pre-Tx (19 patients)	Post-Tx (19 patients)	Delta (19 patients)
Entropy (b = -2.58, p = 0.02), **Info Measure Correlation 1** (b = -0.93, p = 0.03), **Info Measure Correlation 2** (b = 0.69, p = 0.04), **Inverse Variance** (b = -0.68, p = 0.02), **Long Run Emphasis** (b = 0.65, p = 0.02), **Grey Level Non Uniformity (GLRLM)** (b = -1.23, p = 0.05), **Long Run Long Grey Level Emphasis** (b = 0.56, p = 0.04), **Grey Level Non Uniformity (GLSZM)** (b = 0.54, p = 0.02), **Size Zone Non Uniformity** (b = 0.41, p = 0.05), **Size Percentage** (b = 0.41, p = 0.05), **Large Size Low Grey Level Emphasis** (b = 0.35, p = 0.09), **Variation of Intensity**	**Entropy** (b = -2.39, p = 0.06), **Info Measure Correlation 1** (b = -0.96, p = 0.05), **Inverse Variance** (b = -0.81, p = 0.02), **Long Run Emphasis** (b = 0.81, p = 0.02), **Grey Level Non Uniformity (GLRLM)** (b = -1.19, p = 0.06), **Long Run Long Grey Level Emphasis** (b = 0.67, p = 0.04), **Grey Level Non Uniformity (GLSZM)** (b = 0.59, p = 0.02), **Size Zone Non Uniformity** (b = 0.42, p = 0.08), **Size Percentage** (b = 0.43, p = 0.08), **Variation of Intensity** (b = 0.88, p = 0.01) **Variation of Area** (b = 0.42, p = 0.1), **Compactness 1** (b = 0.23, p = 0.1), **Surface Area** (b = 0.45, p = 0.06)	**Compactness 1** (b = 0.45, p = 0.09), **Surface Area** (b = 0.45, p = 0.08)	**Kurtosis**, (b = 0.50, p = 0.05) **Inverse Variance** (b = 0.80, p = 0.04), **Grey Level Non Uniformity (GLRLM)** (b = 0.81, p = 0.02)

(b = 0.87, p = 0.004) **Variation of Area** (b = 0.42, p = 0.05), **Compactness 1** (b = 0.38, p = 0.08), **Surface Area** (b = 0.41, p = 0.05)			

Table 7: Selected Features for Patient Survival Prediction in Second Feature Selection

Pre-Tx (24 patients)	Pre-Tx (19 patients)	Post-Tx (19 patients)	Delta (19 patients)
Entropy (21 times), **Info Measure Correlation 1** (5 times), **Info Measure Correlation 2** (17 times), **Inverse Variance** (8 times), **Long Run Emphasis** (18 times), **Grey Level Non Uniformity (GLRLM)** (3 times), **Variation of Intensity** (23 times)	**Entropy** (16 times), **Info Measure Correlation 1** (2 times), **Inverse Variance** (6 times), **Long Run Emphasis** (17 times), **Grey Level Non Uniformity (GLRLM)** (17 times), **Long Run Long Grey Level Emphasis** (1 time), **Variation of Intensity** (18 times), **Compactness 1** (1 time)	**Compactness 1** (10 times), **Surface Area** (13 times)	**Kurtosis**, (11 times) **Inverse Variance** (12 times), **Grey Level Non Uniformity (GLRLM)** (2 times)

3.3 Pre-Tx Radiomic Patient Survival Model （24 Patients）

The average ROC curves from 24-time cross validation are shown in Figure 16, and AUC values of each classification method with the maximum value showed with underline are shown in Table 8.

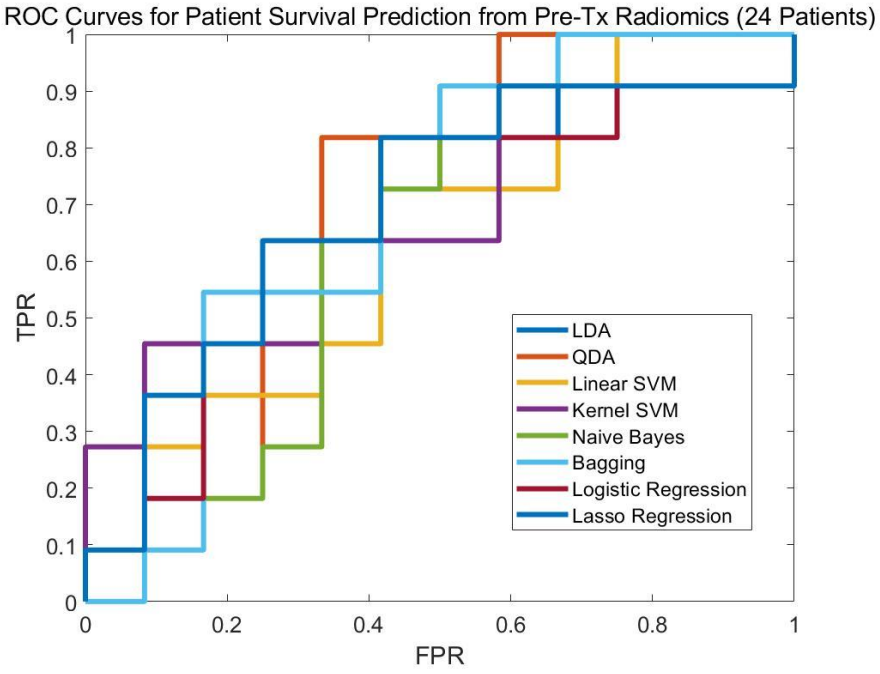

Figure 16: ROC Curves for Patient Survival Prediction (Pre-Tx, 24 Patients)

Table 8: AUC Values for Patient Survival Prediction (Pre-Tx, 24 Patients)

LDA	**0.6439**
QDA	**0.6970**
LSVM	**0.6288**
KSVM	**<u>0.7045</u>**
Naïve Bayes	**0.6439**
Bagging	**0.6970**
Logistic Regression	**0.6667**
Lasso Regression	**0.6970**

3.4 Pre-Tx Radiomic Patient Survival Model （19 Patients）

19 patients used here are the same patients as those used in post-Tx radiomics or delta radiomics. The average ROC curves from 24-time cross validation are shown in Figure 17, and AUC values of each classification method with the maximum value showed with underline are shown in Table 9.

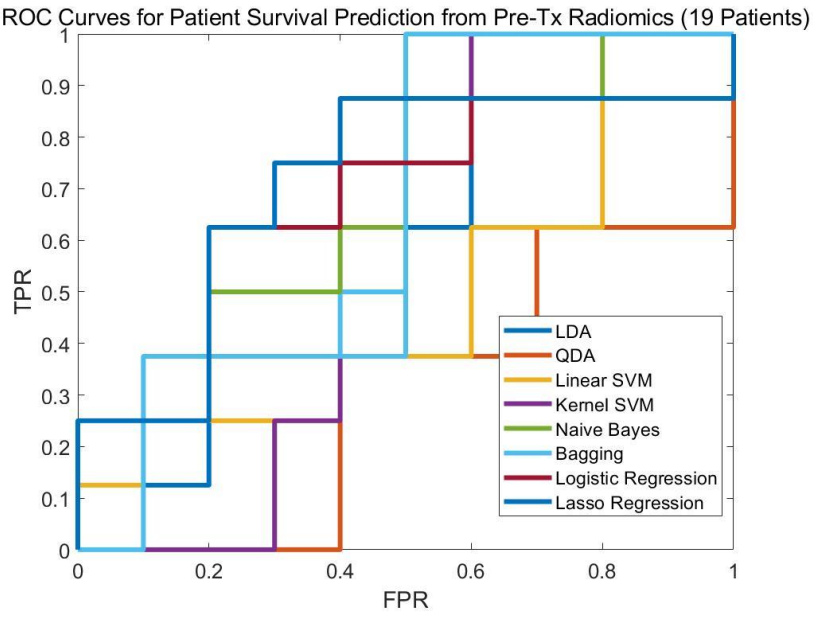

Figure 17: ROC Curves for Patient Survival Prediction (Pre-Tx, 19 Patients)

Table 9: AUC Values for Patient Survival Prediction (Pre-Tx, 19 Patients)

LDA	**0.6500**
QDA	**0.3000**
LSVM	**0.4625**
KSVM	**0.5375**
Naïve Bayes	**0.6875**
Bagging	**0.6500**
Logistic Regression	**0.6750**
Lasso Regression	**0.7125**

3.5 Post-Tx Radiomic Patient Survival Model （19 Patients）

Similar to pre-Tx radiomics, we can implement feature selection and classification methods on post-Tx radiomic features. The average ROC curves are shown in Figure 18, and AUC values are shown in Table 10.

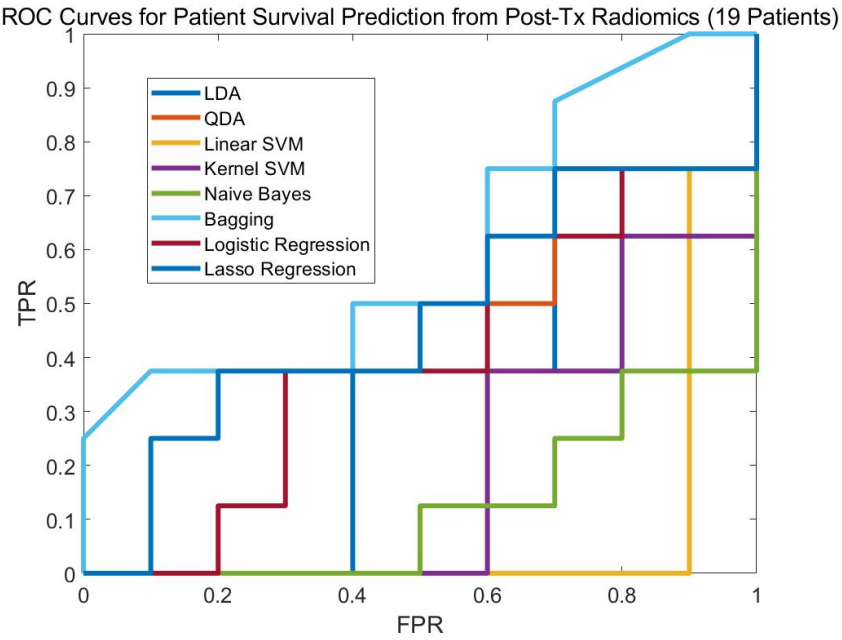

Figure 18: ROC Curves for Patient Survival Prediction (Post-Tx, 19 Patients)

Table 10: AUC Values for Patient Survival Prediction (Post-Tx, 19 Patients)

LDA	**0.3125**
QDA	**0.2625**
LSVM	**0.0750**
KSVM	**0.2000**
Naïve Bayes	**0.1250**
Bagging	**0.6063**
Logistic Regression	**0.4000**
Lasso Regression	**0.4750**

3.6 Delta Radiomic Patient Survival Models (19 Patients)

Delta radiomic features also went through similar feature selection and classification methods. The average ROC curves are shown in Figure 19, and AUC values are shown in Table 11.

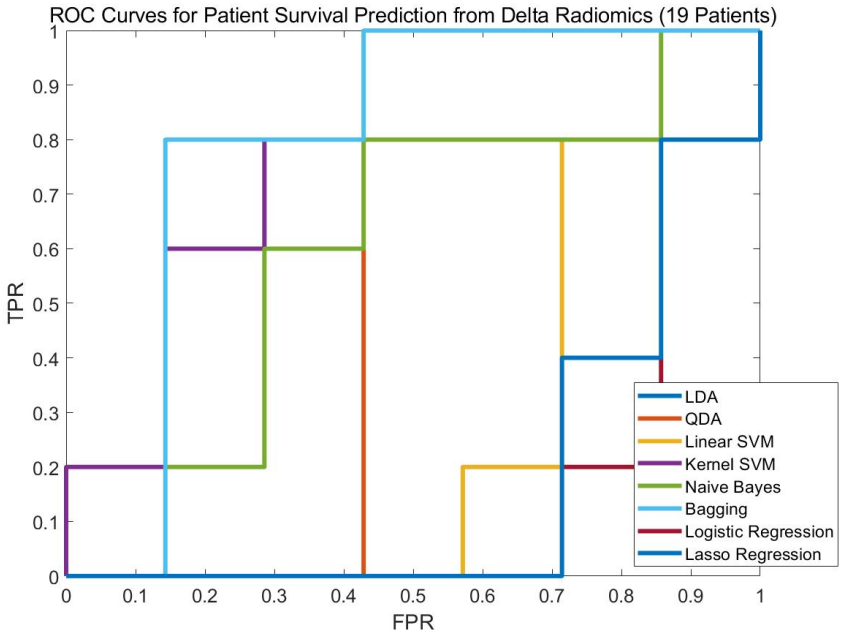

Figure 19: ROC Curves for Patient Survival Prediction (Delta, 19 Patients)

Table 11: AUC Values for Patient Survival Prediction (Delta, 19 Patients)

LDA	**0.1429**
QDA	**0.4571**
LSVM	**0.2857**
KSVM	**0.8000**
Naïve Bayes	**0.6000**
Bagging	**0.8000**
Logistic Regression	**0.1429**
Lasso Regression	**0.1714**

3.7 Genomic Patient Survival Model

The genomic patient survival model only contains one feature, which is p53 mutation. p53 mutation is also used for patient survival prediction for both 24 patients and 19 patients. Since it is a binary feature, the ROC curves are relatively simple. Corresponding ROC curves are shown in Figure 20 and Figure 21, and AUC values are shown in Table 12.

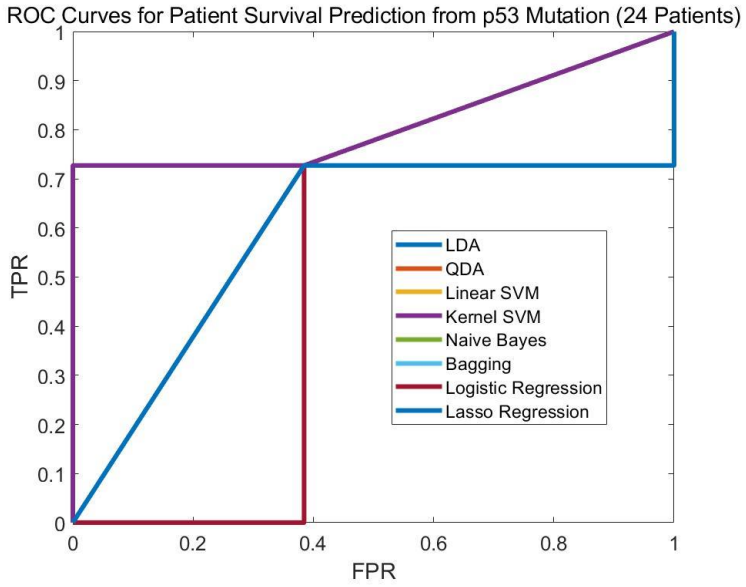

Figure 20: ROC Curves for Patient Survival Prediction from p53 mutation (24 Patients)

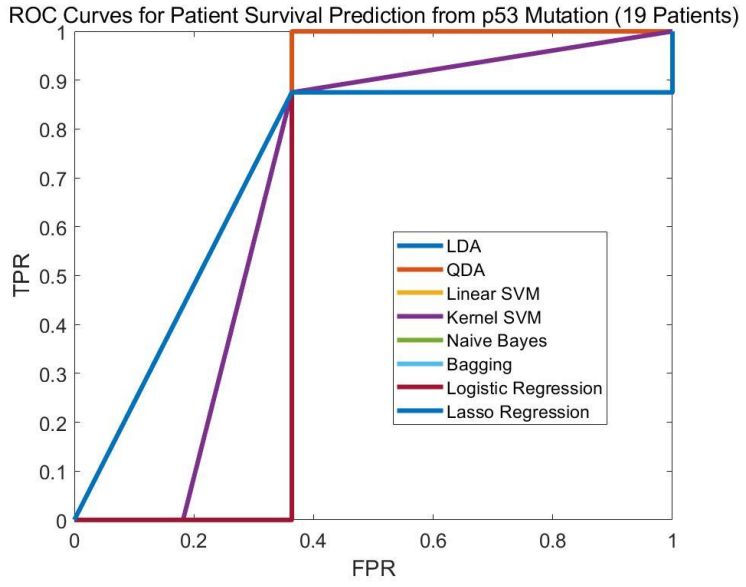

Figure 21: ROC Curves for Patient Survival Prediction from p53 mutation (19 Patients)

Table 12: AUC Values for Patient Survival Prediction (p53 Mutation)

Classification	24 Patients	19 Patients
LDA	0.4476	0.5568
QDA	0.4476	0.6364
LSVM	0.6713	0.7557
KSVM	0.8112	0.6761
Naïve Bayes	0.4476	0.5568
Bagging	0.4476	0.5568
Logistic Regression	0.4476	0.5568
Lasso Regression	0.5874	0.7159

3.8 Pre-Tx Radiogenomic Patient Survival Models (24 Patients)

In this section, p53 mutation is used as a biomarker combined with radiomic features to make a radiogenomic model. There are four radiogenomic models built, which are based on the combination of pre-Tx radiomics and p53 mutation (24 patients and 19 patients respectively), post-Tx radiomics and p53 mutation, delta radiomics and p53 mutations. Figure 22 showed corresponding ROC curves of pre-Tx radiogenomic models (24 patients), and Table 13 showed AUC values as well as corresponding AUC values of radiomic models as comparison.

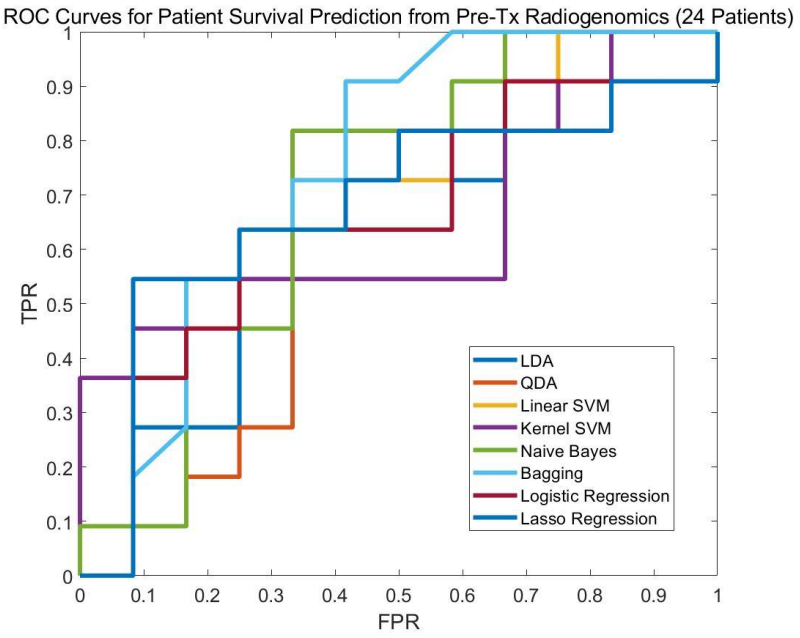

Figure 22: ROC Curves of Pre-Tx Radiogenomics (24 Patients)

Table 13: AUC of Pre-Tx Radiogenomic Model (24 Patients)

Models	Pre-Tx Radiomics	Pre-Tx Radiogenomics
LDA	0.6439	0.6439
QDA	0.6970	0.6136
LSVM	0.6288	0.6894
KSVM	<u>0.7045</u>	0.6515
Naïve Bayes	0.6439	0.7045
Bagging	0.6970	<u>0.7500</u>
Logistic Regression	0.6667	0.6515
Lasso Regression	0.6970	0.6818

3.9 Pre-Tx Radiogenomic Patient Survival Models (19 Patients)

Similarly, ROC curves of pre-Tx radiogenomic models (19 patients) are shown in Figure 23, and Table 14 showed AUC values.

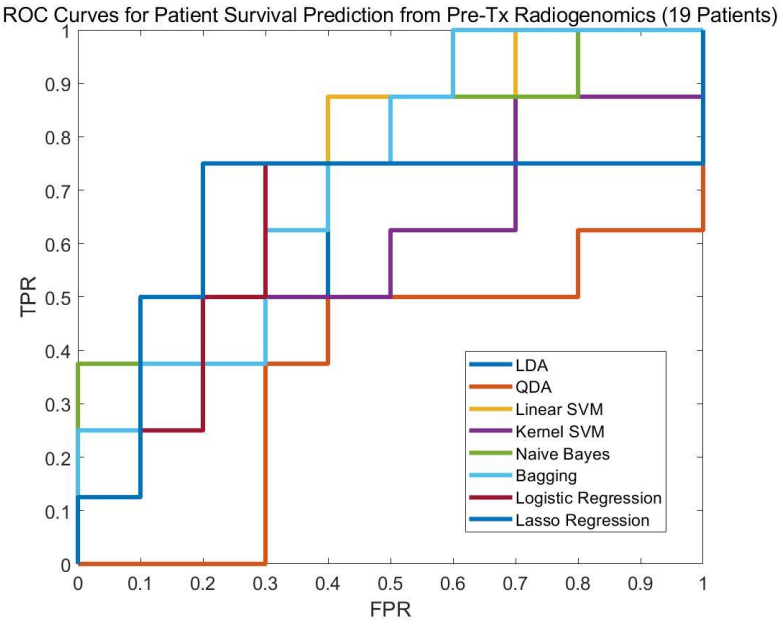

Figure 23: ROC Curves of Pre-Tx Radiogenomic Models (19 Patients)

Table 14: AUC of Pre-Tx Radiogenomic Model (19 Patients)

Models	Pre-Tx Radiomics	Pre-Tx Radiogenomics
LDA	0.6500	0.6500
QDA	0.3000	0.3625
LSVM	0.4625	<u>0.7375</u>
KSVM	0.5375	0.6000
Naïve Bayes	0.6875	<u>0.7375</u>
Bagging	0.6500	0.7250
Logistic Regression	0.6750	0.6125
Lasso Regression	<u>0.7125</u>	0.6625

3.10 Post-Tx Radiogenomic Patient Survival Model

Figure 24 showed ROC curves of Post-Tx Radiogenomic Models, and Table 15 showed corresponding AUC values.

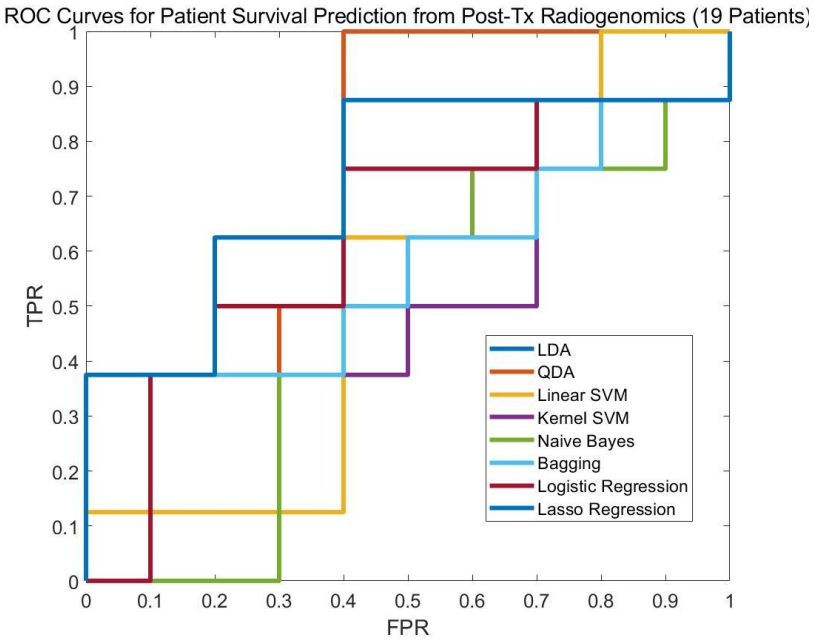

Figure 24: ROC Curves of Post-Tx Radiogenomics

Table 15: AUC of Post-Tx Radiogenomic Models

Models	Post-Tx Radiomics	Post-Tx Radiogenomics
LDA	0.3125	0.4429
QDA	0.2625	<u>0.5857</u>
LSVM	0.0750	0.4000
KSVM	0.2000	0.5571
Naïve Bayes	0.1250	0.4714
Bagging	<u>0.6063</u>	0.5643
Logistic Regression	0.4000	0.4143
Lasso Regression	0.4750	0.4571

3.11 Delta Radiogenomic Patient Survival Model

Similarly, Figure 25 showed ROC curves of delta radiogenomic models and Table 16 showed corresponding AUC values.

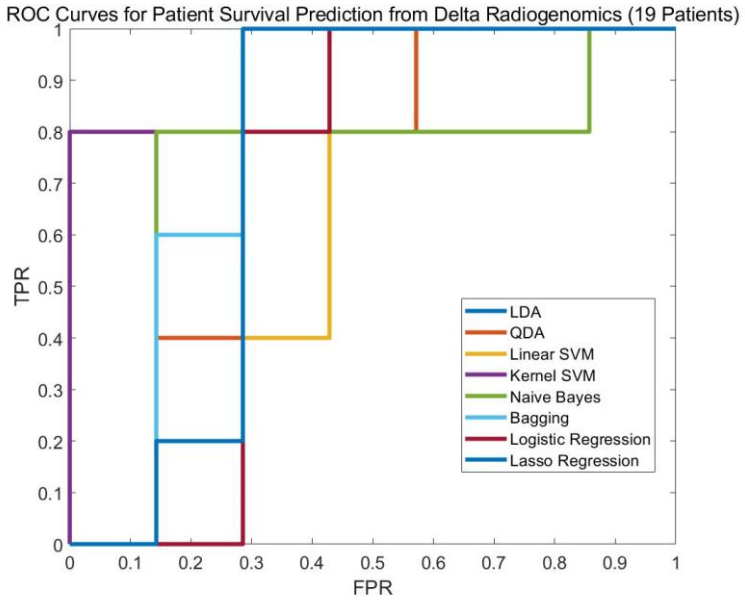

Figure 25: ROC Curves of Delta Radiogenomics

Table 16: AUC of Delta Radiogenomic Models

Models	Delta Radiomics	Delta Radiogenomics
LDA	0.1429	0.7143
QDA	0.4571	0.7143
LSVM	0.2857	0.6286
KSVM	0.8000	0.9143
Naïve Bayes	0.6000	0.7143
Bagging	0.8000	0.8143
Logistic Regression	0.1429	0.6857
Lasso Regression	0.1714	0.7429

3.12 Comparison of Different Models

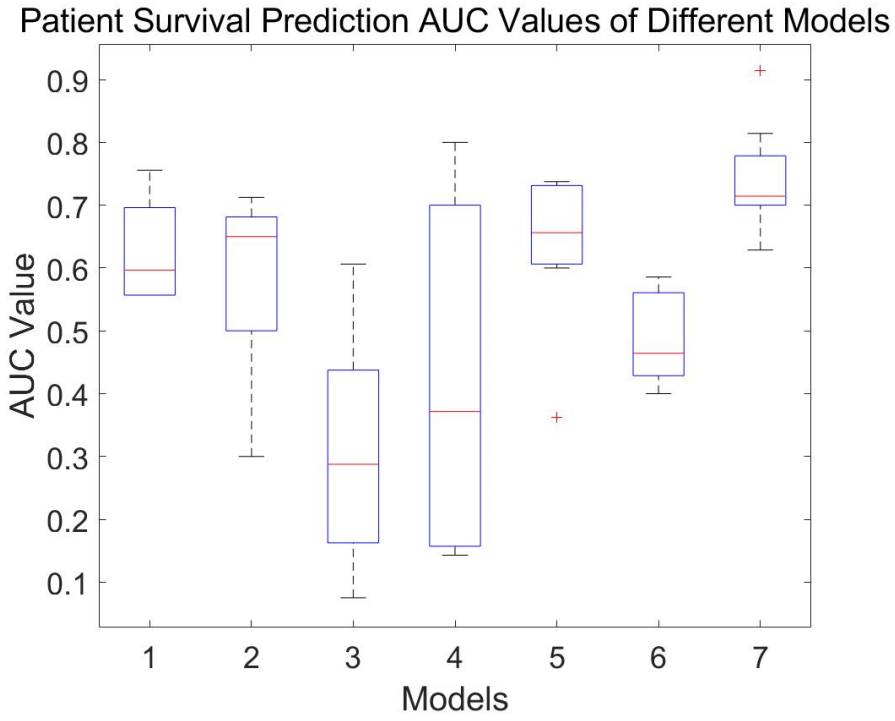

Figure 26 Comparison of Different Patient Survival Prediction Models (19 Patients)

Figure 26 has shown the AUC value comparison of different patient survival prediction models. Model 1 is patient survival prediction from p53 mutation. Models 2 to 4 are patient survival prediction from Pre-Tx, Post-Tx, and delta radiomics, and Models 5 to 7 are patient survival prediction from Pre-Tx, Post-Tx, and delta radiogenomics (combination of radiomics and p53 mutation). All of them only used the 19 patients who had post-Tx contours for comparison.

The comparison result is shown with a boxplot, in which the red line represents the median AUC value of all 8 classification models, the upper line and lower line of the blue box represents the 25th and 75th percentage of all the AUC values, the dashed line

indicates the most extreme points that are not considered outliers, and the red crosses indicate outliers. It is obvious that the addition of p53 mutation in radiogenomic models has a much better performance than original radiomic models or genomic models.

4. Discussion

4.1 p53 Mutation Detection

Since leave-one-out cross validation is used, the selected features in each models may be different. Here all the selected features that have a selected frequency larger than half of the dataset are shown in Table 17.

Table 17: Highly Frequent Features in p53 Mutation Detection

Pre-Tx (24 Patients)	Pre-Tx (19 Patients)	Post-Tx (19 Patients)	Delta (19 patients)
Kurtosis (24 times)	**Variation of Area (17 times)**	**Skewness (19 times)**	**None**
		Long Run Emphasis (17 times)	

In general, 1 feature in pre-Tx radiomics (24 patients), 1 feature in pre-Tx radiomics (19 patients), 2 features in post-Tx radiomics, and 0 features from delta radiomics showed high frequency in being selected for p53 mutation detection. Actually, after the first step of feature selection, there is only one feature left for pre-Tx radiomics (24 and 19 patients). Also, no delta features passed the first round of feature selection.

Kurtosis is selected in pre-Tx radiomics (24 patients). The calculation of kurtosis is shown in Equation 16. N_v is the number of voxels in ROI, μ is mean of intensity

histogram, and X_d is set of discretized grey levels. Kurtosis, first introduced in 1929 (Pearson, 1929), indicates the impact of the tail to the whole intensity histogram of image. It is one of the intensity related features.

$$Kurtosis = \frac{\frac{1}{N_v}\sum_{k=1}^{N_v}(X_{d,j} - \mu)^4}{(\frac{1}{N_v}\sum_{k=1}^{N_v}(X_{d,j} - \mu)^2)^2}$$

Equation 16: Kurtosis Calculation

Variation of Area is selected in pre-Tx radiomics (19 patients), it is also known as zone size variance. It is a texture feature which belongs to Grey Level Size Zone Matrix Features (Thibault, Angulo and Meyer, 2014). The calculation of Variation of Area is shown in Equation 17. i and j are number of grey levels and zone sizes, $p_{ij} = s_{ij}/N_s$, $\mu = \sum_{i=1}^{N_g}\sum_{j=1}^{N_z} jp_{ij}$,$s_{ij}$ is element in the matrix N_z is the size limit of size zone, N_g is the maximal grey level in size zone matrix, and N_s is the total number of elements in grey level size zone matrix.

$$F_{szm.zs.var} = \sum_{i=1}^{N_g}\sum_{j=1}^{N_z}(j - \mu)^2 p_{ij}$$

Equation 17: Calculation of Variation of Area

Skewness is one of the two features showing high frequency in post-radiomics for p53 mutation detection. The calculation of skewness is shown in Equation 18 (Pearson, 1929). N_v is the number of voxels in ROI, μ is mean of intensity histogram, and

X_d is set of discretized grey levels. As an intensity related feature, skewness measures the asymmetry of a histogram.

$$Skewness = \frac{\frac{1}{N_v}\sum_{k=1}^{N_v}(X_{d,j} - \mu)^3}{(\frac{1}{N_v}\sum_{k=1}^{N_v}(X_{d,j} - \mu)^2)^{3/2}}$$

Equation 18: Calculation of Skewness

Long Run Emphasis is a texture feature which belongs to Grey Level Run Length Matrix (Galloway, 1975). It is another feature showing high frequency in post-Tx radiomics. N_s is the sum over all elements in the matrix, N_r is the maximal possible run length along direction, j is run length, and $r_{.j}$ is marginal sum of the runs over the grey values i for run length j.

$$F_{rlm,lre} = \frac{1}{N_s}\sum_{j=1}^{N_r} j^2 r_{.j}$$

Equation 19: Calculation of Long Run Emphasis

If we recall the content of Table 4, we can observe that the performance of pre-Tx (24 patients) is much better than pre-Tx (19 patients). It is reasonable since 24 patients is a larger dataset and an abundant dataset is one of the key for machine learning. Also, selected feature (Kurtosis) in pre-Tx radiomics (24 patients) has a lower than the selected feature (Variation of Area) of pre-Tx radiomics (19 patients), which means that we have more confidence to believe that Kurtosis can be used for p53 mutation detection.

One of the possible reasons that no delta radiomic features were selected may be that delta radiomic features show the change of radiomic features due to the treatment,

but tumor genetic mutation may have existed before the treatment. Therefore, it is reasonable that no delta radiomics is selected for p53 mutation detection.

No morphological features showed high frequency in feature selection for p53 mutation detection, possibly implying that p53 mutation may not correlate with the morphology of tumor. Also, intensity features showed high frequency of being selected in both pre-Tx radiomics (24 patients) and post-Tx radiomics, implying the potential relationship between p53 mutation and intensity histogram of tumor.

After the correlation analysis, we found that all the frequently selected features showed positive correlation with p53 mutation, possibly because of the promotion of tumor heterogeneity due to p53 mutation.

4.2 Patient Survival Prediction

For patient survival prediction, the frequently selected features (Being selected for more than half of the dataset) are shown in Table 18.

Table 18: High Frequently Selected Features in Patient Survival Prediction

Pre-Tx (24 Patients)	Pre-Tx (19 Patients)	Post-Tx (19 patients)	Delta (19 patients)
Entropy (21 times)	**Entropy (16 times)**	**Compactness 1 (10 times)**	**Kurtosis (11 times)**
Info Measure Correlation 2 (17 times)	**Long Run Emphasis (17 times)**	**Surface Area (13 times)**	**Inverse Variation (12 times)**
Long Run Emphasis (18 times)	**Grey Level Non Uniformity (GLRLM) (17 times)**		

Variation of Intensity (23 times)	Variation of Intensity (18 times)		

In leave-one-out cross validation, the selected feature may be different due to the difference of features. However, those features which really has impact on patient survival were selected multiple times. Therefore, here we showed all the frequently selected features.

In pre-Tx radiomics, the most frequently selected features included one intensity feature and three texture features. However, one of the frequently selected features was different for 24 patients and 19 patients. Two morphological features showed high frequency in post-Tx radiomics. One intensity feature and one texture feature showed high frequency in delta radiomics.

Entropy as a highly frequent feature, here referred to Shannon entropy, which indicates the complexity of information one histogram contains (Shannon, 1948). The formula for entropy calculation is shown in Equation 20. p_i is the Co-occurrence probability, where $p_i = \frac{n_i}{N_v}$, N_g is number of grey level bins, N_v is number of voxels in ROI, and n_i is frequency count of each grey level.

$$F_{ih,entropy} = -\sum_{j=1}^{N_g} p_i log_2 p_i$$

Equation 20: Calculation of Entropy

Info Measure Correlation 2 is a texture feature which belongs to Grey Level Co-Occurrence Matrix Features (Haralick, Shanmugam and Dinstein, 1974). It is defined as Equation 21. N_g is the number of grey levels, i, j are grey levels, and p_{ij} are elements in co-occurrence probability matrix.

$$F_{cm.info.correlation.2} = \sqrt{1 - \exp(-2(HXY_2 - HXY))}$$
$$HXY = -\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p_{ij} log_2 p_{ij}$$
$$HXY_2 = -\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p_{i.}p_{.j} log_2 (p_{i.}p_{.j})$$

Equation 21: Calculation of Info Measure Correlation 2

Variation of Intensity is another texture feature showing high frequency in pre-Tx radiomics (24 patients). Also known as grey level variance, it belongs to Grey Level Size Zone Matrix Features (Galloway, 1975). The calculation of variation of intensity is shown in Equation 22. i and j are number of grey levels and zone sizes, $p_{ij} = s_{ij}/N_s$, $\mu = \sum_{i=1}^{N_g}\sum_{j=1}^{N_z} jp_{ij}, s_{ij}$ is element in the matrix N_z is the size limit of size zone, N_g is the maximal grey level in size zone matrix, and N_s is the total number of elements in grey level size zone matrix.

$$F_{szm.gl.var} = \sum_{i=1}^{N_g}\sum_{j=1}^{N_z} (i - \mu)^2 p_{ij}$$

Equation 22: Calculation of Variation of Intensity

Grey Level Non Uniformity is a texture feature showing high frequency in pre-Tx radiomics (19 patients). It is defined in Equation 23 (Galloway, 1975). It describes the

distribution of run lengths. N_s is the total number of elements, N_g is the maximal number of grey levels, $r_{i.}$ is the marginal sum of one grey level.

$$F_{rlm,glnu} = \frac{1}{N_s}\sum_{q}^{N_g} r_{i.}^2$$

Equation 23: Calculation of Grey Level Non Uniformity

Surface Area is a morphological feature showing high frequency in post-Tx radiomics, it describes the total surface area of volume of ROI. The calculation of surface area depends on meshing algorithms.

Compactness I is another morphological feature showing high frequency in post-Tx radiomics. It describes the how compact the ROI is, as shown in Equation 24, in which, V is the volume, and A is surface Area.

$$F_{morph,comp.1} = \frac{V}{\pi^{1/2}A^{3/2}}$$

Equation 24: Calculation of Compactness 1

Inverse Variation is a texture feature showing high frequency in delta radiomics, as shown in Equation 25 (Haralick, Shanmugam and Dinstein, 1973). i, j are grey levels, N_g is the maximal number of grey levels, p_{ij} is the co-occurrence probability matrix element.

$$F_{cm.inv.var} = 2\sum_{i=1}^{N_g}\sum_{j>i}^{N_g}\frac{p_{ij}}{(i-j)^2}$$

Equation 25: Calculation of Inverse Variation

Long Run Emphasis and Kurtosis has already been introduced in Equation 19 and Equation 13.

In this study, four pre-Tx radiomic features (both 24 and 19 patients), two post-Tx radiomic features, and two delta radiomic features showed high frequency for being selected for patient survival prediction. The reason that frequently selected features were different between 24 patients and 19 patients in pre-Tx radiomics is due to the limited data. However, still three of four highly selected features were same for two datasets. Two morphological features showed high frequency in post-Tx radiomics, probably due to the morphology of tumor was directly affected by radiation. Highly frequent selected delta radiomic features showed the change of which feature has impact on patient survival. One intensity feature (Kurtosis) and one texture feature (Inverse Variation) showed high frequency. It might indicate the potential relation between these features and patient survival.

If we compare pre-Tx, post-Tx and delta radiomics for the 19 patients, pre-Tx radiomics on average showed the highest performance. One of the reasons is due to the various dates of post-Tx CT, which may have affected the performance of post-Tx radiomics and delta radiomics. However, the combination of delta radiomics and p53 mutation showed a better result than all other models alone, as shown in Figure 26.

4.3 Relation Between p53 Mutation and Radiomics

As shown in Figure 26, compared with other patient survival prediction models, the combination of delta radiomic features and p53 mutation showed best performance of patient survival prediction. One of the possible reason is the combination of p53 mutation and delta radiomics contains the information of the change of tumor during treatment as well as genetic mutation of tumor. Also, the result may be related to the fact that p53 mutation can increase the radioresistance of tumor cells. (Cuddihy and Bristow, 2004) It means that even tumors receive same dose or show same impact in radiomic features after the treatment, their long-term response may be different due to different genetic mutations.

This result may shed light to future assessment of radiotherapy that the combination of delta radiomics and genetic mutations of tumor can perform a more accurate outcome of patients than traditional radiomics.

4.4 Limitation

The original patient dataset is relatively small, (24 patients for pre-Tx radiomics, and 19 patients for post-Tx radiomics). The follow-up post-Tx dates were variable which may affect post-Tx images. Also, the location of ROI may not be exact the same location as pre-Tx ROI since the location of tumor may has changed after treatment. Also, the genetic mutation of tumor was not testified by biopsy.

5. Conclusion and Future Work

5.1 Conclusion

From limited dataset, it might be feasible to detect p53 mutation by both pre-Tx and post-Tx radiomics. Lasso and LSVM have shown the best performance in classification.

We predicted patient survival with pre-Tx, post-Tx, and delta radiomics, as well as radiogenomics. We noted that different features were selected for predicting the patient overall survival. This may be related to the limited data available for the study. It may also be related to the different characteristics of pre-Tx, post-Tx and delta radiomics. For patient survival, intensity and texture features showed high frequency of being selected for pre-Tx and delta features, and morphological features showed high frequency for post-Tx radiomics. However, we also found that the combination of delta radiomics and p53 mutation showed a better patient survival prediction than pre-Tx, post-Tx, delta radiomics and p53 mutation alone. The reason might be related to the difference of tumor reaction to radiation due to p53 mutation. KSVM and Bagging showed highest performance compared with other classification methods.

5.2 Future Work

Larger dataset of patients should be applied in further study for validation of study and a more generalized conclusion.

p53 has shown its ability to increase the patient survival prediction in radiogenomics. Besides p53, there are also other genes that can be potential biomarkers in radiogenomics. Following research will focus more on other genes, especially various oncogenes and tumor suppressor genes.

So far the modelling hasn't considered clinical factors or dosimetric information, which may also have important impact on patient survival. The combination of radiogenomic models and clinical factors or dosimetric features is another focus of future work.

Also, deep learning may be applied in future study, especially for automatic feature extraction for specific goals.

References

Aerts, Hugo J.W.L., et al. "Decoding tumor phenotype by noninvasive imaging using a quantitative radiomics approach." *nature communications* 7 Aug 2014.

—. "Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach." *Nature Communication* 3 June 2014: 1-8.

Ahrendt, Steven A., et al. "p53 mutations and survival in stage I non-small-cell lung cancer: results of a prospective study." *Journal of the National Cancer Institute* 2 July 2003: 961-970.

American Cancer Society. *https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2018/cancer-facts-and-figures-2018.pdf*. n.d.

Breiman, Leo. "Bagging Predictors." *Machine Learning* 1996: 123-140.

Chang, Yushi. "An Investigation of Machine Learning Methods for Delta-Radiomic Feature Analysis." *Duke Kunshan University Master Degree Thesis* 2018: 1-48.

Cho, Y, et al. "Crystal structure of a p53 tumor suppressor-DNA complex: understanding tumorigenic mutations." *Science* 15 July 1994: 346-355.

Coroller, Thibaud P., et al. "CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma." *Radiotherapy and Oncology* March 2015: 345-350.

Corradetti, Michael N., et al. "Dynamic Changes in Circulating Tumor DNA During Chemoradiation for Locally Advanced Lung Cancer." 2017.

Cox, David R. "Regression Models and Life Tables (with discussion)." *Journal of the Royal Statistical Society* 1972: 187-220.

Cuddihy, A. R. and R. G. Bristow. "The p53 protein family and radiation sensitivity: Yes or no?" *Cancer Metastasis Review* August 2004 : 237-257.

De Perrot, Thomas, et al. "Differentiating kidney stones from phleboliths in unenhanced low-dose computed tomography using radiomics and machine learning." *Imaging Informatics and Artificial Intelligence* 11 January 2019: 1-7.

Fave, Xenia, et al. "Delta-radiomics features for the prediction of patient outcomes in non-small cell cancer." *Scientific Reports* 3 April 2017.

Fund, World Cancer Research. *https://www.wcrf.org/dietandcancer/cancer-trends/lung-cancer-statistics*. 2018.

Galloway, M. M. "Texture analysis using grey level run lengths." *Computer Graphics and Image Processing* (1975).

Green, D. M. and J. A. Swets. *Signal detection theory and psychophysics.* 1966 .

Grossmann, Patrick, et al. "Defining the biological basis of radiomic phenotypes in lung cancer." *eLife* 21 July 2017.

Haralick, R. M., K. Shanmugam and I. Dinstein. "Texture Features for Image Classification." *IEEE Transactions on Systems, Man, and Cybernetics* 1973.

Harris, Sandra L. and Arnold J. Levine. "The p53 pathway: positive and negative feedback loops." *Oncogene* 18 April 2005: 2899-2908.

Hawkins, Samuel , et al. "Predicting malignant nodules from screening CTs." *Journal of Thoracic Oncology* December 2016.

Huret, Jean-Loup . *Haematology, Atlas of Genetics and Cytogenetics in Oncology and Haematology*. August 2014.

Kwon, Deukwoo, et al. "Classification of suspicious lesions on prostate multiparametric MRI using machine learning." *Journal of Medical Imaging, 5(3)* 2018.

Li, Hui, et al. "MR Imaging Radiomics Signatures for Predciting the Risk of Breast Cancer Recurrence as Given by Research Versions of MammaPrint, Oncotype DX, and PAM50 Gene Assays." *Radiology* November 2016: 382-391.

Li, Yiming, et al. "MRI features predict p53 status in lower-grade gliomas via a machine-learning approach." *NeuroImage:Clinical* 29 Oct 2017.

Liang, Cuishan, et al. "The development and validation of a CT-based radiomics signature for the preoperative discrimination of stage I-II and stage III-IV colorectal cancer." *Oncotarget* 22 April 2016: 31401-31411.

Mogi, Akira and Hiroyuki Kuwano. "TP53 Mutations in Nonsmall Cell Lung Cancer." *Journal of Biomedicine & Biotechnology* 18 January 2011.

National Cancer Institute. "Surveillance, Epidemiology and End Results Program." 2016.

National Human Genome Research Institute. *Genome.gov*. n.d.

Oxford English Dictionary. 2019.

Parmar, Chintan, et al. "Machine Learning methods for Quantitative Radiomic Biomarkers." *Scientific Reports* 17 August 2015: 1-11.

Pearson, K. "Editorial note to 'Inequalities for moments of frequency functions and for various statistical constants." *Biometrika* (1929).

Pinker, Katja, et al. "Background, Current Role, and Potential Applications of Radiogenomics." *Journal of Magnetic Resonance Imaging* March 2018: 604-620.

Saha, Ashirbani, et al. "A study of association of Oncotype DX recurrence score with DCE-MRI characteristics using multivariate machine learning models." *Cancer Research* 23 January 2018: 799-807.

Santosa, Fadil and William W. Symes. "Linear inversion of band-limited reflection seismograms." *SIAM Journal on Scientific and Statistical Computing* 1986: 1307-1330.

Shannon, Claude E. "A Mathematical Theory of Communication." *Bell System Technical Journal* (1948).

Sherr, Charles J. "Principles of Tumor Suppression." *Cell* 23 January 2004: 235-246.

Thibault, G., J. Angulo and F. Meyer. "Advanced statistical matrices for texture characterization." *IEEE transactions on bio-medical engineering* (2014).

Tibshirani, Robert. "Regression Shirinkage and Selection via the lasso." *Journal of the Royal Statistical Society* 1996: 267-288.

—. "The Lasso Method for Variable Selection in the Cox Model." *Statistics in Medicine* December 1997: 385-395.

Weinstein , I Bernard and Andrew K. Joe. "Mechanism of Disease: oncogene addiction--a rationale for molecular targeting in cancer therapy." *Nature Review* 28 April 2006: 448-457.

Wu, Shuang, et al. "Radiomics-based machine learning methods for isocitrate dehydrogenase genotype prediction of diffuse gliomas." *Journal of Cancer Research and Clinical Oncology* 4 February 2019.

Zinn, Pascal O., et al. "Distinct Radiomic Phenotypes Define Glioblastoma TP53-PTEN-EGFR Mutational Landscape." *Neurosurgery* 11 May 2017.

Zwanenburg, Alex, et al. "Image biomarker standardisation initiative." *arxiv.org* 17 Sep 2018.