MULTIMODAL IMAGE CLASSIFIERS FOR PROGNOSIS AND TREATMENT RESPONSE PREDICTION FOR LUNG PATHOLOGIES

by

PRANJAL VAIDYA

Submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

Department of Biomedical Engineering
CASE WESTERN RESERVE UNIVERSITY

August, 2022

CASE WESTERN RESERVE UNIVERSITY

SCHOOL OF GRADUATE STUDIES

We hereby approve the dissertation of

Pranjal Vaidya

candidate for the degree of Doctor of Philosophy [1]

Committee chair

Dr. Pallavi Tiwari

Committee member, advisor

Dr. Anant Madabhushi

Committee member

Dr. David Wilson

Committee member

Dr. Amit Gupta

Committee member

Dr. Prantesh Jain

Date of defense

July 13 , 2022

[1]We also certify that written approval has been obtained for any proprietary material contained therein

Contents

List of Tables

List of Figures

Acknowledgments

Among the many people who have been very critical of this work, I would specifically like to thank my advisor Dr. Anant Madabhushi, my committee, and my colleagues at the Center for Computational Imaging and Personalized Diagnostics. I wouldn't be where I am today without your guidance, encouragement, and support. I would like to thank our collaborators from Cleveland Clinic and University Hospitals, specifically Dr. Amit Gupta and Dr. Pradnya Patil, for their invaluable inputs in all my critical projects. I would like to thank my ever-supportive family- my parents and sister, without whom I wouldn't have been able to start my journey in the first place and my husband, who has been my constant support throughout these years and has been my rock through all my highs- and lows.

Lastly, I would like to acknowledge those funding resources which have supported me in my studies. This work is made possible by the National Cancer Institute, National Heart, Lung, and Blood Institute 1R01HL15127701A1, R01HL15807101A1, National Institute of Biomedical Imaging and Bioengineering 1R43EB028736-01, National Center for Research Resources under award number 1 C06 RR12463-01, VA Merit Review Award IBX004121A from the United States Department of Veterans Affairs Biomedical Laboratory Research and Development Service the Office of the Assistant Secretary of Defense for Health Affairs, through the Breast Cancer Research Program (W81XWH-19-1-0668), the Prostate Cancer Research Program (W81XWH-15-1-0558, W81XWH-20-1-0851), the Lung Cancer Research Program (W81XWH-18-1-0440, W81XWH-20-1-0595), the Peer-Reviewed Cancer Research Program (W81XWH-18-1-0404, W81XWH-21-1-0345), the Kidney Preci-

sion Medicine Project (KPMP) Glue Grant, DoD Prostate Cancer Research Program Idea Development Award W81XWH-18-1-0524, Clinical and Translational Science Collaborative (CTSC) Cleveland Annual Pilot Award 2020 UL1TR002548 and sponsored research agreements from Bristol Myers-Squibb, Boehringer-Ingelheim, Eli-Lilly, and AstraZeneca.

Abstract

Multimodal Image Classifiers for Prognosis and Treatment Response Prediction for Lung Pathologies

by

Pranjal Vaidya

Non-small cell lung cancer tumors follow an orderly progression from adenocarcinoma in situ (AIS) to minimally invasive carcinoma (MIA) and invasive adenocarcinoma (INV). Currently, there is no definite biomarker to access the level of invasion and detect invasive disease in these early lepidic lesions using radiographic scans, which would ideally help in surgery planning for these patients. Within the early-stage NSCLC cohort, while all the patients will receive the surgery, a significant portion of patients (up to 50%) will develop recurrence. Although most of these patients are eligible to receive adjuvant chemotherapy (chemo), not all patients will receive the added benefits. In the more advanced NSCLC setting, immunotherapy (IO) has shown promising survival improvement, but only a fraction (20%) of patients will respond to IO, and a fraction of patients (8%) would, in fact, receive adverse effects of it, and cancer would spread rapidly (hyperprogression). Most of the current AI methods developed in this field are based on a single modality. However, information across different modalities and scales may hold complementary information, and integrating

them together may enhance the performance of AI models. In addition, most of the developed AI models lack interpretability, an essential element for successfully transitioning these AI methods into clinical practices.

In this dissertation, we introduced new interpretable AI biomarkers that use textural patterns on radiographic scans, known as Radiomics, and combine these biomarkers across multiple modalities and scales for NSCLC and COVID-19 patients. The Radiomic features were analyzed from inside the tumor region as well as from the area immediately surrounding the nodule. Furthermore, we integrated the clinical features into Radiomics Model by using novel techniques. We also created a human-machine integrated model using Radiologists' scores combined with Radiomic Analysis. Lastly, we used pathology data to create radiology-pathology fusion models and pathology information, along with the Radiogenomic analysis to understand the biological interpretability of the Radiomic Features.

In this dissertation, we have looked at 4 specific use cases (1) predicting the level of invasion in low-risk nodules, (2) predicting the risk of post-surgical recurrence, and subsequently added benefit of adjuvant chemo for early-stage patients, and (3) predicting patients' response to IO. In addition, we developed tools for predicting the COVID-19 patients at severe risk who would end up using mechanical ventilation. The development and validation of these approaches were performed on data from about 2500 patients across 9 different sites.

Chapter 1

Introduction and Previous Work

1.1 Overview

With the advances in technology, it has become easier to gather multi-modal information for each patient, including both multi-scale imaging and non-imaging modalities. One of the major challenges to overcome during the integration of heterogeneous data types (which is difficult to directly combine) lies in the fact that there exist significant differences in scale and dimensionality between modalities. Artificial intelligence (AI) approaches are increasingly being used for diagnosis, prognosis, and treatment decisions of various diseases. While most of the current AI methods are developed in a single modality of data, fusion approaches integrating data from multiple sources, modalities, and scales can provide further complementary information and help in enhancing the performance of these approaches.

Radiomics refers to looking at the textural patterns on Radiographic Scans, which look at pixel-level patterns that are not very evident to the human eye. Radiomics has shown diagnostic, prognostic, and predictive applications in various lung cancer space problems. The tumor region and the region surrounding the nodule have shown promising results. Although there has been significant progress within the overall Radiomics Space, the interpretability of these radiomic features hasn't been explored extensively, which is an essential parameter for

successfully transitioning these features into routine clinical practices. In addition, there has been comparatively limited work on identifying the integration of these Radiomic Features with clinical parameters.

With the digitization of tissue scans, it has become easier to look at computerized patterns of tissue scans. These patterns show cell-level interactions, and pathomics refers to looking at computerized patterns of these tissue scans. Pathomics has shown diagnostic, prognostic, and predictive performance in lung cancer. Even though these imaging modalities (radiography and histology) have individually shown promising results, the combination approaches for these techniques have been limited. The previous work by Feng. Et. Al. and Yang et al. show that different modalities of images at different scales may provide complementary information within the rectal cancer space. But these approaches haven't been explored within the lung cancer space even though almost all lung cancer patients have access to these multi-imaging modality datasets. These multi-imaging modalities also provide a unique opportunity to correlate these images with each other for interpretable AI analysis. In addition, a fusion of clinical variables into these AI models has further proven to improve the model accuracies since these are the routine clinical variables used by clinicians on a regular basis to make decisions regarding diseases. Unfortunately, there's a lack of validated imaging-based biomarkers taking advantage of multimodality information and at the same time providing interpretable AI. In this dissertation, we have developed imaging-based biomarkers for various lung cancer domain problems that take multi-modal information and also provide the biological interpretability of the AI models.

1.2 Background

1.2.1 Non-Small Cell Lung Cancer Space

Lung cancer is the leading cause of cancer mortality in the United States, accounting for more than a quarter of all cancer deaths for both men and women1. Non-small cell lung

cancer accounts for 85% of total lung cancer cases. Depending on the stages of the tumor, non-small cell lung cancer (NSCLC), patients will have different treatment regimens. These tumors follow an orderly progression from the adenocarcinoma in situ (AIS) to minimally invasive (MIA) and finally invasive adenocarcinoma (INV). The timing and nature of surgical intervention for semisolid abnormalities are dependent upon distinguishing between adenocarcinoma-in-situ (AIS), minimally invasive adenocarcinoma (MIA), and invasive adenocarcinoma (INV). Currently, there is no definite biomarker to access the level of invasion and detect invasive disease in these early lepidic lesions using radiographic scans, which will potentially help in their surgery planning. When the tumor progresses to a relatively early stage, they form a subset of NSCLC known as early-stage (stage 1A to 2B cases). While all early-stage NSCLC will receive the surgery, a significant portion of patients (up to 50%) will develop recurrence. Although most of these patients are eligible to receive adjuvant chemotherapy (chemo), not all patients will receive its added benefit of it. Separately, in the more advanced NSCLC setting, immunotherapy (IO) has shown promising survival improvement, but only a fraction (20%) of patients will respond to IO, and a fraction of patients (8%) would in fact receive an adverse effect of it, and cancer would spread rapidly (hyperprogression). There is a paucity of validated prognostic, predictive biomarkers, and companion diagnostic tools to predict the (1) level of invasion in low-risk nodules, (2) risk of post-surgical recurrence, and subsequently added benefit of adjuvant chemo, and (3) response to IO.

1.2.2 The Coronavirus Disease of 2019

The novel SARS-Cov2 virus or Coronavirus 19 (COVID-19) pandemic has led to widespread deaths due to respiratory complications. While the confirmatory test currently implemented is RT-PCR-based assays, early evidence suggests that the sensitivity of the test might be wanting, especially for early detection of the disease when the patient is mainly asymptomatic. While the FDA has approved both molecular and serological tests to diagnose the

virus, several challenges still exist in using these tests as the gold standard for diagnosing the virus. With multiple tests being approved by the FDA, there is a relative lack of standardization amongst the tests. Moreover, with point-of-care testing, the validity and accuracy of the tests are also dependent on extraneous factors such as specimen handling, how the swab was taken, and the expertise of the test administrator, among others.

While there is evidence that COVID-19 is a multisystem disorder, the disease has primarily been shown to affect the lungs and cause respiratory symptoms leading to Acute Respiratory distress syndrome (ARDS) in the severe phenotype. Chest radiographs and chest CT scans which allow for efficient visualization of the lungs have been shown to be an effective screening tool for suspected COVID-19 patients even those in the early asymptomatic stage. Previous data have shown the presence of lung changes in CT scans in some COVID-19 + patients who initially tested negative for the virus on RT-PCR. CT findings in patients with mild COVID-19 have reportedly resembled ground glass-like opacities in bilateral lungs usually distributed peripherally, while a more consolidative pattern akin to community-acquired pneumonia (CAP) was seen in critical patients in the ICU. While these visual findings might suggest the presence of COVID-19, the relative lack of specificity of these findings means that it is currently difficult to differentiate it from other pneumonia from imaging alone (12). This has led to guidelines not recommending routine use of CT scans as a screening tool but recommending its application in COVID-19 positive patients for treatment assessment and monitoring.

1.3 Novel Contribution

The work presented here is based on radiomics implementation within various applications. We have used radiomic features from various regions of interest to go beyond the traditional radiomic analysis. We have used different multi-scale datasets and integration techniques to create combined models. We have used survival analysis and various supervised and unsu-

pervised techniques based on the application that we are targeting to understand the desired population more accurately. And finally, we have attempted to understand the biological meaning of the CT-based features by correlating them with pathology and radiogenomic data with it.

1.3.1 Radiomics AI approaches: challenges and novel contribution

The majority of AI tools within the lung pathologies domain have been focused on using deep learning-based techniques for disease diagnosis, prognosis, and predicting treatment responses. But these techniques work on the black-box-based approaches where the underlying interpretability of the model remains limited. Radiomics is known as one of the interpretable AI approaches. Radiomic texture features are defined as computing higher-order statistical patterns of image voxels within a local region or measuring response to a filter designed to capture patterns of interest. Generally, there are two types of features extracted in radiomics: "semantic" and "agnostic" features. Semantic features are used to describe regions of interest, while agnostic features are used to capture lesion heterogeneity through quantitative descriptors. For example, heterogeneity that is a symptom of malignant nodules reflects the areas of high cell density, necrosis, and myxoid change and can be captured by radiomic texture features. Various Radiomic-Based textural features have shown diagnostic, prognostic, and predictive ability within the non-small cell lung cancer space. These include features from Gabor feature family, Laws, Laplace, Haralick and CoLlAGe Feature families.

Most of the previous work within the radiomic within NSCLC space has been focused on looking at the tumoral region on the CT scans. Recently, there has been an interest in looking beyond the tumor region and at the region immediately surrounding the nodule. In this work, Radiomics was analyzed from the tumor region and from the region outside the nodule in a novel annular peritumoral ring-shaped fashion. Each ring of 3mm distance, leading up to 15 mm radially outside the nodule, was analyzed.

One of the major challenges while analyzing radiomics on CT scans is that these radiomics

hand-crafted features get affected by the variations within the raw parameters of CT scans. Acquisition parameters not only vary across different sites but also within the same site, as protocol standardization is clinically challenging. Differences in slice thickness and reconstruction algorithm can affect the produced image quality. Additionally, the amount of radiation dose has an effect on CT appearance. Noise within the image increases with the reduction of dose, which further affects the performance of disease detection. Furthermore, parameters such as automated tube current and operator present noise index can affect the final appearance of CT images. While developing the models, we made sure to standardize these parameters as much as possible to produce reproductive results.

1.3.2 Multi Scale Integration: challenges and novel contribution

In this dissertation, another major contribution revolves around the multi-scale integration approaches. We have integrated information from radiographic scans, histopathological scans, and clinical information to improve the model performances and subsequently make it easier to transition into routine clinical practices.

In this dissertation, we have explored multiple data integration methods. We have developed a novel nomogram approach that visually shows the relationship between various features (imaging and non-imaging) integrated within the model. We have analyzed the datasets within supervised and non-supervised approaches to see the effects of various parameters on each other. A few unsupervised techniques used across multiple projects include – K-means clustering, hierarchical clustering, and consensus clustering. These parameters were also correlated with each other to find the association between them.

The major challenge in multi-scale data integration lies in the difference between scales and dimensionality between these modalities. While developing and implementing the multi-modal fusion models, we require access to matched pathology, radiology, and clinical datasets. Sometimes when the patients have multiple nodules, correlating this information with each other gets difficult. There are multiple ways in which this multi-modal information could be

fused together. Some examples of fusion strategies include the fusion of classifier probabilities, the fusion of features extracted from multiple modalities, etc. Choosing an appropriate fusion strategy that would be generalizable across multiple datasets is an essential step in this analysis. Another challenge lies in ensuring that we are selecting complementary information for the analysis and removing the redundant information without losing the critical information from these datasets. Within the pathomic analysis, the quality of histopathology scans affects the feature analysis. Most existing slide scanners have a maximum capability to scan at $\times 40$. Higher resolution images (¿$\times 20$) can be scaled down to be used by an algorithm trained at a resolution of $\times 20$, but the use of an AI approach developed at $\times 40$ when the maximum scanning resolution available is $\times 20$ would likely result in a loss of data fidelity. In addition, the creation of accurate, manually annotated reference datasets by expert pathologists is an essential step in order to standardizing the evaluation of the performance of AI algorithms. For clinical features, the major challenge lies in analyzing the missing values and categorical variables. Multiple data imputation methods could be implemented for analyzing the missing values, and choosing an appropriate method becomes a challenge. Prior to clinical adoption, AI-based and ML-based tools need to be sufficiently validated using multi-institutional data in order to ensure the generalizability of the approaches. One of the critical reasons for attempting to validate AI approaches using separate test sets independently is to ensure that these approaches are resilient to pre-analytical sources of variation. In our analysis, all models were independently validated on external validation sets to ensure the generalizability of the models.

1.3.3 Survival Analysis: challenges and novel contribution

We have extensively explored survival analysis techniques in this dissertation. Some of the clinical problems that we were targeting in this dissertation were handled using continuous survival data rather than binary analysis. The feature selection and model construction were performed using the disease-free survival information and the censoring information to better

represent the patient population. Specifically, within the early-stage NSCLC cohort, we constructed a model using disease-free survival and were able to identify the population that specifically would benefit from adjuvant chemotherapy. To the best of our knowledge, one of the first attempts was to create a predictive model for early-stage cases. We implemented hazard ratios, Kaplan-Meier curves, and cox-proportional hazard models in the analysis.

One of the major challenges in the survival analysis lies in getting the correct survival information and understanding the censored population. In addition, selecting an appropriate threshold to differentiate the high-risk population from the low-risk population remains challenging. There are several methods through which you can select an appropriate threshold, but selecting the method that would be reproducible and generalizable is important. We have developed a technique to select appropriate thresholds based on the hazard ratios in this work.

1.4 Major Goals of dissertation

This dissertation includes several inter-and intra- modal fusion strategies within the lung space. Within the non-small cell lung cancer (NSCLC) space, we looked at minimally invasive adenocarcinoma, early-stage NSCLC cases as well as late-stage NSCLC cases. In addition to cancer, we also looked at the Coronavirus Disease of 2019 (COVID-19) prognosis. The developed experiments demonstrate the utility of different fusion strategies as well as to creates interpretable AI models: a) combining radiomics with the radiologists' scores for predicting minimally invasive adenocarcinoma from invasive adenocarcinoma, b) Creating a CT-based signature for predicting response to adjuvant-chemotherapy for early-stage NSLCLC and understanding biological underpinning of these radiomic features using histopathology scans, c) Using Radiomics on CT scans for predicting hyperperprogressive disease for late-stage NSCLC, d) Combining imaging features and non-imaging features for predicting COVID-19 disease prognosis e) Combining Radiomic and Pathomic Features for

predicting recurrence in Early-Stage NSCLC and predicting immunotherapy response in late-stage NSCLC

1.4.1 Combining radiomics with the radiologists' scores for predicting minimally invasive adenocarcinoma from invasive adenocarcinoma

There have been previous attempts at identifying the level of invasion using radiomic features, but most of them focus on radiomic textural analysis solely within the tumor (20, 21). The peritumoral microenvironment has emerged as a promising candidate location for identifying the level of invasion, although it has been relatively unexplored (22). In addition to using the novel Radiomic Features, to the best of our knowledge, no previous work has created a human-machine integrated AI-based model for predicting the level of invasion. In this experiment, we took input from two radiologists in a blinded fashion and integrated their scores into our radiomics-based model. Additionally, we also divided patients into different subgroups based on the diameter of the nodule and evaluated classifier performance within nodules with different sizes.

1.4.2 Creating a CT-based signature for predicting response to adjuvant-chemotherapy for early-stage NSLCLC and understanding the biological underpinning of these radiomic features using histopathology scans

Within the Early-Stage domain, the current developed AI models have been focused on predicting recurrence for these patients. To the best of our knowledge, none of the models have specifically looked at predicting the added benefit of adjuvant chemotherapy for these patients. Additionally, very few radiomic-based models have looked at the interpretability

of the radiomic features. In this study, we looked at the unique aspect of not only predicting recurrent but also finding the subset of patients directly benefitting from adjuvant chemotherapy. We also used histopathology features and radiogenomic analysis to understand the biological interpretability of these top radiomic features observed on the CT scans.

1.4.3 Using Radiomics on CT scans for predicting hyperperprogressive disease for late-stage NSCLC

Radiomic approaches have shown utility for predicting response to immune checkpoint inhibitors in advanced-stage NSCLC. To the best of our knowledge, there hasn't been enough work within this newly defined category of immune checkpoint inhibitors, known as hyperprogressors, which has a paradoxical acceleration of tumor growth after initiation of immunotherapy. To develop an AI model for predicting these Hyperprogressor patients, we extracted features from inside the nodule and outside the tumor region and also looked at quantitative vessel tortuosity features (QVT) on CT scans representing the curvedness of blood vessels feeding the tumor. The analysis was performed both within supervised and unsupervised approaches, and we looked at survival information to analyze and compare the models.

1.4.4 Combining Radiomic and Pathomic Features for predicting recurrence in Early-Stage NSCLC and predicting immunotherapy response in late-stage NSCLC

Previous studies show the utility of whole-slide tissue-based features for diagnostic, prognostic, and predictive applications within the non-small cell lung cancer domain applications. Especially, within the Early-Stage domain, the spatial architecture of tumor-infiltrating lymphocytes (SPATIL) features have shown superior performance for predicting recurrence, and PhenoTIL features with late-stage NSCLC have shown Good performance. Similarly, on

the radiography front, as explained above, the radiomic features have shown the ability of lung CTs to predict recurrence in Early-Stage NSCLC and predicting response in late-stage NSCLC. However, to the best of our knowledge, no previous studies have combined biomarkers from different scales such as radiology (lung CT) and pathology (HE slides) for early-stage NSCLC recurrence prediction and response to immunotherapy in late-stage NSCLC. Therefore, as part of this objective, we combine biomarkers from both CT and digitized HE slides, and train in ML classifiers to predict recurrence in Early-stage cases and predict response to immunotherapy in late-stage NSCLC.

1.4.5 Combining imaging features and non-imaging features for predicting COVID-19 disease prognosis

Most of the AI analysis within the COVID-19 space has focused on chest x-rays (CXRs) (1, 2), though more recently, more and more works on AI for CT scans have also been published. However, many of the proposed models are poorly reported and are at high risk of bias, and at present, it is not recommended to use any of the reported prediction models for use in clinical practice. To the best of our knowledge, none of the Radiomics-based COVID-19 studies have focused analysis on 1k patients, integrated with clinical parameters, and validated on an independent cohort. Therefore, in this work, we present a novel AI approach combining Radiomic Features with clinical parameters, in the fashion of nomogram for predicting RT-PCR COVID positive patients would have a severe disease phenotype and end up needing invasive mechanical ventilation.

Chapter 2

Combining Radiomics with the Radiologists' Scores for Predicting Minimally Invasive Adenocarcinoma from Invasive Adenocarcinoma

2.1 Overview

Lung cancer is the leading cause of cancer related deaths in the world. Adenocarcinoma is the most common lung cancer histologic type.[1] With the increase in diagnostic imaging methods such as low-dose chest CT screening, there has been an increase in the detection of lung cancers at earlier stages often presenting as small solid/semisolid nodules or ground-glass opacities (GGOs)[2] [3]. The new IASLC guidelines [5 and the AJCC-defined 8th edition staging guidelines,[3] along with the WHO classification of adenocarcinomas,[4] have divided the adenocarcinoma into three broad categories: preinvasive adenocarcinoma [including adenocarcinoma in situ (AIS)], minimally invasive adenocarcinomas (MIA) and invasive adenocarcinoma (INV).[5] Histopathologically, lepidic growth (defined as growth along the alveolar

walls) is a hallmark of non-invasive lesions.[5] An invasive component in the new classification system is defined as either any cellular histologic subtype other than lepidic or invasion of malignant cells into myofibroblastic stroma.[6] Lepidic cancers are observed to follow an orderly progression from the AIS to MIA before becoming INV.[7]

Outcomes of adenocarcinomas following surgical resection are dependent on the initial stage. Resected stage IA non-small cell lung cancer (NSCLC) has a five-year overall survival rate of about 75%.[8] In comparison, the five-year disease-specific survival rate for resected MIA is nearly 100%.[9] The surgical approach and extent of lung resection for these lung nodules can be dictated by the adenocarcinoma histologic subtype.[10] Sublobar resection can produce equivalent results to lobectomy in patients with non- or minimally invasive adenocarcinomas, with the benefit of preservation of lung parenchyma and potential eligibility for repeat resection in the case of subsequent primary tumor.

At present, there are no definite radiographic biomarkers to identify the extent of invasion prior to surgical resection. Although the invasive portion of the cancer is typically solid and non-invasive (lepidic portion) is ground glass in appearance on the CT scan, there is substantial overlap in the imaging findings between different subcategories. Furthermore, traditional CT scan evaluation can be subjective, and interpretations tend to vary widely depending on the experience of the reading radiologist.[11] This coupled with other variables such as scan parameters, slice thickness, etc. limits reliable differentiation on routine radiologic assessment. Fine needle aspiration and imaging is inaccurate in determining the degree of invasion.[12] Hence, there is a critical need to create an accurate model to non-invasively assess the level of invasion on imaging in these early-stage adenocarcinomas prior to surgical resection.

Radiomic textural features represent high-throughput quantitative imaging data extracted from radiographic scans to investigate subtle patterns within a region of interest (ROI).[13] These textural patterns extracted from inside and outside the nodule have been shown to have diagnostic, prognostic, and predictive utility in the lung cancer domain.[14]

These features are known to capture the underlying tumor biology and morphology of the tissue[15].[16] There have been previous attempts at identifying the level of invasion using radiomic features, but most of them focus on radiomic textural analysis solely within the tumor[17].[18] The peritumoral microenvironment has emerged as a promising candidate location for identifying the level of invasion, although it has been relatively unexplored [22].

In this study, we constructed a non-invasive radiographic biomarker based on baseline chest CT scan-guided radiomics to distinguish MIA from INV for stage I NSCLC patients with tumor diameter less than 3 cm. We evaluated these radiomics features via supervised and unsupervised approaches to identify specific patterns associated with INV and MIA nodules. We also divided patients into different subgroups based on the diameter of the nodule and evaluated classifier performance within nodules with different sizes. Finally, we compared our model with the performance of two radiologists and integrated the radiologists' score with the corresponding machine classifier performance to assess combined human and machine classification performance.

2.2 Methods

2.2.1 Study Population

We performed a retrospective, multi-cohort study of patients with resected MIA and stage 1A INV cases. A total of 268 patients from four different institutions were included in the study, all of whom had baseline (pre-treatment) CT scans. Based on our inclusion criteria, we selected a cohort of patients who had tumor size less than or equal to 3 cm with a special focus on a subset of 1 to 2 cm nodules.

The patients were randomly divided into training (D_{Train}=40%) and validation (D_{Test}=60%) cohorts. The D_{Train} was selected to keep the same number of invasive and non-invasive lesions for training the model.

2.2.2 Procedure

CT Segmentation and Radiomic Textural Feature Extraction

The index pulmonary lesions on these baseline CT scans were annotated using a freehand tool on 3D slicer software by an expert radiologist.

After the tumor was annotated, the area of the nodule was calculated using MATLAB 2015. The tumor area was calculated upon identification of the CT slice with the largest tumor region and was used for subgroup analysis and for creating a combined radiomics area-based model.

These annotated nodules were used to extract the intra- and peri-tumoral texture features. The peri-tumoral compartment around the nodule was defined via quantitative morphological operations (dilation) as a region extending radially from the nodule boundary up to roughly 15 mm, since a resection margin larger than 15 mm for lung nodules is considered not to confer additional benefit in terms of invasive lesions. The program was modified to eliminate skin, air, or fat when the mask was extended. Radiomic peritumoral features were extracted in an annular ring-shaped fashion. Five annular rings peritumorally were analyzed, each with 3-mm increments leading up to a maximum radius of 15 mm from the nodule periphery.

Features from Gabor, Laws, Laplace, Haralick and Collage were extracted for the analysis. Haralick and Collage features are based on constructing a gray-level co-occurrence matrix and are known to capture the general disorganized and chaotic microarchitecture of the annotated region of interest[19].[20] The Laws and Laplace features focus on the high-frequency content of the image, focusing on the boundary of the ROI.[21] Gabor features are wavelet-based features.[22]

Classifier Construction

All patients included in the study were divided into two groups: pre-invasive/minimally invasive lesion group (AIS, MIA) and frank invasive group (invasive pulmonary adenocarcinoma [IPA]). These two groups were used as a clinical endpoint for the classification problem.

First, all the radiomic features were analyzed using an unsupervised clustering approach to evaluate the ability of the radiomic features to differentiate the two different diagnostic categories blinded to prior pathology results or clinical outcome. First, the PCA was used on an entire feature pool and the top three principal components were used within K-Means clustering analysis. In addition, the hierarchical clustering was performed on an entire cohort.

Next, a supervised machine learning based logistic regression classifier, MR, was constructed using the top selected features from the training cohort, D_{Train}, and then was validated on an independent and blinded validation set DTest. Further, DTest was divided into 3 different subsets based on the nodule size (less than 1 mm, 1 mm-2 mm, 2 mm-3 mm) and the performance of the model was observed on these various subgroups defined using nodule sizes.

Next, another supervised machine classifier model was constructed using the tumor areas, MA, and further integrated with radiomic features to construct the combined tumor area-radiomics based model (M_{R+A}).

Human Reader Experiment

The patients from D_{Test} were individually assessed by two radiologists with 12 and 21 years of experience, respectively, being blinded to the ground truth pathologic diagnosis of the nodules. The two readers scored each tumor from 1 to 3; 1 suggesting the nodule was MIA, 2 being indeterminate, and 3 being INV. We calculated the accuracy of the radiologists' scores and further compared our radiomics model, M_R, with the results from the radiologists (M_{HR}). Finally, we integrated the probability obtained from the radiomics model, M_R, with the radiologists scoring (1 to 3) to obtain the combined human and machine-based

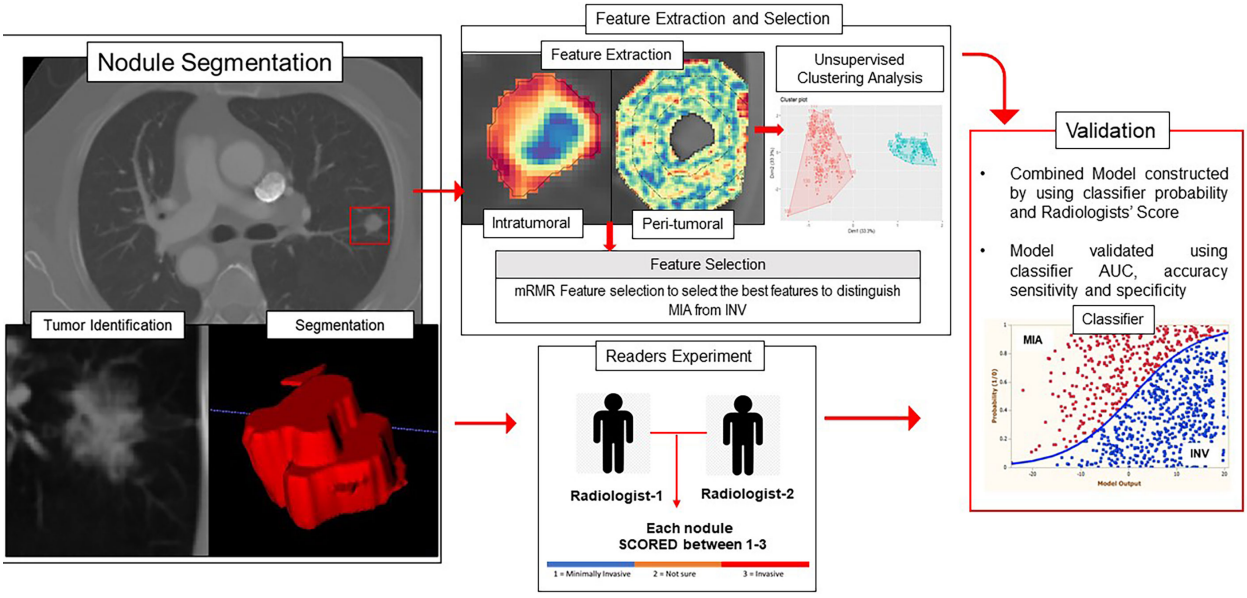

Figure 2.1: Overall workflow diagram. The nodules were segmented on the CT scans, and intratumoral and peritumoral features were extracted using MATLAB 2015. The top features were selected using the mRMR feature selection method. The validation of the radiomics model was performed using unsupervised clustering and supervised classification-based approaches.

interpretations (M_{R+HR}).

2.2.3 Statistical Analysis

Statistical analysis was performed using MATLAB 2015 and R. version 3.5.3. A two-sided p-value (¡0.05) was considered significant for all the statistical analyses.

Looking at the radiomic feature pool, radiomic feature stability and reproducibility were evaluated using the RIDER test-retest dataset.[23] This dataset contains 31 lung cancer patients - scanned two times, 15 min apart. These scans were used for calculating the intraclass correlation coefficient (ICC) for each feature vector, which measures the similarity between two feature vectors. Considering the threshold of 0.85, all feature vectors having a value less than this threshold were removed from the analysis.

Within an unsupervised clustering analysis, hierarchical clustering and principal component analysis (PCA) combined with K-means clustering was performed on D_{Train}. The clustering results were compared against ground truth for calculating the clustering accu-

racy.

For feature selection and building a classifier, 300 iterations of threefold cross-validation were performed within the training dataset, D_{Train}. The minimum redundancy maximum relevance (mRMR) feature selection algorithm[24] was implemented within the cross-validation setting to select the top-performing radiomic features that discriminate INV from MIA/AIS. MRMR identifies a set of features that maximally distinguished two classes while minimizing intra-feature correlation. A maximum of five features was selected to prevent overfitting due to the curse of dimensionality arising from an overabundance of features relative to the sample size. mRMR was performed using MATLAB software with a feature selection toolbox for C. The top radiomic feature set was further analyzed using box-and-whisker plots and qualitative feature maps comparing feature expressions between MIA/AIS and invasive adenocarcinomas.

To evaluate classifier performance, the area under the receiver operating curve (AUC), accuracy, sensitivity, and specificity were calculated for training and validation datasets. The significance of the addition of a nodule area to the radiomic model was calculated using DeLong's test and the corresponding p-value.[25] Figure 2.1 shows the overall pipeline of the procedure.

2.3 Results

2.3.1 Baseline Characteristics

Of the 268 nodules, 103 nodules were pathologically confirmed as pre-invasive lesions (AIS, n = 2) and minimally invasive lesions (MIA, n = 101), whereas 165 were confirmed as invasive lesions (INV = 165). Figure 2.2 shows the datasets and patient inclusion criteria along with training and testing set distributions.

Figure 2.3 shows an example of CT scans with INV and MIA lesions.

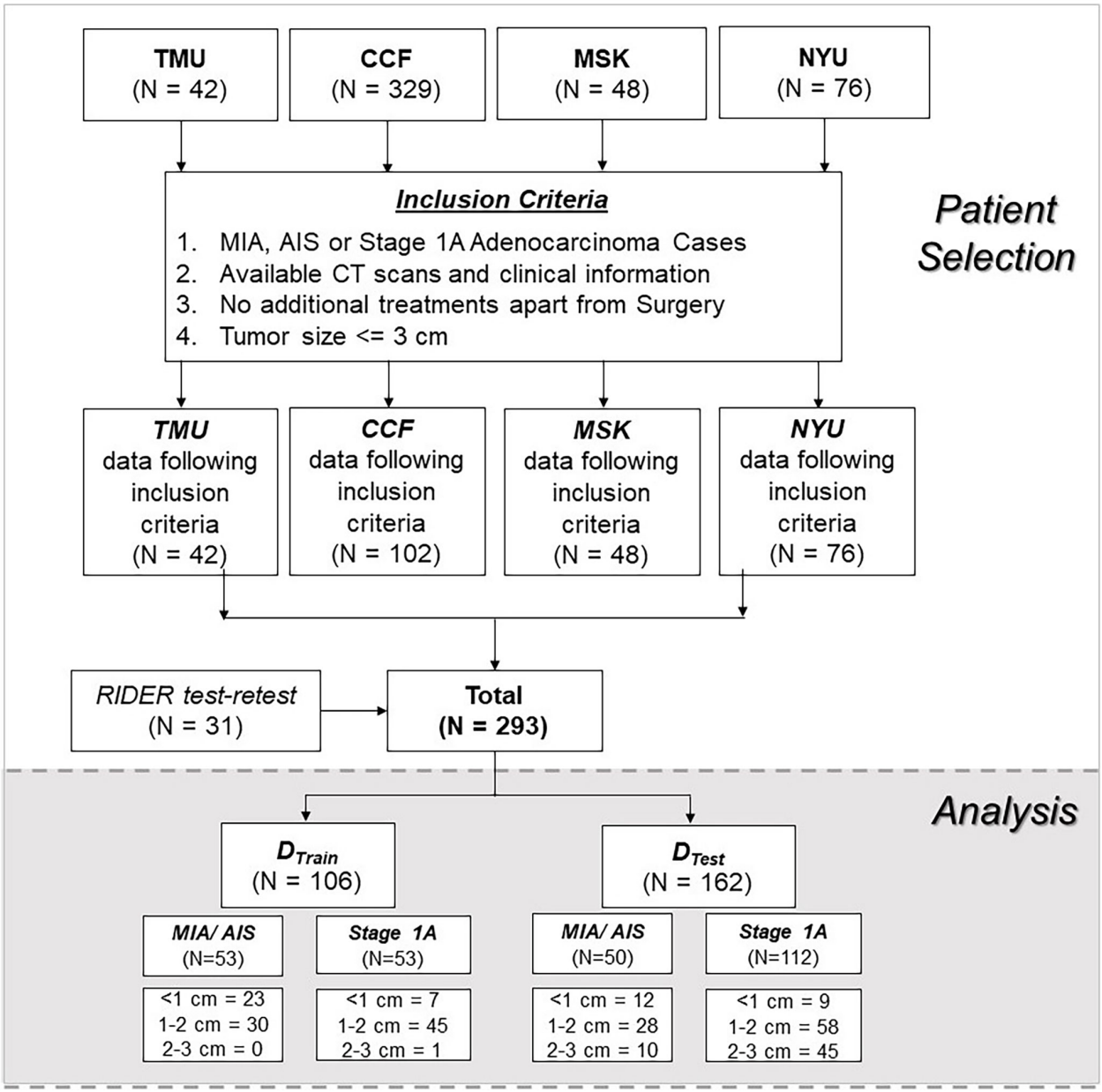

Figure 2.2: Data source and CONSORT diagram for patient selection.

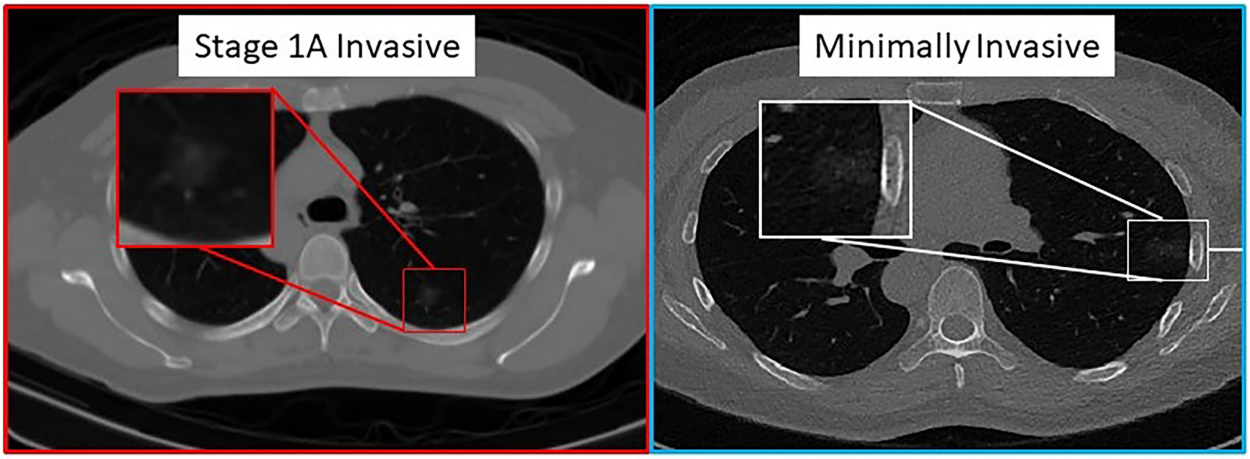

Figure 2.3: Pathologically proven INV (left image) and MIA (right image) cases presenting as predominantly ground-glass nodular densities which are indistinguishable on CT imaging.

2.3.2 Experiment 1 – Differentiating Minimally Invasive Adenocarcinoma From Invasive Adenocarcinoma

Unsupervised Clustering

The extracted radiomic feature pool, that is, the combination of intratumoral textural and peritumoral textural radiomics features, was used within the principal component analysis (PCA) and k-means clustering to perform unsupervised clustering analysis. The optimal number of clusters was two using the first three principal components on DTrain. The constructed clusters had an accuracy of 73.1%. The compactness within the clusters, that is, how similar the members within the same group are, was 62.8%. The validation of the constructed cluster was performed using the silhouette coefficient (silhouette width). The silhouette plot[26] suggests that the clustering using the two groups was optimal with no negative silhouette width and most cluster values > 0.5.

Using the entire extracted radiomic feature pool, within the hierarchical clustering analysis, we observed the 4 obvious clusters of patients. Cluster 1 and Cluster 3 were associated with INV cases (cluster 1 = 100%, cluster 3 = 62.5% INV cases), whereas clusters 2 and 4 were associated with MIA cases (cluster 2 = 71.4%, cluster 4 = 75% MIA cases). The results of unsupervised clustering analysis are shown in Figure 2.4.

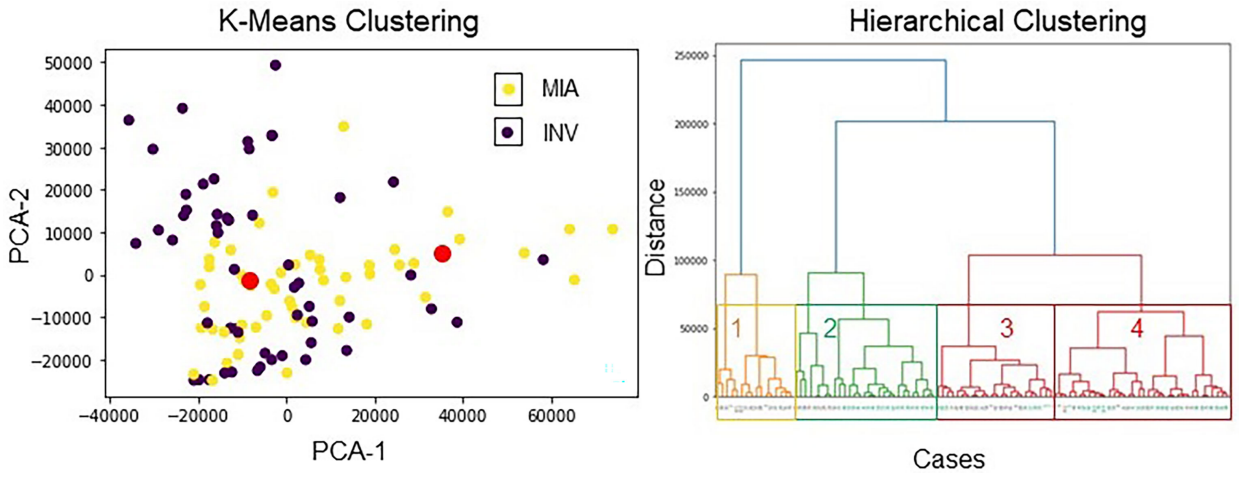

Figure 2.4: Unsupervised clustering analysis using radiomic features. (left image) K-means clustering with 4 clusters. The red dots show the centroids of the three clusters obtained via K-means clustering. The violet points represent INV patients, and the yellow points depict MIA patients. The two distinct clusters had an accuracy of 73.13% to distinguish MIA from INV cases. (right image) Hierarchical clustering using all features. On the x-axis, black color stands for the INV cases, and aquamarine color stands for the MIA cases.

Table 2.1: Model notations.

Model	Notations
Radiomics Model	$\mathbf{M_R}$
Clinical Model	$\mathbf{M_A}$
Radiomics-Clinical Model	$\mathbf{M_{R+A}}$
Human Reader Model	$\mathbf{M_{HR}}$
Integrated Human Reader and Radiomics Model	$\mathbf{M_{HR+R}}$

The unsupervised clustering analysis suggests that the majority of INV adenocarcinoma cases were clustered together, and MIA/AIS patients were clustered together. Collectively, these results suggest that these specific patient groups have distinct radiomic signatures.

Supervised Analysis and Selecting the Top Differentiating Features

During feature discovery for the model M_R within D_{Train}, the top 5 features identified included a peritumoral (CoLlAGe feature family) and 4 intratumoral features (Laws, Laplace, and Haralick feature family). INV cases were observed to have a higher expression of intratumoral features compared to MIA cases. Figure 5 shows the feature expression maps for the INV and MIA cases. The notations of various models constructed using these features are explained

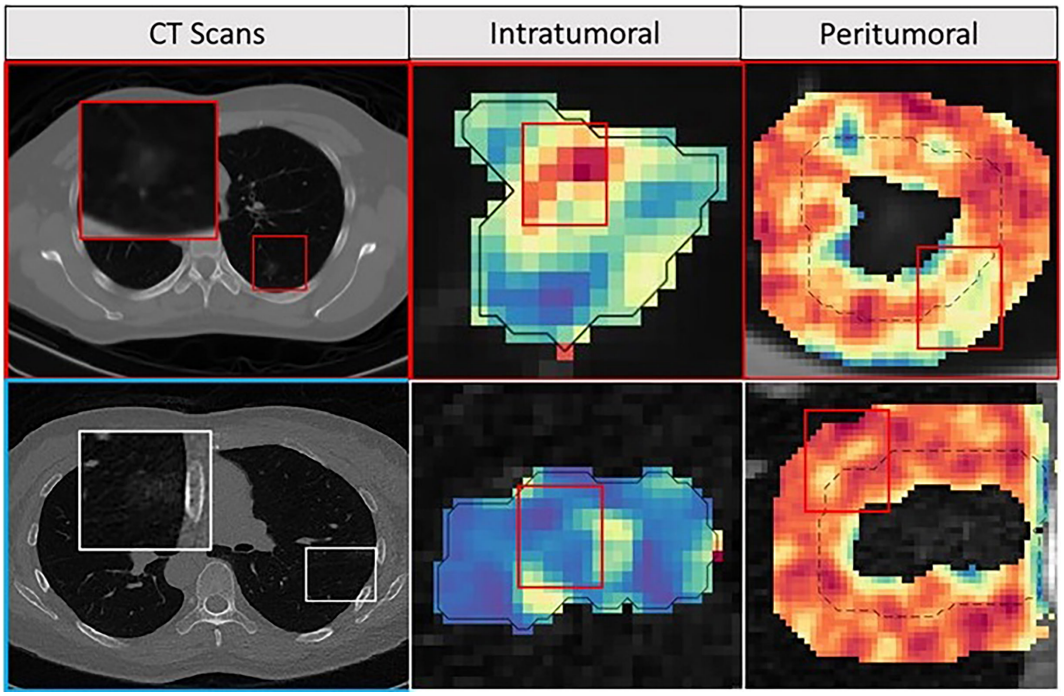

Figure 2.5: Feature maps. The first row depicts INV patient, and the bottom row depicts an MIA patient with an axial CT image as well as corresponding peritumoral and intratumoral feature maps. For the INV case, the feature maps had a higher feature expression compared to the MIA cases suggesting association between chaotic/disturbed microarchitecture and tumor invasiveness.

in Table 2.1.

On the training cohort (D_{Train}, N=106), the logistic regression AUC for MR was 0.917 [0.87-0.97]. The same classifier, within an independent blinded test set (D_{Test}, N=162), M_R yielded an AUC of 0.88 Table 2.2.

Next, within the subgroup analysis, we noticed the radiomic model, M_R, was consistent in distinguishing INV from MIA. Further, M_R is largely unaffected by the size of the nodule (Figure 2.6).

Table 2.2: Model notations.

		No. of Cases	**Area M_A**	**Radiomics M_R**	**Rad+Area of M_{R+A}**	**P wrt Area**
Training			0.73[0.64-0.83]	0.917[0.87-0.97]	0.95[0.916-0.987]	3.013e(-6)
Testing	**All**		0.665	0.862	0.869	1.362e(-5)
	0-1cm	22	0.79	0.759	0.713	0.492
	1-2cm	87	0.61	0.919	0.926	1.057e(-5)
	2-3cm	45	0.57	0.954	0.836	2.136e(-5)

		Radiologist 1 M_{HR1}	Radiologist 2 M_{HR2}	Classifier M_R	Combined M_{R+HR}	P-Value	
						M_R	M_{HR1}
D_{Test} **Entire dataset**	AUC	0.815	0.796	0.861	**0.909**	0.041	4.1 e(-5)
	Accuracy	0.748	0.742	0.788	0.828		
	Sensitivity	0.800	0.792	0.820	0.760		
	Specificity	0.723	0.640	0.772	0.861		
D_{Test} **0-1 cm**	AUC	0.528	0.505	0.759	**0.796**	0.289	0.031
	Accuracy	0.571	0.571	0.667	0.761		
	Sensitivity	0.833	1.00	0.583	0.750		
	Specificity	0.223	0.00	0.778	0.778		
D_{Test} **1-2 cm**	AUC	0.831	0.794	0.916	**0.928**	0.267	0.0015
	Accuracy	0.771	0.747	0.843	0.892		
	Sensitivity	0.857	0.818	0.893	0.893		
	Specificity	0.727	0.607	0.818	0.891		
D_{Test} **2-3 cm**	AUC	0.856	0.898	0.953	**0.963**	0.624	0.025
	Accuracy	0.800	0.844	0.756	0.911		
	Sensitivity	0.861	0.889	1	0.945		
	Specificity	0.556	0.667	0.694	0.778		

Bold numbers represent the AUCs.

Figure 2.6: AUC comparison for logistic regression model trained with radiologists' interpretations, M_{HR}., radiomic features, M_R, and combined radiomic and area-based models, M_{R+HR} on the test set. P-value is calculated to observe the added benefit of M_{HR} in the radiomics model, M_R.

Further, when the area of the nodule was integrated within the logistic regression classifier along with the radiomic features, M_{R+A}, there was no statistically significant improvement in AUC on the validation set as compared to M_R standalone.

2.3.3 Experiment 2 – Comparing the Radiomics Analysis With Readers

We performed the analysis with individual radiologists (M_{HR}) along with the combined performance with the classifier (M_{R+HR}). Reader 1 had an AUC of 0.815 and an accuracy of 0.748 for predicting MIA cases from INV cases, whereas Reader 2 had AUC and accuracy of 0.796 and 0.742, respectively (Figure 2.6).

Within nodules <1 cm size, the classifier demonstrated an improvement over the radiologists' interpretations.

Finally, we combined the classifier predictions with the radiologists' scores and constructed the combined model M_{R+HR}. M_{R+HR} achieved an average AUC of 0.909 on D_{Test}, corresponding to the highest AUC among all models (M_R= 0.861, p=0.041; M_{HR1} = 0.815, p<0.001; M_{HR2} = 0.796, p<0.001).

2.4 Discussion

Current CT technologies have improved and expedited early lung nodule diagnosis. Patients diagnosed as MIA survive well, postoperative recurrence and lymph node metastasis are rare, and 5-year survival rate is close to 100%. In contrast, patients with INV have reduced five-year survival[27][28].[29] Lobectomy is considered the standard surgical treatment for INV patients.[10] Prior studies using CT scans features of air bronchograms and borders have not been able to accurately distinguish invasive lesions.[28] An accurate way to determine the lesion's invasiveness pre-operatively on routine chest CT scans would be beneficial in guiding the need for the timing of resection and potentially amount of resection.[10]

In our work, we developed a computerized model using textural patterns known as radiomics to accurately differentiate MIA from INV cases from pre-treatment baseline CT scans from four different institutions. We observed that radiomic features extracted from intra- and peritumoral regions of these lung nodules harbor information related to nuances of the tissue properties not apparent to the naked eye. Additionally, in our analysis, two radiologists examined these scans in a blinded fashion. They scored them visually, and the integration of radiologists' interpretation with the classifier performance yielded the highest diagnostic accuracy on the test set (AUC = 0.909).

Although there have been previous successful attempts to examine GGOs via radiomics analysis[17][18][30],[31] most studies focus on textural patterns extracted from within the lung lesions to differentiate MIA from INV lesions. Specifically, most of them employed features focused on the gray level co-occurrence-based matrix and wavelet-based feature families for identifying INV cases[17].[32] A few studies have further integrated clinical and morphological features into the radiomics model to improve model accuracy[17].[33]

Two of the top five features identified by our radiomics based supervised approach corresponded to the gray-level co-occurrence-based feature (GLCM) families which is in line with previously published results (20). In addition, we also noticed Laws and Laplace features extracted from within the nodule to be among the top set of discriminating features. These

two feature families (Laws and Laplace) examine higher-order frequency content of the given region of interest.[21] We noticed a higher expression of all intratumoral features for INV when compared to MIA nodules. The elevated expressions of these radiomic features could reflect more chaotic and haphazard microarchitecture within the comparatively high-risk invasive tumors (Figure 2.5).

In our work, we also interrogated the tumor environment (TME) surrounding the nodule (i.e., peritumoral region) to evaluate its utility in providing complementary information with respect to disease diagnosis. We defined the radiomic profile of these GGO nodules during the feature discovery portion using a combination of intra- and peritumoral regions. Within our analysis, we noticed one of the top five features was from the peritumoral region. The feature was observed from within the 3 to 6 mm region outside the nodule. Recent studies have shed new light on this complex interaction between tumor and host immune cells and immune responses. In work by Altorki et al. (22), the authors demonstrated the role of TME for progression for pre-invasive to invasive adenocarcinoma lesions. They observed a dominant regulatory T cell-mediated immune suppression initiated at the precursor level sustained with rising intensity throughout malignant progression. Few studies also show that these perinodular radiomic features may reflect tumor microarchitecture changes or be capturing the presence of tumor-infiltrating lymphocytes (TILs).[15] We noticed an increased peritumoral CoLlAGe feature[20] expression for MIA cases.

Specifically with respect to the perinodular region, in work by Wu G. et al.,[32] the authors did not observe an improvement in AUC with the addition of radiomic features from the perinodular region to differentiate INV cases from MIA and AIS ($p = 0.11$). They observed the most predictive features to emanate from the ground-glass and solid regions of the nodule. Whereas in the work by Wu L. et. al.,[34] the authors show the utility of perinodular features for the same clinical problem. However, in our analysis, we noticed CoLlAGe peritumoral radiomic features to be statistically significant between the training and testing cohorts ($p < 0.01$). CoLIAGe captures higher-order co-occurrence patterns of local gradient tensors at

a voxel level and has been shown to be diagnostic and prognostic for a variety of disease indications[1415].[20] Additionally, in our analysis, we included the complete GGOs in addition to semisolid nodules unlike in the study by Wu et al.[32]

We further evaluated and compared our radiomic model with the tumor diameter. Studies show the two-dimensional diameter of the nodule to be one of the strongest predictors for pulmonary nodule risk classification in the quantitative CT image analysis. In work by Xu et al.,[30] the authors noticed the diameter of GGOs to be significantly different in MIA and INV nodules, and a conventional model constructed using clinical and quantitative features (such as age, diameter, and density) yielded the best AUC (0.848; 95% CI = 0.750-0.946). The authors observed that the addition of radiomic features to the clinical and quantitative models did not improve the performance of the combined model.[30] In contrast, multiple studies have reported the added benefit of radiomics to clinical and quantitative models[33].[17] In a study by Weng et al.,[17] the authors constructed a nomogram using lesion shape, solid component, and radiomics features from the nodule to obtain an AUC of 0.88. Similarly, Luo et al.[33] used three CT features (pleural indentation, solid component size, and solid component proportion) and one radiomic feature to help differentiate invasive pulmonary adenocarcinoma (IPA) from non-IPA to achieve a final AUC of 0.903. Interestingly, in our analysis, the radiomic model was superior to the model constructed with the nodule area in both training and testing sets. The addition of the nodule diameter to the radiomics model did not improve the performance especially in the independent validation set (D_{Train}:0.95 [0.92-0.98] from 0.92 [0.87-0.97], p=0.03; D_{Test}:0.869 [0.80-0.93] from 0.862 [0.79-0.93], p=0.86) even though individual tumor diameter was statistically significant in differentiating MIA and INV nodules (D_{Train}¡ 0.05; D_{Test}¡ 0.05). We further created a subset of nodules with a diameter of less than 10 mm. We noticed that our radiomics classifier was prognostic even within the smaller nodules, giving an AUC of 0.76 [0.53-0.98] on these smaller lesions.

Another unique aspect of our study included integrated classifier performance with expert

radiologists' visual assessment of the tumors. We noticed that the classifier had an overall improvement of 4.5% compared to the radiologists' interpretations. We noticed that the radiologists had high sensitivity, but poor specificity. After combining the probabilities of the machine learning classifier with the radiologists' score, the model AUC improved to 0.909 from 0.867 of the classifier model ($p<0.05$) and 0.816 of the radiologists' model ($p<0.05$).

Overall, our study has three main novel contributions including the multi-institutional nature, the addition of novel radiomics descriptor in the analysis, and human-machine comparison and integration to create consensus and accurate models.

Despite the progress made in this study, our work has some limitations. First, the developed model is completely retrospective in nature. For a successful transition into the clinically deployable model, a prospective evaluation will be required. Second, even though the analysis had multiple institutions, we did not truly validate the model independently since all the cases from individual sites were collapsed and subsequently randomly divided into training and testing sets. Future work will entail prospective data as well as validation on data from sites independent from those employed for developing the model.

Chapter 3

Radiomics for Predicting the Added Benefit of Adjuvant Chemotherapy for Early-Stage NSCLC

3.1 Overview

The pathological staging system proposed by the International Association for the Study of Lung Cancer is considered the gold standard in treatment planning in non-small cell lung cancers (NSCLC).[35] Stages I and II NSCLC are considered early stage, and surgical resection is the standard of care for patients with these cancers.[36] In addition, adjuvant cisplatin-based chemotherapy is currently recommended for patients with stage II NSCLC, whereas patients with stage I NSCLC continue to be treated with surgery alone[36][37][38].[39] These recommendations are based on large clinical trials showing a 5-year overall-survival benefit in the adjuvant chemotherapy group compared with the surgery alone group of 4% (hazard ratio [HR] 0·86, 95% CI 0·76–0·98; International Adjuvant Lung Cancer Trial [IALT], n=1867)[38] and 15% (0·69, 0·52–0·91; JBR.10 trial, n=482).[39] The Lung Adjuvant Cisplatin Evaluation[40]—a pooled meta-analysis that included 4584 patients across five clinical trials—found

a HR for surgery with versus without adjuvant chemotherapy of 0·89 (0·82–0·95; p=0·005), corresponding to a 5-year survival absolute benefit of 5·4% from adjuvant chemotherapy. However, other clinical trials, including the Big Lung Trial (n=381)[41] and the Adjuvant Lung Cancer Project Italy (n=1088) 8, found no significant benefit for either 5-year overall survival or disease-free survival. The inconsistency between the results from different trials might suggest the need for a biomarker that could identify patients who would benefit from adjuvant chemotherapy.

Stratified subgroup analysis of these trials based on the eighth edition of the tumour, node, metastasis (TNM) staging system found that adjuvant chemotherapy does not significantly improve overall survival in patients with stage IB (T2aN0M0) NSCLC (HR 0·93, 95% CI 0·78–1·10)[3839].[40] The CALGB9633 trial[42] (n=344), which included only patients with stage IB NSCLC showed no survival benefit with adjuvant chemotherapy (0·83, 0·64–1·08, p=0·12) with a median follow-up of 74 months. Since adjuvant chemotherapy has been shown not to lead to a significant survival benefit in stage I NSCLC (and sometimes to have detrimental effects on survival in stage IA NSCLC [ie, HR¿1]), it is currently not recommended. For stage IB patients, the American Society of Clinical Oncology guidelines do not recommend adjuvant chemotherapy following surgery in routine practice. However, National Comprehensive Cancer Network guidelines suggest that adjuvant chemotherapy is an appropriate option when considering high-risk factors such as poorly differentiated tumours, vascular invasion, wedge resection, tumour size greater than 4 cm, and pleural vascular invasion.

Even after curative resection, about 30–40% of patients with stage I NSCLC recur with post-operation observation alone, suggesting that these are high-risk patients who are at an increased risk of disease recurrence and, therefore, might benefit from adjuvant chemotherapy. Single gene-based and multigene-based expression assays, as well as traditional clinicopathological factors, have shown prognostic value in early-stage NSCLC, but only a select few have been useful for the identification of patients who might derive added benefit from

adjuvant chemotherapy[43].[44]

In the past 5 years, the use of CT-based radiomics[13] (computer-extracted quantitative imaging features derived from radiographic images) in lung cancer for disease diagnosis and prognosis has received increasing attention[13144546].[47] A few research groups have shown that imaging-based prognostic biomarkers can be used to predict disease recurrence and survival in the context of NSCLC, but most of them have little clinical utility because they are black-box approaches, and thus lack an underlying biological rationale—especially deep learning-based models. Although radiomics itself can be considered as a black-box approach, it offers the opportunity to study the association between specific image features with underlying morphological and molecular attributes of the disease, thereby providing a stronger correlative link with the underlying tumour biology. However, radiomics approaches have not been evaluated in their ability to predict the added benefit of adjuvant chemotherapy in early-stage NSCLC.

In this study, we aimed to construct a quantitative radiomic risk score (QuRiS) that employs quantitative texture features from within and outside the primary lung nodule derived from routine CT scans to predict disease-free survival and response to adjuvant chemotherapy for patients with stage I and II NSCLC. Furthermore, as some studies have suggested that signatures derived from a combination of biomarkers might have improved prognostic and predictive capabilities compared with individual analytical approaches and might therefore be better suited for informing clinical decision making[4849] we also aimed to construct a nomogram (QuRNom) integrating QuRiS with tumour and node descriptors and lymphovascular invasion status for the estimation of disease-free survival. We also aimed to explore the biological underpinning of prognostic radiomic features by exploring the association of QuRiS with genomic and pathomic data.

3.2 Methodology

3.2.1 Study design and participants

We did a retrospective multicohort study of patients with stage I and II NSCLC admitted to three independent sites, treated either with surgery plus adjuvant chemotherapy or surgery alone. We included all patients with available pretreatment diagnostic CT scans and corresponding survival and recurrence information. We excluded patients who died, recurred, or left the study within 3 months after surgery.

We reviewed the charts of patients with NSCLC admitted to the Cleveland Clinic Foundation (Cleveland, OH, USA) between April 14, 2004, and April 24, 2015. Eligible patients with resected stage I and II (early-stage) NSCLC with available diagnostic CT scans were included in cohort D_1 after pathological confirmation of the disease and its stage.

The second cohort, cohort D_2, included eligible patients identified from a retrospective chart review of patients with NSCLC continuously admitted at the University of Pennsylvania (Philadelphia, PA, USA) from Nov 10, 2004, to Feb 7, 2017. The third cohort, cohort D_3, included eligible patients for whom CT scans were available on The Cancer Imaging Archive (TCIA).

The study conformed to Health Insurance Portability and Accountability Act guidelines and was approved by the Institutional Review Board at University Hospitals Cleveland Medical Center (Cleveland, OH, USA), number 02–13–42C. Informed consent requirement was waived as the study used archival tissue.

3.2.2 Procedures

Radiomic features were extracted using software developed at the Center of Computational Imaging and Personalized Diagnostics, Case Western Reserve University (Cleveland, OH, USA), implemented on a MATLAB release 2016a platform, to capture textural heterogeneity from within and outside the nodules from the chest CT scans for all patients with early-stage

lung cancer. An experienced cardiothoracic radiologist (PR) identified the target nodule on CT scans in cohorts D_1 and D_2. A second cardiothoracic radiologist (RG) identified the nodules on the scans in cohort D_3. Each nodule was further manually segmented across all the sections where it was visible in the axial view using a manual annotation tool in (3D Slicer, version 4.6). These annotations were considered for intranodular and perinodular radiomic analysis (figure 6.3).

Following annotations, the peritumoral mask was selected using morphological dilation and erosion operations. An annular 3-mm wide area at a radial distance of 15 mm outside the nodule was selected for extracting peritumoral features.

Radiomic features from Gabor, Haralick, Laws, Laplace, and Collage feature families were extracted from each annotated region. To ensure the stability and reproducibility of radiomic features, we used the test-retest Reference Image Database to Evaluate Response lung CT dataset.19

The most useful predictive features from the patients who underwent surgery alone in the discovery cohort D1 were used to construct QuRiS, and QuRNom was then constructed using QuRiS combined with prognostic clinical variables identified using the training (D_1) cohort. QuRNom was used to show an incremental value of QuRiS compared with traditional clinicopathological factors for personalising disease-free survival prediction.

Haematoxylin-and-eosin stained whole-slide tissue scans obtained from surgical resections were available for a subset of patients in cohort D_1. These scans were used to explore the biological underpinnings of the prognostic radiomic features used for developing QuRiS. Tumour infiltrating lymphocytes are known to carry prognostic ability in early-stage lung cancer domain. In this analysis, we explored associations between tumour infiltrating lymphocytes and prognostic radiomic features to assess whether these radiomic features were potentially capturing information regarding tumour infiltrating lymphocytes on CT scans.

We used sequencing data from frozen-mRNA samples from cohort D3 from the TCIA Pan-Cancer project to find associations between prognostic radiomic features and gene expression

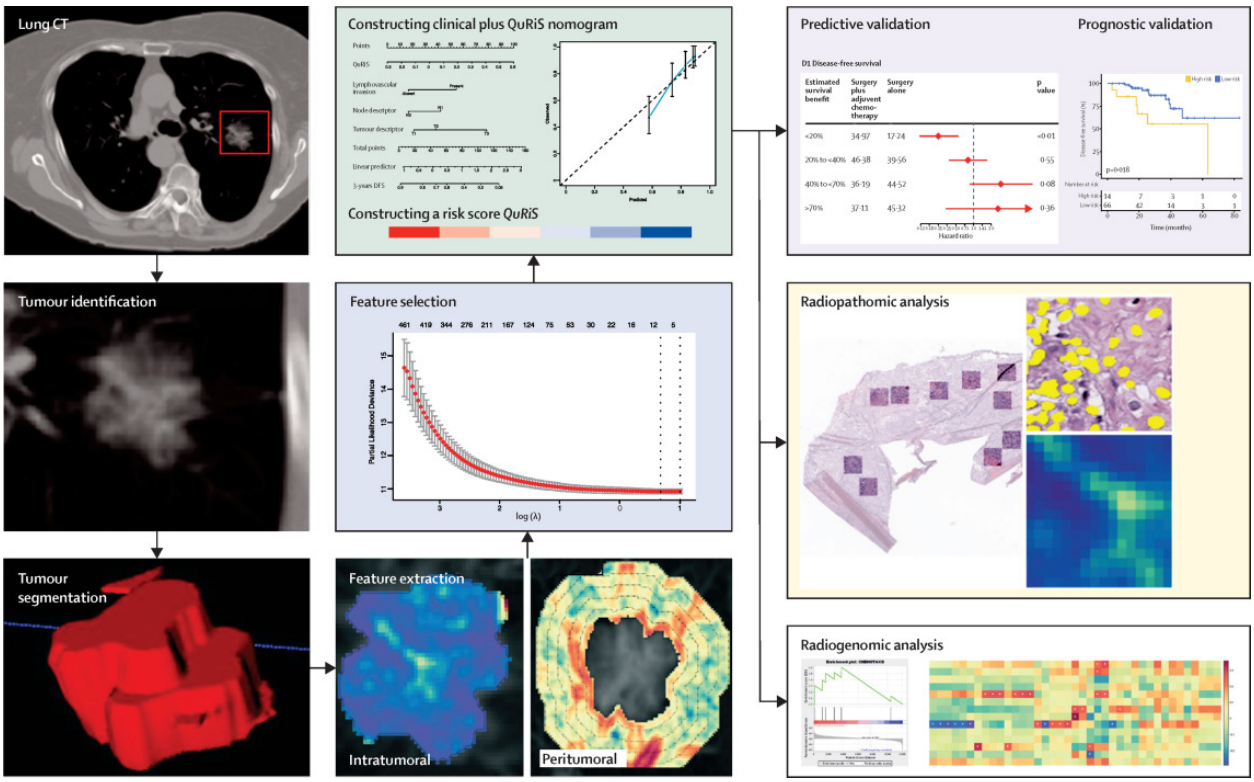

Figure 3.1: Overall workflow and pipeline of the project. The first step involved identifying and annotating the primary nodule on the CT scan. Intratumoral and peritumoral textural features were extracted using Matlab 2016a. For the peritumoral region, features were extracted from 0–15 mm region outside the tumour and divided into five 3-mm peritumoral rings. Feature statistics included mean, median, SD, skewness, kurtosis, and range and were calculated for each of the individual annular rings. Top features were selected using the LASSO feature selection method and used for constructing QuRiS. QuRNom was constructed using prognostic clinical features and QuRiS. QuRiS and QuRNom were validated for prognostic performance and predicting added benefit of adjuvant-chemotherapy. Associations between QuRiS features and spatial patterns of tumor-infiltrating lymphocytes on whole-slide tissue scans were also evaluated, as were associations with mRNA data and underlying immune specific biological pathways. DFS=disease-free survival. LASSO=least absolute shrinkage and selection operator. QuRiS=quantitative radiomic risk score. QuRNom=quantitative radiomic nomogram.

data.

3.2.3 Outcomes

The primary endpoint of the study was the predictive and prognostic performance of QuRiS and QuRNom in regard to disease-free survival, which was measured from the date of surgery to the time of disease relapse (tumour recurrence within or immediately adjacent to the treated field, mediastinal relapse, or distant relapse) or the time of death, whichever occurred earlier; patients who were alive and did not relapse were censored at the date of last follow-

up.[50] We focused our analysis on disease-free survival instead of overall survival because, in patients with stage I and II NSCLC, overall survival is longer than in patients with later stages of disease and we aimed to find a correlation with disease-specific survival in the developed model.

Other objectives of our study were to assess whether QuRiS and QuRNom could be used to identify patient cohorts with early-stage disease that would receive added benefit from adjuvant chemotherapy after surgery and whether QuRiS features were associated with the spatial architecture and arrangement of cancer nuclei and tumour-infiltrating lymphocytes derived from haematoxylin-and-eosin stained tissue images of corresponding surgical specimens (radiopathomic analysis) and with underlying biological pathways in cancer progression by using mRNA sequencing data (radiogenomic analysis).

3.2.4 Statistical analysis

The least absolute shrinkage and selection operator (LASSO) method, which is suitable for high-dimensional data,21 was used to select the most useful predictive features from the patients who underwent surgery alone in the discovery cohort D_1. After selecting the top features, the corresponding LASSO coefficients were used for QuRiS construction. QuRiS was calculated for each patient via a linear combination of selected features that were weighted by their respective coefficients. The value of the tuning parameter in the LASSO-Cox model (λ) was averaged out by 10 cross-validations to minimise error.

QuRNom, was constructed using the discovery cohort D_1 and comprised the radiomics model as well as significant clinical-pathological risk factors found on multivariate Cox regression analysis.[51]

For prognostic stratification, QuRiS values were used to divide cohort D_1 into two groups for which disease-free survival and HRs were calculated. The prognostic performance of QuRiS was validated using Kaplan-Meier survival analysis, log-rank test, HR (95% CI), and Harrell's concordance index (C index [95% CI]). Univariate analysis of QuRiS and the

clinical–pathological variables was done. Multivariable Cox-regression analysis was used to investigate the relationships between the various covariates and 3-year disease-free survival. Further subset analyses involved determining survival differences in different stages. To measure nomogram performance, C indices were calculated from the nomogram for QuRiS, clinical factors alone, and QuRiS plus clinical factors. The calibration plot for the nomogram was plotted by reviewing the plots of nomogram-predicted survival probabilities with Kaplan-Meier estimated probabilities along with the Hosmer-Lemeshow test. Bootstraps with 1000 resamples were used to quantify model overfitting and calculate Kaplan-Meier estimates.

For predictive validation, QuRis was used to divide our patient dataset into three risk groups (high, intermediate, and low risk). These groups were selected using thresholds determined by calculating and comparing the HRs predicting 3-year disease-free survival in patients in cohort D_1 who received adjuvant chemotherapy versus those who received surgery alone. Forest plots were constructed to show the HRs comparing disease-free survival between the patients receiving adjuvant chemotherapy and those undergoing surgery alone in all the cohorts amongst the QuRiS risk groups. For predictive validation of combined QuRNom, nomogram-based estimates of survival probability were used for comparing HRs between adjuvant chemotherapy and surgery alone groups.

For finding the histomorphometric correlation of QuRiS, after extracting the spatial architecture of tumour infiltrating lymphocyte features from surgically resected whole-slide tissue scans, we calculated Pearson correlation coefficient (r) between prognostic radiomic features used for constructing QuRiS and the spatial architecture of tumour infiltrating lymphocyte features for the radiopathomic analysis.[52]

Radiogenomic analysis was done using mRNA sequencing data obtained with Illumina Genome Analyzer Sequencing version 2 (Illumina, San Diego, CA, USA).24 Top genes were selected by correlating mRNA sequencing data with QuRiS. These genes were used in gene ontology analysis to identify molecular pathways indicating biological processes associated with the QuRiS.

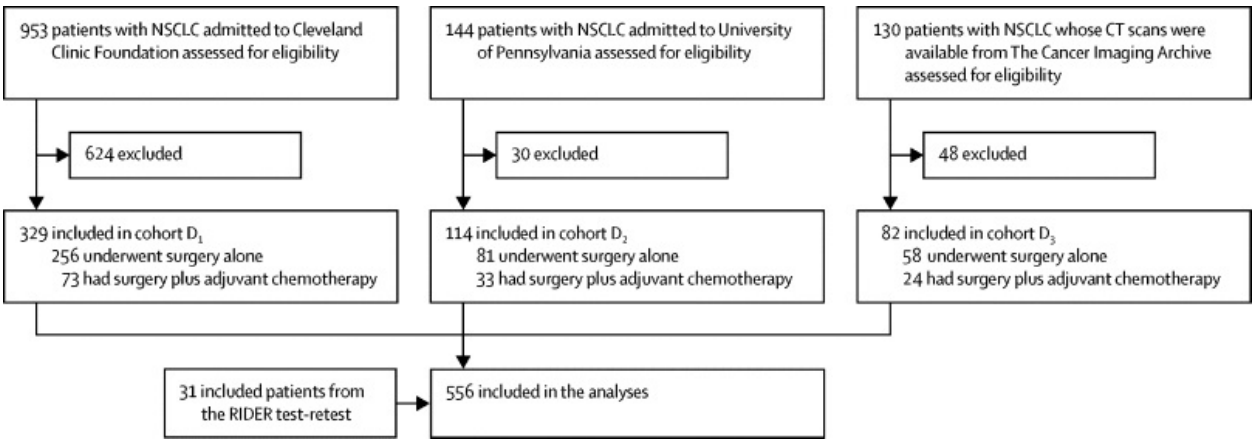

Figure 3.2: Study profile (NSCLC= non-small cell lung cancers)

All statistical analyses was done with R version 3.5.3.

3.3 Results

After reviewing 700 charts of patients with NSCLC admitted to the Cleveland Clinic Foundation between April 14, 2004, and April 24, 2015, we included 329 eligible patients in cohort D_1 (figure 6.1). 73 (22%) of these patients received adjuvant chemotherapy and 256 (78%) underwent surgery alone. We included 114 eligible patients in cohort D_2 identified from a retrospective chart review of 128 patients with NSCLC continuously admitted at the University of Pennsylvania from Nov 10, 2004, to Feb 7, 2015. 33 (29%) of 114 patients in the D_2 cohort received adjuvant chemotherapy and 81 (71%) patients underwent surgery alone. The third cohort, cohort D_3, included 82 eligible patients from TCIA. 58 (71%) of 82 patients in cohort D_3 underwent surgery alone, whereas 24 (29%) patients received adjuvant chemotherapy. Baseline characteristics of patients in each cohort are shown in the table.

Cohort D_1 (n=329) was used as the discovery cohort to train QuRiS and to construct QuRNom. For external validation of QuRiS and QuRNom as prognostic and predictive biomarkers, D_2 and D_3 were used.

The 13 most discriminative features from the radiomic feature pool were selected to calculate the QuRiS using LASSO analysis. The top 13 features included three from the

Table 3.1: Baseline characteristics

	Cohort D_1 (n=329)	**Cohort D_2** (n=114)	**Cohort D_3** (n=82)	**p value**
Treatment				
Surgery Alone	256 (78%)	81 (71%)	58 (71%)	0.21
Adjuvant Chemotherapy	73 (22%)	33 (29%)	24 (29%)	..
Age at admission, years				
≤ 65	86 (26%)	46 (40%)	34 (41%)	0.79
> 65	142 (43%)	68 (60%)	48 (58%)	..
unknown	101 (31%)	0	0	..
Gender				
Women	113 (34%)	76 (67%)	26 (32%)	< 0.0001
Men	115 (35%)	38 (33%)	56 (68%)	..
unknown	101 (31%)	0	0	..
Smoking status				
Current	85 (26%)	6 (5%)	17 (21%)	< 0.0001
Previous use	209 (64%)	82 (72%)	47 (57%)	..
Never used	32 (10%)	26 (26%)	18 (22%)	..
Unknown	3 (1%)	0	0	..
Tumor Stage				
Stage I	276 (84%)	89 (78%)	57 (70%)	< 0.0001
Stage II	53 (16%)	33 (28%)	25 (30%)	..
Lymphovascular invasion status				
Present	90 (30%)	Unknown	5 (6%)	0.011
Absent	214 (65%)	Unknown	74 (90%)	..
Unknown	28 (9%)	Unknown	0	..
Node descriptor				
0	255 (78%)	Unknown	69 (84%)	0.89
1	46 (14%)	Unknown	13 (16%)	..
Unknown	28 (9%)	Unknown	0	..
Tumor descriptor				
1	158 (48%)	Unknown	38 (46%)	0.026
2	88 (27%)	Unknown	31 (38%)	..
3	8 (2%)	Unknown	13 (16%)	..
Unknown	75 (23%)	Unknown	0	..
Margin status				
Negative	323 (98%)	144 (100%)	Unknown	0.22
Positive	214 (65%)	Unknown	74 (90%)	..
Unknown	28 (9%)	Unknown	0	..

Table 3.2: Baseline characteristics

	Cohort D_1 (n=329)	**Cohort D_2** (n=114)	**Cohort D_3** (n=82)	**p value**
Tumor location				
Right upper lobe	128 (39%)	35 (31%)	26 (32%)	0.32
Right lower lobe	45 (14%)	22 (19%)	11 (13%)	..
Right middle lobe	15 (5%)	4 (4%)	8 (10%)	..
Left upper lobe	94 (29%)	32 (28%)	23 (28%)	..
Left lower lobe	47 (14%)	21 (18%)	14 (17%)	..
Type of surgery				
Lobectomy	64 (19%)	55 (48%)	Unknown	< 0.0001
Sublobar	234 (71%)	49 (43%)	Unknown	..
Pneumonectomy	15 (5%)	1 (1%)	Unknown	..
Segmentectomy	16 (5%)	0	Unknown	..
EGFR mutation				
Mutant	25 (8%)	36 (32%)	17 (21%)	0.16
Wildtype	95 (29%)	78 (68%)	54 (66%)	..
Unknown	209 (64%)	0	11 (13%)	..
KRAS mutation				
Mutant	2 (1%)	32 (28%)	17 (21%)	0.66
Wildtype	3 (1%)	82 (72%)	54 (66%)	..
Unknown	324 (98%)	0	11 (13%)	..
Histology				
Adenocarcinoma	211 (64%)	114 (100%)	82 (100%)	< 0.0001
Squamous Cell Carcinoma	93 (28%)	0	0	..
Other	21 (6%)	0	0	..
Unknown	4 (1%)	0	0	..
Left lower lobe	47 (14%)	21 (18%)	14 (17%)	..
Recurrence				
Yes	71 (22%)	25 (22%)	26 (32%)	0.14
No	158 (78%)	89 (78%)	56 (68%)	..

intratumoral and ten from the peritumoral region. All the top features were observed from Haralick or Collage feature families representing the grey-level co-occurence matrix. QuRiS could predict disease-free survival in cohorts D_1 (HR 1·56, 95% CI 1·08–2·23, p=0·016), D_2 (2·66, 1·24–5·68, p=0·011), and D3 (2·67, 1·39–5·11, p=0·0029). Among patients who received surgery alone, we noticed two definite groups with high and low risk of recurrence (figure 3b). These groups were defined using a QuRiS threshold of 0·606, below which patients were observed to have low risk of recurrence (and hence increased disease-free survival) and above which patients were observed to have high risk of recurrence (hence decreased disease-free survival). By contrast, no significant difference was observed between high-risk and low-risk patients who received the adjuvant chemotherapy, potentially suggesting that a subset of patients in this cohort benefited from chemotherapy and had improved disease-free survival (figure 3C). In the subgroup analysis, within stage I cases, we could still observe two distinct predicted high-risk and low-risk groups according to QuRiS (HR 1·89, 95% CI 1·17–3·05, p=0·008 for cohort D_1; 3·87, 1·53–9·81, p=0·002 for cohort D_2; and 2·42, 1·01–5·79, p=0·047 for cohort D_3;).

Univariate analysis on cohort D_1 revealed that—among all clinicopathologic factors considered—tumour and node descriptors (as per the TNM staging system), lymphovascular invasion status, pathological stage, and margin status are prognostic factors. We included tumour and node descriptors, lymphovascular invasion status, and QuRiS to construct QuRNom for predicting 3-year disease-free survival . We could not include margin status as another independent parameter in the nomogram as almost all the patients had a negative margin. The stage was mainly a representation of tumour and node descriptors and hence was excluded from the final model.

The calibration curve for QuRNom for predicting 3-year disease-free survival showed agreement between predicted survival and actual survival. The Hosmer-Lemeshow test yielded a p value of 0·68, suggesting that predicted and observed disease-free survival were similar. The C index for QuRNom was 0·74 (95% CI, 0·72–0·76), which was confirmed via

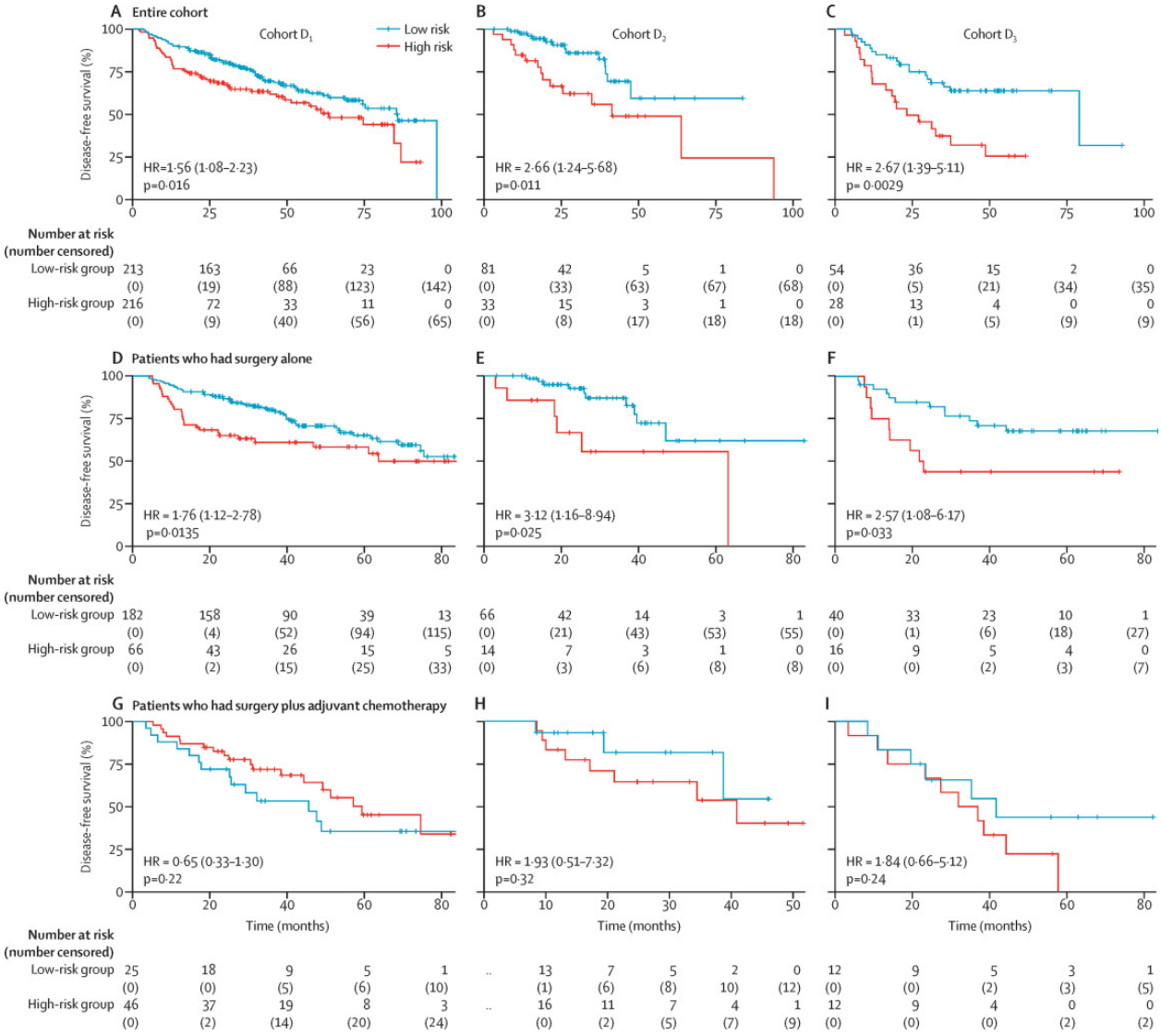

Figure 3.3: Kaplan-Meier plot for disease-free survival according to QuRiS-based risk groups. Disease-free survival for the entire (A) D_1, (B) D_2, and (C) D_3 cohorts, patients who received surgery alone in cohorts (D) D_1, (E) D_2, and (F) D_3, and patients who had surgery plus adjuvant chemotherapy in cohorts (G) D_1, (H) D_2, and (I) D_3. The QuRiS threshold of 0·0646 was developed using training cohort D_1 and validated in cohorts D_2 and D_3. HR=hazard ratio. QuRiS=quantitative radiomic risk score.

bootstrap validation.

For predictive validation, QuRiS values were used to classify patients into three risk groups (high, intermediate, and low risk) on the basis of HRs from the D_1 cohort. Patients with QuRiS score below 0·093 were considered low risk and above 0·099 were considered as high-risk patients. All the patients with a QuRiS score of 0·093 to 0·099 were considered at intermediate risk.

Patients in the high-risk group showed significantly prolonged survival when patients received adjuvant chemotherapy (HR 0·27, 95% CI 0·078–0·95, p=0·042 for cohort D_1 and 0·076, 0·013–0·42, p=0·0029 for cohorts D_2 and D_3 combined), whereas there was no added benefit for patients in the low-risk group and, in fact, the HR point estimate indicates negative effects of adjuvant chemotherapy in both training and validation cohorts (2·15, 1·13–4·08, p=0·019 for cohort D_1 and 2·25, 0·95–5·03, p=0·064 for cohorts D_2 and D_3 combined; figure 4A, 4B). Stratified subgroup analysis within stage IB showed a similar difference in treatment-related effects between different QuRiS groups.

Furthermore, QuRNom was used for predicting the added benefit of adjuvant chemotherapy. The QuRNom-estimated survival benefit was used for differentiating patients into different subgroups (patients with survival benefit 20%, patients with survival benefit ¿20% to 40%, patients with survival benefit ¿40% to 70%, and patients with survival benefit ¿70%). Patients with survival benefit less than 20% showed an improved median disease-free survival of about 17 months in cohort D_1 (HR 0·27, 95% CI 0·11–0·54, p¡0·0003) and 27 months in cohort D_3 (0·13, ¡0·01–0·99; p=0.0019). As clinical information regarding lymphovascular invasion status and tumour and node descriptors were not available for cohort D2, a nomogram could not be constructed. In patients with survival benefit 40–70% or greater than 70%, the patients who received adjuvant chemotherapy showed poorer survival than those who underwent surgery alone (2·6, 0·34–19·24, p=0·36 for cohort D_1 and 1·5, 0·33–6·5, p=0·62 for cohort D_3; figure 4C, 4D).

Histomorphometric analysis revealed QuRiS to be significantly correlated with the clus-

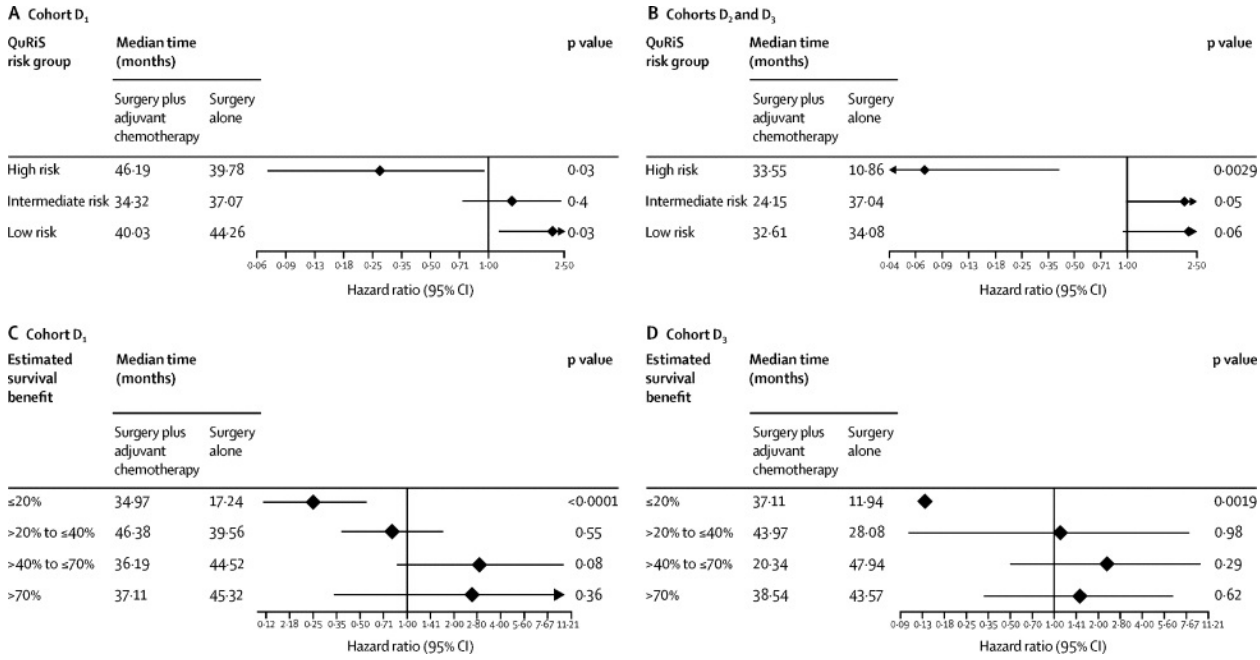

Figure 3.4: Forest plots of disease-free survival. Hazard ratios based on QuRiS comparing disease-free survival in patients who received surgery plus adjuvant chemotherapy and those receiving surgery alone in (A) cohort D_1 and (B) cohorts D_2 and D_3. Hazard ratios by nomogram-estimated survival benefit from adjuvant chemotherapy in (C) cohort D_1 and (D) cohort D_3. The survival benefit of adjuvant chemotherapy was calculated as the difference between nomogram-estimated 3-year disease-free survival for patients who received surgery plus adjuvant chemotherapy and those who had surgery alone. QuRiS=quantitative radiomic risk score.

tering patterns of the cancerous cells observed on haematoxylin-and-eosin-stained whole-slide scans of 70 patients in cohort D1 (Pearson correlation coefficient r=0·23; p=0·036).

Individual radiomic feature analysis with the spatial tissue-infiltrating lymphocytes features revealed the intratumoral Haralick energy feature to be associated with various features explaining the tumour-infiltrating lymphocyte–nuclei interplay (such as the ratio of cancerous clusters to lymphocyte clusters or the density of cancerous clusters). QuRiS predicted that high-risk patients would have a more chaotic and disturbed microarchitecture on CT scans. For these patients, the corresponding whole slide tissue images revealed very dense, tightly bound cancer nuclear clusters (figure 5A–H).

In the radiogenomic analysis of 82 patients in cohort D_3, QuRiS was observed to be associated with biological pathways implicated in angiogenesis, proliferation, cellular differentiation, T-cell and lymphocyte activation, and chemotaxis . Although these relevant pathways included less than 100 genes, their fold enrichment (ie, how much more these pathways were correlated with the radomic risk score than they would be by random chance) was always greater than two.

Specifically, the intratumoral Haralick energy feature was observed to be inversely correlated with macrophage chemotaxis.

3.4 Discussion

In this Article, we present the first-of-its-kind CT radiomics-based score (QuRiS) that is predictive of benefit of adjuvant chemotherapy and associated with survival in early-stage NSCLC. Current treatment guidelines (released by the International Association for the Study of Lung Cancer, the American Society of Clinical Oncology, and National Comprehensive Cancer Network) do not recommend the addition of adjuvant chemotherapy following curative resection in stage IA NSCLC, whereas the potential use of adjuvant chemotherapy in stage IB NSCLC is fraught with controversy, mainly due to ambiguous results from large-

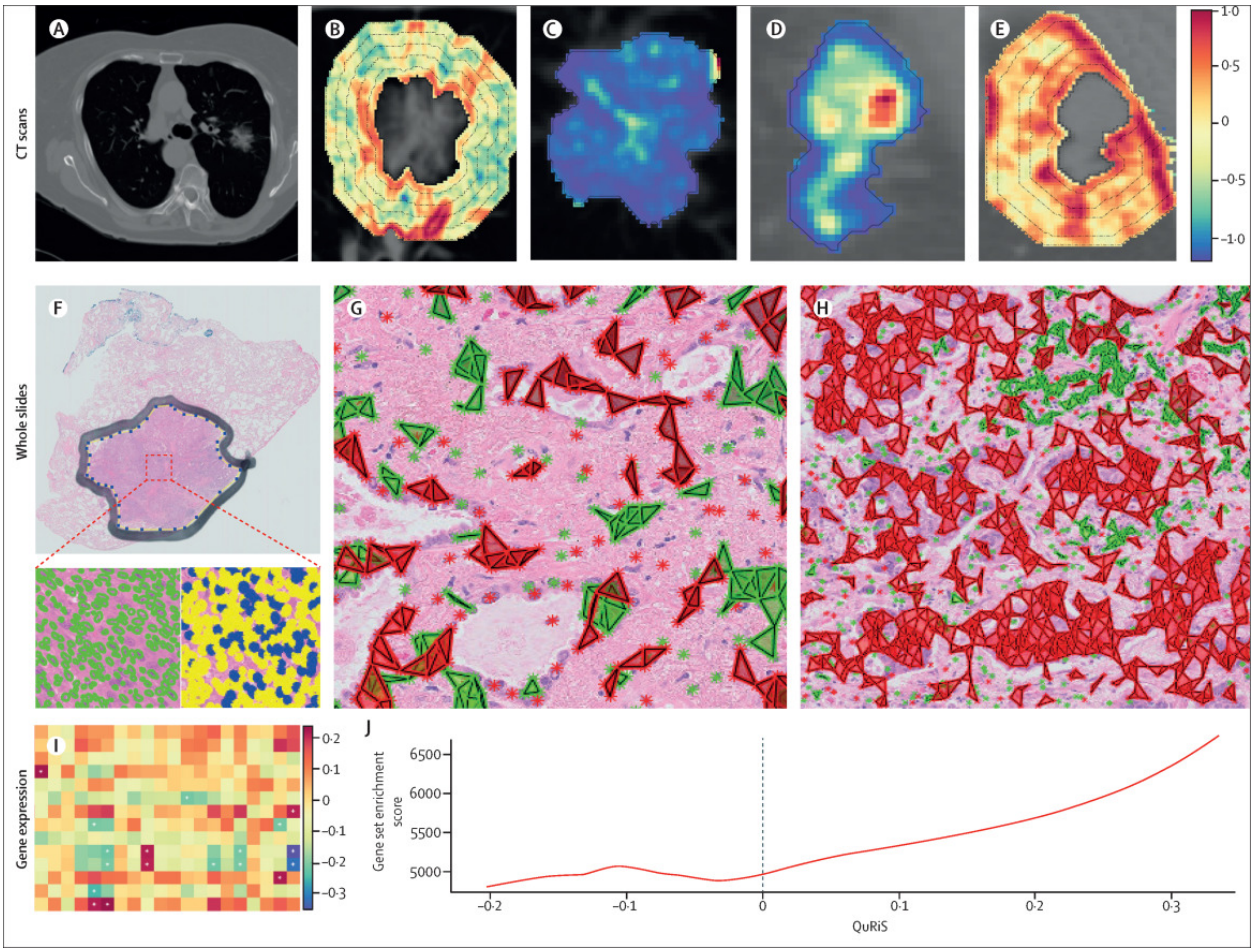

Figure 3.5: Association between radiomic, pathomic, and genomic data. (A) CT scan of patients with lung cancer. (B) Intratumoral and (C) peritumoral radiomic Haralick textural feature for low-risk nodule. (D) Intratumoral and (E) peritumoral Haralick feature for high-risk nodule. Panels (B–E) are shown at a magification of 103 (zoomed in from panel A). (F) Whole-slide tissue scan corresponding to the CT scan in (A); green colouring of cells indicates the first segmentation of all nuclei on the whole-slide scans; yellow cells are non-lymphocyte cells and blue are lymphocytes. Spatial tissue infiltrating lymphocytes feature on whole-slide scan corresponding to (G) low-risk tumours identified in panels (B) and (C), and (H) high-risk tumours identified in panels (D) and (E). In (G) and (H), red stars represent cancer cells and red triangles represent cancerous clusters; green stars represent individual tumour infiltrating lymphocytes and green triangles represent tumour infiltrating lymphocyte clusters. Panels (G, H) are shown at a magification of 107. (I) Radiogenomic analysis. (J) Chemotaxis gene set enrichment analysis. QuRiS=quantitative radiomic risk score.

scale clinical trials investigating the added benefit of adjuvant chemotherapy in patients treated with curative resection. By contrast, in stage II NSCLC, adjuvant chemotherapy is currently recommended universally without any possibility of risk stratification. These diverse recommendations highlight the need for a predictive biomarker that could potentially identify patients who are at increased risk of recurrence, even with cancers clinicopathologically defined as low stage (stage I). These patients would be ideal candidates for adjuvant chemotherapy following resection. Conversely, the identification of patients who are at low risk of recurrence, even among those who have higher-stage NSCLC, might prevent them from having the toxic effects of adjuvant chemotherapy without gaining any additional benefit over resection.

Currently available biomarkers in NSCLC include largely prognostic signatures reliant on clinical factors such as circulating tumour DNA, tumour macrophages, single-gene and multigene expression profiles, and next-generation sequencing. For example, a 14-gene expression signature was developed in one study with quantitative PCR analysis of DNA from 361 patients with non-squamous NSCLC.[53] The assay was independently validated on two cohorts of 433 patients (5-year overall survival was 71·4% in low-risk patients and 49·2% in high-risk patients; p=0·0003) and 1006 patients (5-year overall survival was 74·1% in low-risk patients and 44·6% in high-risk patients) with stage I–III non-squamous NSCLC. However, to offer support in patient management and achieve maximum clinical benefit, signatures need to be not only prognostic but also predictive of benefit of administering adjuvant chemotherapy. The few existing predictive signatures are based on gene-expression assays, which are expensive and time-consuming and disrupt clinical workflow as tissues need to be shipped to a central laboratory. In another study, a 15-gene predictive signature was developed using tissue microarray samples from 133 patients (62 who underwent surgery alone and 71 who also received adjuvant chemotherapy).[43] The signature was developed from patients who received surgery alone and was then applied to the cohort of patients who also received adjuvant chemotherapy; the high-risk patients showed improved overall survival (HR 0·33, 95%

CI 0·17–0·63; p=0·0005), whereas overall survival of low-risk patients was worsened (3·67, 1·22–11·06; p=0.013).

The model we developed, QuRiS, comprises 13 radiomic textural features, three features from the intranodular region, and 10 from the perinodular region characterising the tissue microenvironment. The tissue microenvironment is the immediate cellular area outside the tumour and typically includes a plethora of cells, such as fibroblasts, dendritic cells, and inflammatory markers. Research from the past 5 years has suggested that this area contains information that relates to drug resistance and the effectiveness of chemotherapy and immunotherapy[545556].[57] The resection status of tumours can have a role in the solicitation of the immune response as resection must be complete, with no residual tumour cells, for the immune system to work. Previous investigations of radiomic features from the tissue microenvironment have shown their utility in diagnostic and prognostic lung cancer domain. The most discriminative features from the tissue microenvironment included Haralick and Collage features from 15 mm outside the tumour, which capture textural heterogeneity and were found to be high in poor responders (high-risk patients). This finding could potentially reflect the increased heterogeneity in rapidly dividing tumours, which are more aggressive tumour variants with more chaotic and disordered growth.

QuRiS was also developed to predict benefit to adjuvant chemotherapy in early-stage NSCLC patients. The QuRiS high-risk group showed a significant disease-free survival benefit in internal and external validation sets, with an estimated 70% survival benefit with an improved median disease-free survival of about 17 months in cohort D_1, 24 months in cohort D_2, and 27 months in cohort D_3 for patients who received adjuvant chemotherapy versus those who received surgery alone. In the QuRiS low-risk groups, however, all three cohorts showed an increase in the HR point estimate, although the results were not significant, for patients who received adjuvant chemotherapy. This finding suggests that at least in the QuRiS low-risk groups, there is no added benefit of instituting potentially toxic adjuvant chemotherapy.

Subanalysis of individual cancer stages revealed that QuRiS was prognostic even within stage I. The IALT and JBR.[43] trials were among the first large randomised clinical trials to show that adjuvant chemotherapy, after complete surgical resection, improved 5-year overall survival compared with observation alone. But even in these trials, there were no significant differences in survival of patients with stage I cancer treated with adjuvant chemotherapy versus surgery alone (HR¿1 for patients with stage IA cancer, meaning that adjuvant chemotherapy had a detrimental effect on survival, possibly because of toxicity). The results of our QuRiS-based analysis could explain the clinical trial results, as there seem to be two distinct risk groups within each clinically defined risk group (ie, low-risk [stage I] early-stage disease and high-risk [stage II] early-stage disease).

To include QuRiS within a clinical model, we constructed QuRNom, which was a combination of QuRiS with tumour and node descriptors and and lymphovascular invasion status. QuRNom showed significantly superior performance for predicting disease-free survival in early-stage NSCLC compared with the eighth edition of the TNM classification (C index 0·65, 0·64–0·66) and with a combined clinical–pathological model (C index 0·71, 0·69–0·72). Additionally, QuRNom was also validated for predicting added benefit of adjuvant chemotherapy.

To the best of our knowledge, QuRiS is the first radiomics-based risk model that is both prognostic and predictive of adjuvant chemotherapy benefit in early-stage NSCLC patients treated primarily with curative resection. One study defined a radiomic signature using a LASSO-Cox regression model constructed using 132 radiomic texture features from the histogram and grey-level co-occurrence matrix families, from within the node of 282 patients with early-stage NSCLC.[46] The study showed that incorporating the radiomic signature into a radiomics-based nomogram (radiomics and significant clinicopathologic factors) resulted in a better performance for the estimation of disease-free survival (C index 0·72, 95% CI 0·71–0·73) than with the clinical-pathological nomogram alone (0·691, 0·68–0·70; p¡0·0001). Another study showed that in 186 patients with NSCLC, a random forest-based classifier using intranodular radiomic features could predict patients who died versus those who were

alive with an area under the curve of 0·93.[58] Our study differs from these works in that it used perinodular radiomic features up to 15 mm outside the tumour in addition to intranodular features, it used independent validation sets for testing the developed radiomic signature, and it confirmed the ability of the signature and constructed nomogram to predict benefit of adjuvant chemotherapy.

Our study also differed from previous works in that it investigated the underlying biological, including genomic, characteristics behind the QuRiS score and the associated prognostic radiomic features. Using quantitative histomorphometric features related to the spatial interplay and arrangement of cancer nuclei and tumour-infiltrating lymphocyte clusters, we found that QuRiS was correlated with a spatial feature that represented the clustering pattern of cancerous cells, with more chaotic and closer arrangement of clusters in the QuRiS high-risk group than in the low-risk group. Furthermore, an analysis of differentially expressed genes in QuRiS risk groups revealed that QuRiS was significantly correlated with biological pathways related to cellular differentiation and angiogenesis, among others. Potential markers of angiogenesis, including microvessel density found on histological analysis, have shown to be prognostic across multiple cancer types. QuRiS association with angiogenesis pathways suggests that the perinodular texture features might also be defined by more disordered and heterogeneous blood vessel formation in high-risk patients, as compared with the better formed and more ordered blood vessel arrangements in low-risk patients. Similarly, the higher intranodular texture feature in high-risk patients is possibly due to the presence of cellular invasion and poor differentiation in the more aggressive tumour variant, regulated by the biological pathway of cellular differentiation.

Our study has some limitations. First, QuRiS was not implicitly tested for preanalytical sources of variation, such as scanner manufacture, reconstruction kernels, and slice thicknesses. However, since our study involved cohorts from multiple institutions, it relied on images from multiple scanners and with different kernels and slice thicknesses. Second, the developed nomogram, QuRNom, showed prognostic and predictive performance but was not

explicitly shown to rely on a causal association. Third, there was no way of controlling for the time between image collection and treatment delivery. Although not explicitly evaluated in this study, these temporal differences could affect the performance of QuRiS. Finally, the treatment cohorts (surgery alone and surgery plus adjuvant chemotherapy) were not homogeneously determined since they were not obtained from a controlled clinical trial. To ensure the validity of QuRiS for clinical use, patients would need to be prospectively and randomly assigned to either surgery plus adjuvant chemotherapy or surgery alone on the basis of QuRiS scores, and survival of these patients should then be analysed to confirm the efficacy of QuRiS in identifying the most appropriate treatment for patients with early-stage NSCLC.

In summary, we developed a quantitative radiomic risk score (QuRiS), based on 13 features, which was prognostic and predictive of benefit of adjuvant chemotherapy following curative resection in early-stage NSCLC. Further multisite validation including retrospective validation of archived samples from completed clinical trials followed by large prospective clinical trial evaluation, is necessary to validate QuRiS as a non-invasive biomarker to risk-stratify patients and a prognostic and predictive complementary test for early-stage NSCLC, especially given its low cost, which could represent a non-disruptive option within the clinical workflow.

Chapter 4

Predicting Hyperprogressors using Radiomic Features

4.1 Overview

The addition of immune checkpoint inhibitors (ICIs) to the armamentarium of cancer therapies has resulted in unprecedented improvement in survival outcomes for a wide range of malignancies, including non-small cell lung cancer (NSCLC). With the approval of ICIs for clinical use in these malignancies, a phenomenon of paradoxical acceleration of disease progression after initiation of immunotherapy has been recently described by a few groups[59,60,61].[62] This phenomenon was labeled 'hyperprogressive disease' with the reported incidence varying from 8.0% to 25.7%, depending on tumor histology and criteria used to identify hyperprogressors (HPs)[60,62,63].[64]

The biological and clinical factors that contribute to the development of hyperprogressive disease with ICIs are yet to be understood. Evidence from studies published to date have shown associations between hyperprogression and advanced age[64],[65] a higher number of metastatic sites at baseline4 and specific genetic alterations, such as murine double minute-2 (MDM2) amplification or epidermal growth factor receptor (EGFR) mutations.[60] However,

these observations have not been consistently noted in all the studies, and some of them have reported unexpected rapid disease progressions.[66] There is, therefore, a pressing need to identify biomarkers that could identify and help stratify this unique group of patients in whom ICI therapy not only may lack efficacy but also could lead to rapid disease progression and worse clinical outcomes.

CT images of tumors contain a vast amount of valuable information in the form of subtle variations in shape, intensity, gradient, and texture beyond the semantic features that are routinely used by radiologists to describe radiographical appearances of tumors.[14] 10 Radiomics[13] involves the use of computer vision and machine learning methods to quantitatively interrogate the subtle subvisual characteristics of radiographical images in a high-throughput manner in order to answer relevant clinical questions. Radiomic features have shown prognostic and predicting response to multiple different treatments across a wide variety of cancers, including lung[67,68,69,70,54],[15] breast[71],[72]18–21 brain[73,74,75],[76] prostate[77,78] and colorectal[46] cancers. Specifically, in lung cancer,[79] radiomics approaches have been used to predict the benefit of adjuvant chemotherapy, prognosticate disease risk in early-stage lung cancer,[15] predict treatment response to concurrent chemoradiation in locally advanced disease[70]15 and to predict response to immune checkpoint inhibition in advanced NSCLC[54,80].[81]

These features are sought to capture the extent of heterogeneity and other biologically relevant features, such as interaction with stromal or vascular components within the given region of interest. Our group and several others have examined the region beyond the tumor boundaries for this approach and interrogated the peritumoral or space immediately adjacent to the tumor. In our recently published work, we have shown that these peritumoral radiomic textural patterns can determine the response to ICI in patients with late-stage NSCLC.[54]

Another important hallmark of more aggressive cancers is the extent of neovascularization and angiogenesis present in the tumor microenvironment (TME). Our groups have looked at quantifying this vessel tortuosity as a radiomic feature—quantitative vessel tortuosity (QVT) on the CT scan as a way of representing the curvedness of the blood vessels feeding

the tumor. We have previously shown that QVT features could be used to distinguish malignant adenocarcinomas with more tortuous vasculature versus benign granulomas from non-contrast chest CT scans.[82] A marked decrease in QVT in patients with advanced NSCLC treated with ICI has been associated with response to therapy.[83]

In this study, we sought to evaluate the ability of radiomic features pertaining to intratumoral and peritumoral textural patterns and tortuosity of tumor-associated vasculature to further characterize the response patterns seen in patients with NSCLC treated with ICI. In this work, we hypothesize that hyperprogressive disease would have a unique radiomics pattern associated with it when compared with other response patterns, such as responders (stable disease (SD), partial response (PR), complete response (CR)) and non-responders (progressive disease (PD)) as determined by RECIST V.1.1.[83] We studied the performance of each feature family (intratumoral, peritumoral texture, and vessel tortuosity) within unsupervised as well as within the unified supervised classification model in predicting HPs from other response patterns. We developed our radiomic model on a training set and validated our findings on a validation set in a blinded manner. Additionally, we also performed a survival analysis based on our radiomic model using the log-rank test to stratify predicted HP patients based on overall survival (OS).

4.2 Methods

4.2.1 Study design and subjects

In this study, we retrospectively reviewed electronic medical records of 524 consecutive patients with NSCLC who received monotherapy with either a PD1 or PD-L1 inhibitor between January 1, 2015, to April 30, 2018, at Cleveland Clinic. Data pertaining to demographics, smoking history, histology, and molecular testing, number of prior lines of therapy, performance status per the Eastern Cooperative Oncology Group (ECOG) scale, response to ICI on first radiographical assessment, follow-up, and vital status were extracted from electronic

medical records. We identified patients who received ≤ 3 cycles of therapy and also noted the reason for discontinuation of treatment.

4.2.2 Identifying HPs, responders and non-responders

All CT scans were independently reviewed by two clinicians (PDP and PJ) and a senior radiologist (AG). To be eligible, patients had to have CT images available at a minimum of three time points at least 2 weeks apart—a baseline CT scan immediately before starting immunotherapy, a prebaseline CT scan at diagnosis, and the first post-treatment scan after initiating the ICI. Patients without measurable disease per RECIST V.1.1 or lung lesions on baseline pretherapy scans were excluded from the analysis. RECIST V.1.1 criteria were used to calculate the sum of the largest diameters of the target lesions on each of the CT scans. New lesions on treatment were excluded from calculations.

Response assessment on all patients was performed as per RECIST V.1.1 criteria. Specifically, two target lesions per organ to a maximum of five lesions were identified and measured by a thoracic radiologist (10 years of experience), on prebaseline, baseline, and post-ICI therapy scans for every patient. The four RECIST categories were defined using the target lesions observed on three consecutive CT scans. The sum of diameters of the target lesions (longest for non-nodal lesions and the short axis for the nodal region) were compared for prebaseline, baseline, and post-ICI therapy scans. The four categories were defined as CR, that is, the disappearance of all the lesions, PR, that is, $\geq 30\%$ decrease in the sum of the longest diameters of target lesions compared with baseline; PD, that is, $\geq 20\%$ increase in the sum of the longest diameter of target lesions; and SD, that is, neither PR or PD.[84] For the present study, the responders to ICI therapy were identified as those patients with CR, PR, and SD, and non-responders were patients with PD. To determine hyperprogressive disease, tumor-growth kinetics, which measures the change in tumor size per unit time, were calculated for all the PD lesions using the previously defined methodology.[84] The details regarding the calculations of tumor growth kinetics are described in online supplemental

appendix 1. TGKpre was defined as the change in tumor size per unit of time (months) between the baseline and prebaseline scans.

Meanwhile, TGKpost was defined as the change in tumor size per unit time (months) between the baseline and post-ICI therapy scans. HPs were defined when TGKpost was twice of TGKpre. The details about three class selections are described in the flowchart and online supplemental appendix 1 (pp3).

4.2.3 Radiomic feature extractions

Radiomic features were extracted using software developed at the Center of Computational Imaging and Personalized Diagnostics, Case Western Reserve University (Cleveland, Ohio, USA), implemented on a MATLAB release V.2015 platform.

Textural radiomic features from Gabor, Haralick, Laws, Laplace, and Collage feature families were extracted from each annotated intratumoral and peritumoral region. To ensure the stability and reproducibility of radiomic features, we used the test–retest Reference Image Database to Evaluate Response lung CT dataset (online supplemental appendix 1, pp2).

A set of 74 QVT features were extracted from the segmented vasculature surrounding the nodule. These features pertain to the tortuosity, curvature, and branching statistics as well as the volume of the vasculature. In addition, we measured the angles of each three consecutive points of the vasculature and computed the distribution of these angles33 (online supplemental appendix 1, pp1).

4.2.4 Feature evaluation and classifier construction

First, to evaluate and analyze the structure of an entire dataset, unsupervised clustering was performed on the radiomics feature pool using heatmaps and K-mean clustering. External validation of the clustering analysis was performed by comparing the clustering results against ground truth to identify high-risk patient cohort within the dataset without prior knowledge of biology or outcome.

Second, the supervised experiment was performed by considering the primary endpoint of the study to differentiate HPs from the other response patterns. A radiomics model was developed using the training cohort. For the purposes of feature and classifier selection, 300 iterations of threefold cross-validation were used within the training dataset. The minimum redundancy maximum relevance feature selection algorithm was implemented within the cross-validation setting to select the top-performing radiomic features. Top features were analyzed with box and whisker plots, Wilcoxon rank-sum tests, p values, as well as feature maps.

The five machine learning classifiers implemented within the training set included Random Forest (RF), linear discriminant analysis, diagonal linear discriminant analysis, quadratic discriminant analysis, and support vector machine. The best performing classifier modeled on the training set was then validated on an independent and blinded validation set.

4.2.5 Statistical analysis

The statistical analysis was performed using MATLAB V.2015 and R V.3.5.3. A two-sided p value of <0.05 was considered significant for all the statistical analyses. Differences in distributions between the patient dataset and variables were assessed with the unpaired, two-tailed 2 test, or the Fisher exact test as appropriate.

The selected top features were analyzed using box and whiskers plots, along with the Wilcoxon rank-sum tests. Next, the feature maps of the top features were compared against three classes. To validate the classifier performance, the area under the receiver operating curve (AUC), accuracy, sensitivity (true positive rate; ie, predicting HPs accurately from the rest), and specificity (true negative rate; ie, predicting other classes accurately) were calculated for training as well as validation datasets.

Further, the patients were stratified into two groups based on the labels/output predicted by the locked-down machine learning classifier. The OS was compared against the two predicted groups. The OS was defined as the time between the initiation of ICI to the death

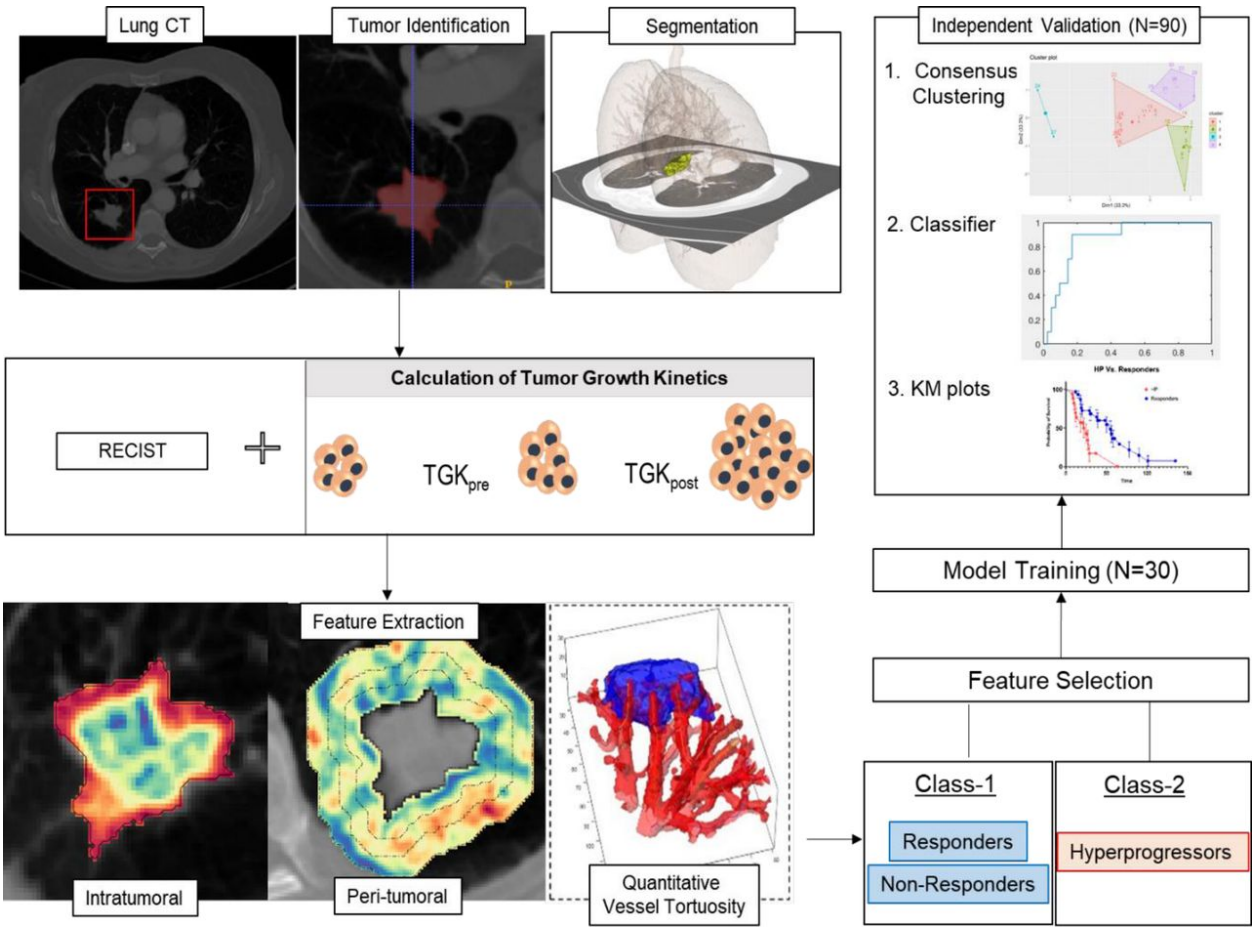

Figure 4.1: Overall workflow and pipeline of the project. The first step involves identifying and annotating lung nodules on prebaseline, baseline, and post-ICI therapy scans. The next step involves calculating TGKs and defining responders, non-responders, and hyperprogressors. After defining the three groups, intratumoral and peritumoral textural features are extracted using MATLAB V.2015a. For the peritumoral region, features were extracted from 0 to 15 mm region outside the tumor and divided into three 5 mm peritumoral rings. Feature statistics included mean, median, SD, skewness, kurtosis, and range and were calculated for each of the individual annular rings. The whole dataset was divided into training and validation sets. Top features were selected using the minimum redundancy maximum relevance feature selection algorithm using the training dataset. A classifier was trained using the training cohort, and the performance was validated using the validation set using clustering, classifiers, and KM plots. ICI, immune checkpoint inhibitor; KM, Kaplan-Meier; TGK, tumor growth kinetic; TGKpre, change in tumor size per unit time (months) between the baseline and pre-baseline scans; TGKpost, change in tumor size per unit time (months) between the baseline and post-ICI therapy scans.

of the patient. The patients were censored if the date of death was unknown. The Kaplan-Meier (KM) survival curves were plotted for two groups using OS. The survival analysis was performed using log-rank tests and HRs with corresponding CIs.

The survival analysis and classification were further performed within individual groups of response patterns (responders, non-responders, and HPs). The classifier performance was compared for differentiating HPs from responders, HPs from other non-responders and responders from non-responders. The three models were compared with classifier AUC as well as HRs and KM curves.

The overall work pipeline is explained in figure 4.1.

4.3 Results

4.3.1 Patient cohort

A total of 524 patients with advanced NSCLC who received ICIs were analyzed in the study, out of which 315 received $\leq$ 3 cycles of an ICI. The patients who either did not have CT images available for analysis, or no available OS or had an unmeasurable disease or had no target lung lesions on baseline scans or poor image quality as determined by a radiologist were excluded from the analysis. The study included the remaining 109 patients after excluding these cases. After implementing RESIST V.1.1 criteria and tumor growth kinetics, a total of 19 HPs were identified. For the cohort with other response patterns (CR, PR, SD, and PD), we identified n=90 based on their first imaging assessment (usually post two cycles) after immune checkpoint blockade (figure 4.2).

The clinical characteristics are listed in table 1. The distribution of patients in the training and testing sets is listed in online supplemental appendix 1 (pp1).

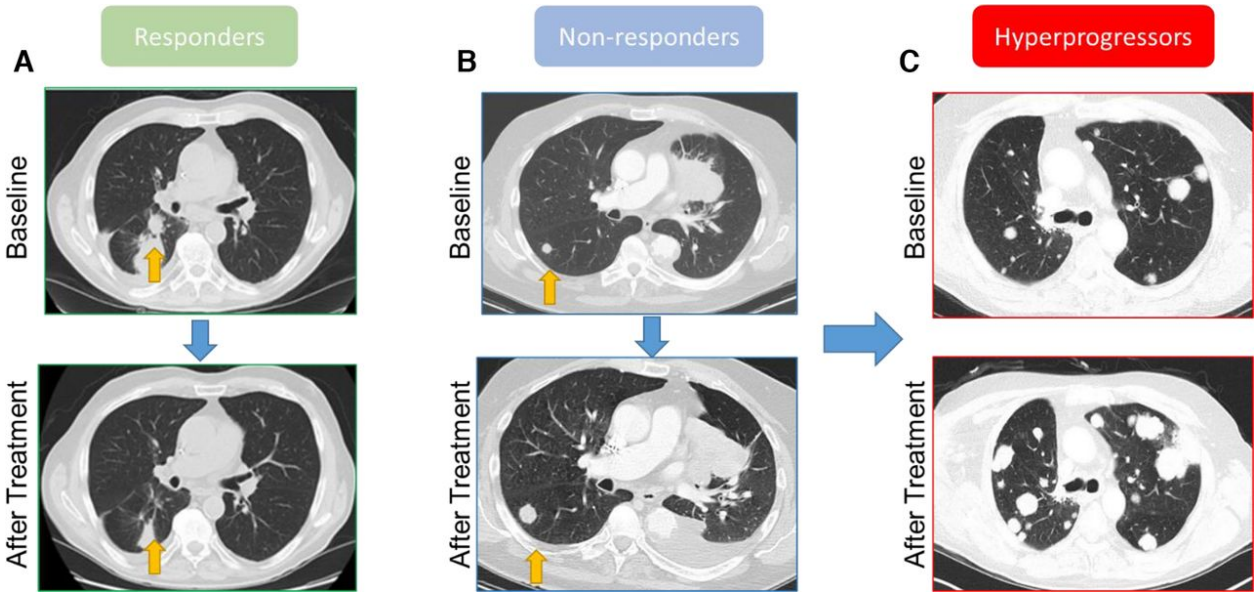

Figure 4.2: CT scans: (A) responders, (B) non-responders and (C) hyperprogressors. The first row shows baseline CT scans and the second row represents CT scans for the same patient after two cycles of immunotherapy.

4.3.2 Experiment 1: analyzing radiomic features

Unsupervised clustering analysis on the feature set

First, the patterns of radiomic features were analyzed using clustering-based heatmap analysis. The approach involves plotting out the individual feature values for each patient within a matrix and performing unsupervised clustering. Heatmap analysis on intratumoral and peritumoral textural patterns resulted in 78.9% of HPs being colocated within cluster 1. Similarly, heatmap analysis with QVT features revealed the HPs aggregating together (figure 4.3).

Next, the entire feature pool, that is, the combination of intratumoral textural, peritumoral textural, and QVT radiomics were used with principal component analysis and k-means clustering to identify four clusters. The compactness within the clusters, that is, how similar are the members within the same group, was observed to be 78.1%. The validation of the constructed cluster was performed using heatmaps and silhouette coefficient (silhouette width). The silhouette plot36 suggests that the clustering using the four groups was optimal with no negative silhouette width and most cluster values of >0.5 (online sup-

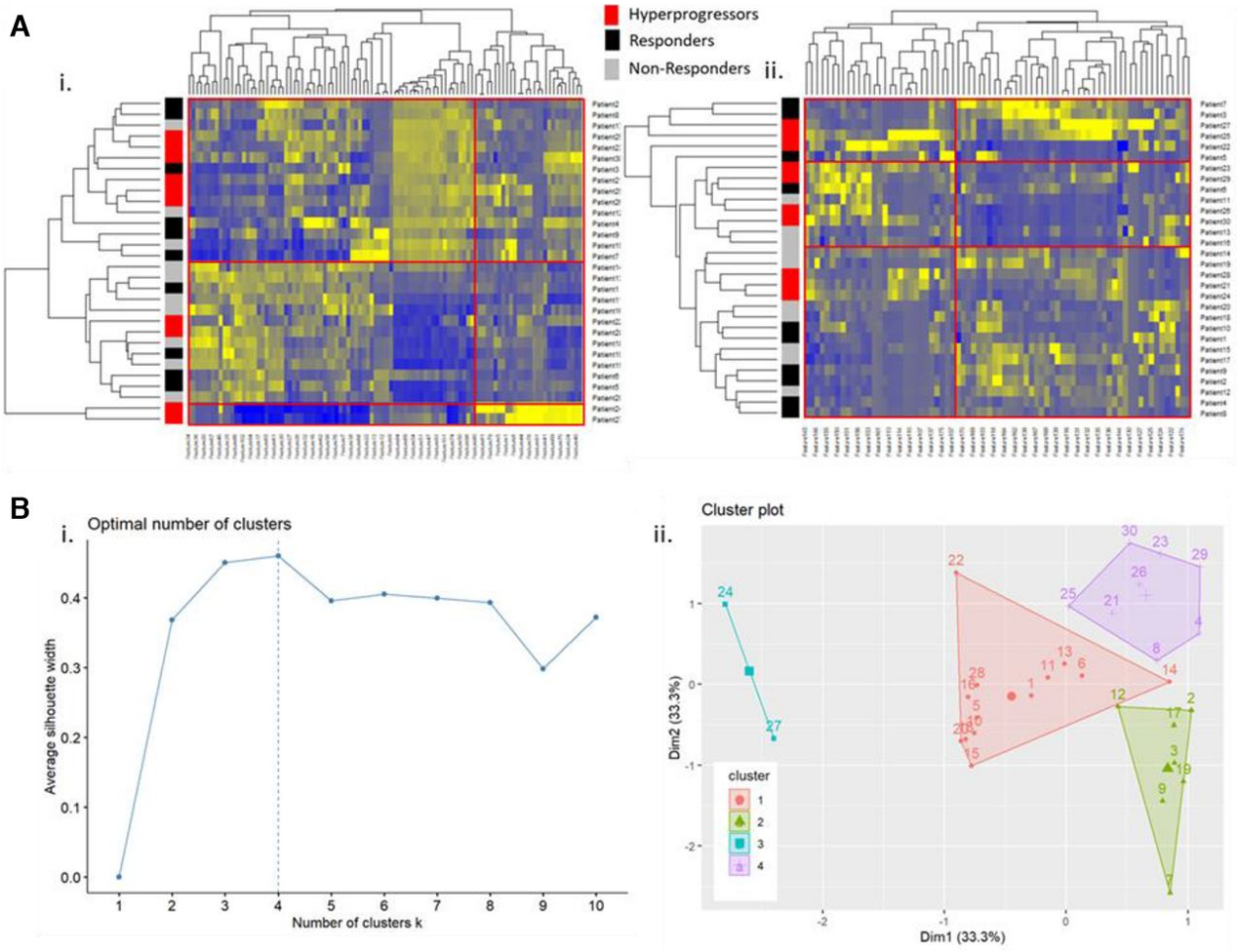

Figure 4.3: Unsupervised clustering analysis: (A) heatmaps for (i) radiomics intratumoral and peritumoral features where 78.9% HPs clustered together in cluster 1, (ii) quantitative vessel tortuosity features. (B) K-mean clustering analysis: (i) elbow curve representing an optimum number of clusters formed using the top three principal components after performing principal component analysis on the entire feature cohort. The optimum number of clusters were observed to be four. (ii) clusters after performing clustering using k=4. These clusters had 78.1% compactness within the clusters. Almost all the HP (80%) were clustered in clusters 3 and 4. Responders and non-responders formed clusters 2 and 1. HP, hyperprogressor.

plemental appendix 1, pp4).

Within the constructed clusters, 80% of combined cluster 3 and cluster 4 comprised 80% of HPs. The remaining 20% HPs were part of cluster 1, making up 15.38% of cluster 1. Twenty percent of responders were clustered together with HPs in cluster 4. The remaining two clusters (clusters 1 and 2) comprised primarily of the responders and non-responders (figure 4.3). Specifically, cluster 2 did not have a single HP within it. The detailed results showed that 53.85% cluster 1 and 42.85% cluster 2 contained non-responders, whereas 30.76% cluster 1 and 57.14% cluster 2 had responders. The unsupervised clustering analysis suggests that high-risk HPs were all clustered together, potentially suggesting that these specific patient groups had a distinct radiomic signature when compared against responders and non-responders.

4.3.3 Experiment 2: supervised classifier for distinguishing HPs from other response patterns

Supervised analysis for selecting the top features The top three features during feature discovery within the training cohort included one peritumoral texture and two QVT features. The peritumoral feature observed was from the Gabor feature family from a 5–10 mm peritumoral region. The other two features were observed from the QVT feature family explaining local curvature and tortuosity of the vessels surrounding the nodule. Hyperprogressive patients were observed to have high feature expressions as compared with responders and non-responders (figure 4.4).

The selected top features were further compared against clinical variables and tumor volume using Pearson's correlation coefficients. The maximum correlation was observed within the Radiomics Gabor feature and tumor volume (rho=0.588).

Table 4.1: Cross validation and hold-out test set performance metrics (AUC, accuracy, sensitivity, specificity) of networks C_A (trained on A_{train}) and C_B (trained on B_{train}) in detecting clinically significant prostate cancer slices, evaluated on three different b-value settings a) B_{4b900}: (0, 300, 500, 900 s/mm^2) b) B_{4b2000}: (0,900,1100,2000 s/mm^2 and c) B_{2b1300}: (0, 1300 s/mm^2). N=112 patients scheduled for prostatectomy underwent two prostate MR examinations (S_A and S_B) performed on the same day approximately 15 minutes apart. The scans, S_A and S_B, were divided into training set (A_{train} and B_{train}), N=78, and test set (A_{test} and B_{test}) N = 34.

b value setting	**Cross-validation**	**AUC (95% CI)**	**Accuracy**	**Sensitivity**	**Specificity**	***p value (AUC)***
B_{4b900}	C_A	0.81 (0.79–0.83)	74%	86%	69%	p = 0.64
	C_B	0.82 (0.79–0.84)	76%	70%	79%	
B_{4b2000}	C_A	0.82 (0.79–0.84)	78%	69%	82%	p = 0.18
	C_B	0.80 (0.78–0.83)	71%	77%	69%	
B_{2b1300}	C_A	0.85 (0.83–0.87)	78%	79%	77%	p = 0.39
	C_B	0.84 (0.80–0.86)	76%	76%	76%	
b value setting	**Hold-out test set (S_{test})**	**AUC (95% CI)**	**Accuracy**	**Sensitivity**	**Specificity**	**p value (AUC)**
B_{4b900}	C_A	0.78 (0.75–0.81)	70%	88%	69%	p = 0.58
	C_B	0.79 (0.76–0.82)	76%	75%	76%	
B_{4b2000}	C_A	0.80 (0.77–0.84)	76%	81%	74%	p = 0.11
	C_B	0.78 (0.75–0.82)	74%	74%	73%	
B_{2b1300}	C_A	0.84 (0.81–0.87)	79%	72%	82%	p = 0.58
	C_B	0.85 (0.83–0.88)	80%	77%	81%	

Table 4.2: Top Features

Feature Family		**Descriptor**	**Location**	**Statistics**
QVT7	Mean Curvature of branches surrounding the tumor	Mean	Surrounding the tumor	Mean
QVT44	Curvature values associated with vasculature	Entropy	Surrounding the tumor	Mean
Radiomics	Gabor	Freq-pi/2 theta-90	5-10mm	Median

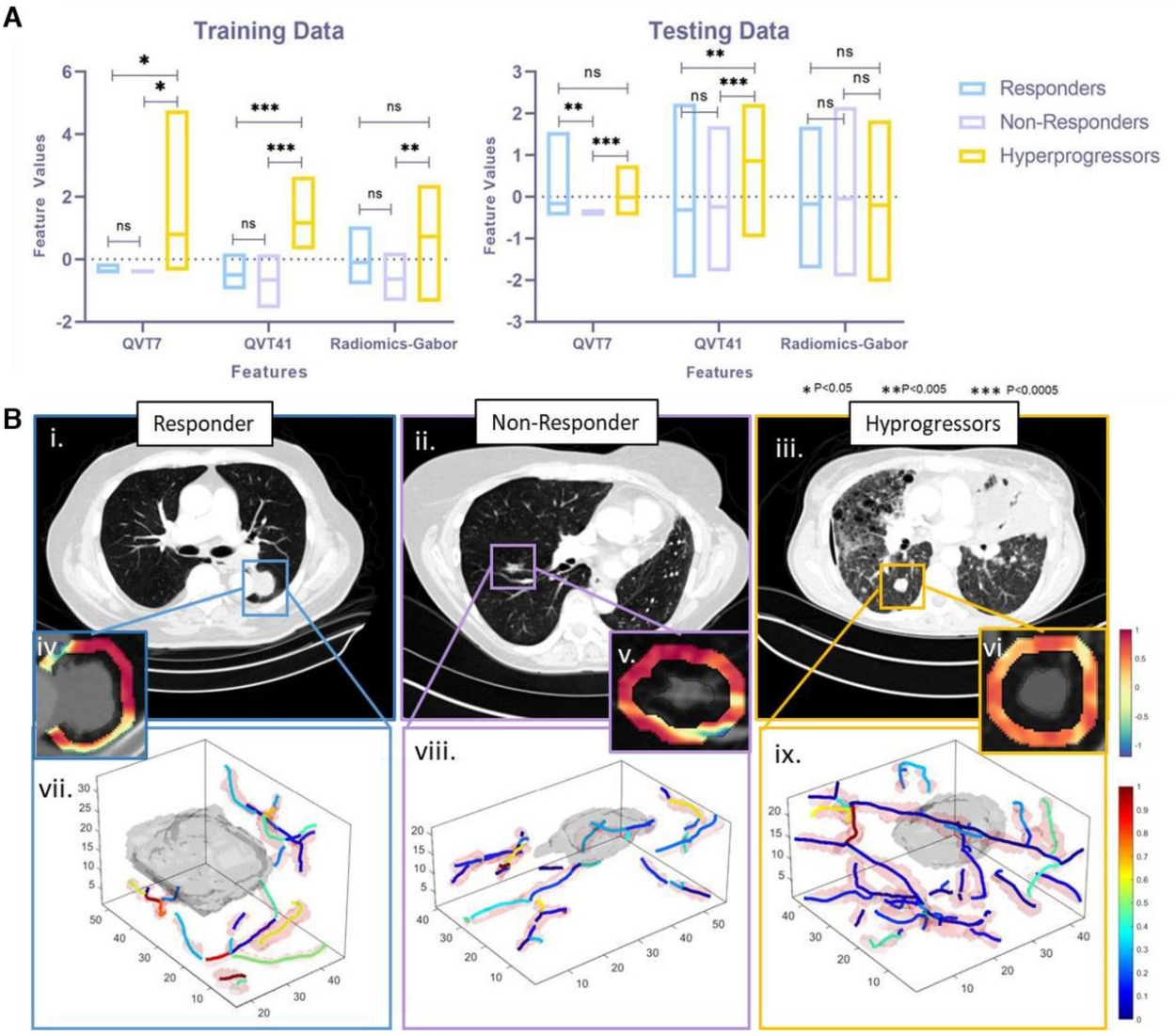

Figure 4.4: Top feature analysis: (A) The box plots for the top three selected features for (i) training cohort and (ii) validation cohort. HPs had statistically significant high-feature values when compared against both responders and non-responders in both pieces of training as well as validation sets. (B) Top feature expression maps with corresponding CT scans for (i) responders, (ii) non-responders and (iii) HPs.Corresponding peritumoral Gabor feature maps are represented in (iv) responders, (v) non-responders and (vi) HPs. similarly, corresponding vessel tortuosity expressions and expressed for (vii) responders, (viii) non-responders and (ix) HPs. HPs were observed to have more convoluted vessel maps. Similarly, radiomic Haralick texture maps represented chaotic peritumoral microarchitecture of HPs. HP, hyperprogressor; (ns, not significant * = $p<0.05$, ** = $p<0.005$, *** = $p<0.0005$).

Supervised classifier performance

Optimal classifier performance with the highest specificity within the training set D1 was achieved using the RF classifier with 50 trees. On the training cohort (D_1, n=30), 300 iterations of threefold cross-validation yielded an average AUC of 0.85±0.06 and an accuracy of 0.86±0.06 for predicting HPs from the other response patterns. The sensitivity, that is, identifying HPs, was observed to be 0.78±0.11, whereas specificity was observed to be 0.91±0.10. The performance of all the five classifiers, along with the feature selection method, is listed in online supplemental appendix 1 pp4. The same classifier was further used for the independent validation set. Within an independent blinded test set, (D_2, n=79), the same RF classifier yielded an AUC of 0.96. The accuracy of the classifier was 0.83, whereas the sensitivity and specificity were 1.0 and 0.81, respectively. The F1 score was observed to be 0.58. In the testing set, all the HPs were identified correctly. Among the remaining cases, eight responders and six non-responders were mistakenly classified as HPs. The confusion matrix is reported in online supplemental appendix 1, pp4.

Within D_2, a subset analysis for differentiating responders against HPs, AUC was observed to be 0.96, and sensitivity and specificity were 1 and 0.91, respectively. The accuracy was observed to be 0.86. For differentiating non-responders against HPs, AUC, accuracy, sensitivity, and specificity were 0.97, 1, 0.86 and 0.89, respectively.

The specific selected top feature set, along with the RF classifier, could not differentiate responders and non-responders within the validation set D2 (AUC 0.43).

4.3.4 Experiment 3: predicted radiomic response groups can also stratify patients with NSCLC treated with ICI based on OS

On D_2, the two predicted groups by the radiomic model, RF classifier, had a statistically significant difference for predicting OS (HR=2.66, 95% CI 1.27 to 5.55, p=0.009). The predicted HPs had significantly lower OS compared with those patients identified as non-

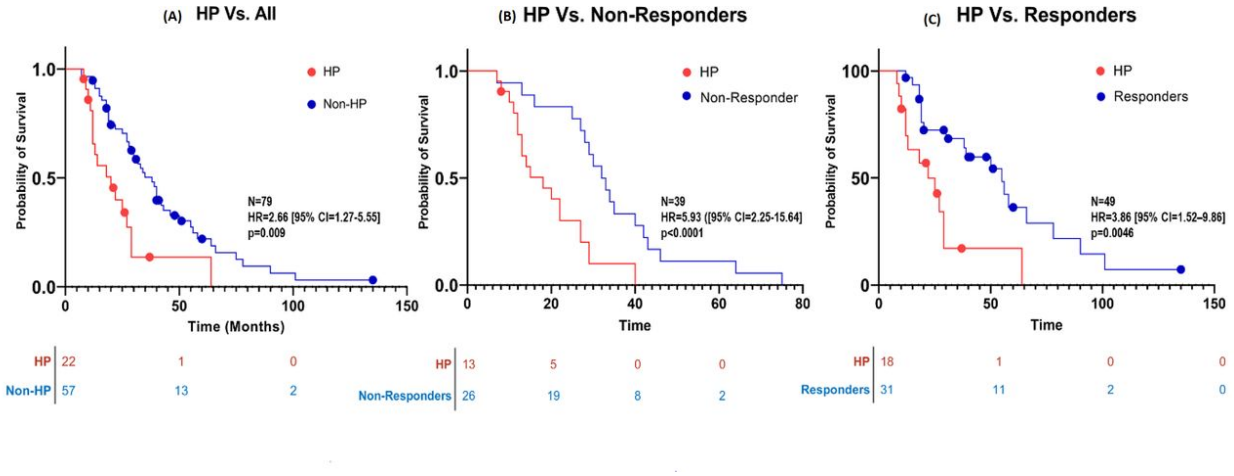

Figure 4.5: Kaplan-Meier survival curves for OS according to predicted labels by random forest classifiers: OS for the test set D2 consisting of patients who were (A) combined responders, non-responders, and HPs. Within predicted two groups, predicted HPs had significantly shorter OS compared with predicted non-HP (B) subset of D_2 consisting only HPs and non-responders and (C) subset of D_2 consisting HPs and responders. HP, hyperprogressor; OS, overall survival.

HPs by the RF classifier. The mean survival time for predicted HPs was 20 months, whereas the predicted survival meantime for non-HPs was 38 months.

Within the subset analysis of D_2, radiomic analysis on predicted HPs had an HR for OS of 3.86 (95% CI 1.52 to 9.86, p=0.0046) when compared with responders alone, whereas predicted HPs had an HR of 5.93 (95% CI 2.25 to 15.64, p¡0.0001) with respect to non-responders (figure 4.5). These results suggest that predicted hyperprogressors had statistically significant worse OS when compared against either responders or even non-responders.

Meanwhile, a comparison between predicted responders and non-responders based on selected top three radiomic features did result in a statistically significant difference between the two groups (HR=1.29, 95% CI 0.49 to 3.35, p=0.59)

4.4 Discussion

The introduction of ICIs has led to a paradigm shift in the management of a vast range of malignancies, including NSCLC. However, ICIs have been associated with atypical response patterns such as hyperprogression, a novel pattern of disease acceleration after the use of

PD1/PD-L1 inhibitors,1 2 which has been reported across different tumor types such as melanoma, head, and neck squamous cell cancer, NSCLC, and urothelial carcinoma, to name a few. It has been postulated that this paradoxical acceleration of disease with ICI may, in fact, be responsible for the higher progression rates and mortality noted early on in patients undergoing immunotherapy in comparison to the chemotherapy, as shown in the phase III trials such as CheckMate 057 that led to the approval of these agents.[85] Due to the lack of a standard definition and different criteria used to define hyperprogression, the reported incidence varies significantly among these studies. For patients with NSCLC, the previously reported incidence ranges between 8% and as high as 25.7%.2 4

In our study of patients with advanced NSCLC treated with a PD1 or PD-L1 inhibitor, we observed hyperprogression in 19 patients. Since most patients receiving ICIs at our institution have a response assessment with imaging performed within 9 weeks or three cycles of therapy, we used this time point as a cut-off for our definition of hyperprogression. Although this varies slightly from previous studies that have used a 2-month or 8-week cut-off, we believe that using the 9-week limit allowed us to capture the true incidence of hyperprogression in our cohort.

The biology of hyperprogressive disease is yet to be understood. In a previous study by Kato et. al.,[60] the authors noted an association with certain genetic alterations such as EGFR mutations and MDM2 amplification.[60] In our study, a majority of the patients did not have any genetic alterations detected on our in-house lung hotspot panel of actionable mutations in NSCLC. Although nine patients did have EGFR sensitizing mutations, we did not observe any significant correlation with hyperprogression. The association of certain clinical variables with hyperprogression such as age,[59] the number of distant metastatic sites[86] has not been consistently confirmed across different studies. There is, therefore, an unmet need for biomarkers that could potentially identify patients at risk of worse clinical outcomes with therapy.

Imaging-based response assessment in patients receiving immunotherapy is fraught with

challenges due to the inability of traditional criteria (RECIST) to account for atypical responses observed with these agents. In some of our previously published work,31 we have identified novel imaging-based radiomic biomarkers to assess responses to ICIs in patients with advanced NSCLC. The foundation of radiomics is that CT images contain a vast amount of information in the form of subtle variations in shape, intensity, gradient, and texture beyond the semantic features that are routinely used by radiologists to describe radiographic appearances of tumors. In the aforementioned studies, we are also tried to understand the morphological/pathological correlates of the predictive radiomic features that correspond to responses to immunotherapy. In doing so, we have identified certain radiomic features that correlate with tumor-infiltrating lymphocyte density on digitalized histopathology specimens from patients treated with ICIs[5962].[65]

While the biological underpinnings of hyperprogression are yet to be uncovered, many of the current theories postulate that dysregulation of various immunoregulatory cells in the tumor microenvironment may be responsible for this phenomenon. One hypothesis that has been put forth is that of 'contrasuppression' or activation and proliferation of regulatory T cells in the presence of ICIs.[86] Other theories suggest that an imbalance in the cytokine milieu and resultant immunosuppression may play a role in paradoxical disease progression with ICIs[70].[87] Lo Russo et al[64] demonstrated the role of tumor-associated macrophage enrichment in immunodeficient mice injected with patient-derived xenografts belonging to patients with NSCLC with hyperprogression after PD1/PD-L1 blockade.

In this study, we evaluated the performance of a new radiomic model using the integration of intratumoral and novel peritumoral texture and vessel tortuosity metrics (QVT) in predicting hyperprogressive disease using only pretreatment CT scans. Our approach is novel being the first Radiomics study that included intratumoral, peritumoral textural patterns along with vessel tortuosity features to predict specific patterns associated with hyperprogressive disease.[88] Our results suggest that these peritumoral texture and vasculature patterns are significantly different in HPs when compared with either responders or other

non-responders.

The peritumoral area is an integral part of the TME. Recent work has suggested that TME embeds information that relates to drug resistance, and the effectiveness of chemotherapy and immunotherapy[695471].[89] Previous investigations of radiomic features from the TME have shown their utility in differentiating lung adenocarcinomas from granulomas, predicting response to neoadjuvant chemoradiation and surgery, pemetrexed chemotherapy in locally advanced NSCLC.14 17 18 Sun et al40 have used a radiomic approach on tumor region and the area surrounding it to detect CD8 cells and used that signature to predict ICI response in multiple retrospective cohorts. For predicting hyperprogressive disease, our results are in line with the recent study by Tunali et al,[88] where top radiomic features reported in the study were also observed from the tumor boundary.[86] Our top radiomic features were also observed from peritumoral regions and represent peritumoral heterogeneity within the textural patterns. These features had higher expression in the baseline CT scans of HPs when compared with responders or non-responders. One of the unique strengths of our study was that in addition to textural feature analysis, we also used a novel approach to quantify the blood vessel morphology (QVT) in the peritumoral area. These QVT features were two of the top three distinguishing radiomic features, emphasizing the importance of peritumoral vasculature in the phenomenon of hyperprogression. QVT features showed more tortuous and disordered vessel architecture for HPs compared with responders or non-responders. We also evaluated the prognostic ability of the radiomic features by performing a survival and classification analysis within the three response categories: responders, non-responders, and HPs. The radiomics classifier correctly predicted that the HP patients would have worse OS when compared with responders and non-responders.

We acknowledge the limitations of our study, many of which are a consequence of the retrospective nature of this study and a limited number of HPs cases. Since the HP phenomenon is quite rare and observed within less than 8% cases, we were limited in the total number of HP cases in the analysis. Further, ours is a single-institution study, but further validation

in independent cohorts of patients is warranted. Since the hyperprogression phenomenon is very rare and observed within less than 8% cases, we were limited in the HP cases. Further, the exclusion of patients without adequate scans for analysis could have affected the incidence of hyperprogression and the performance of our discriminative radiomic signature. While there is no standardized definition of hyperprogression, some previous studies[626386] have used volumetric changes to identify hyperprogressive disease. We acknowledge the limitations of using non-volumetric tumor growth kinetics for identifying hyperprogression, but believe that this would be a methodology that clinicians could easily replicate in practice to identify this subset of patients. PD-L1 expression data were unavailable for a majority of patients in our cohort. The correlation between PD-L1 expression, which is a clinically validated biomarker of benefit from PD1/PD-L1 blockade and hyperprogression, poses an interesting question which could not be addressed in our study.

4.5 Conclusions

Our findings suggest that radiomic analysis of pretherapy CT scans of patients with NSCLC who are being considered for immune checkpoint blockade could be used to identify patients who are at a higher risk of hyperprogression with this treatment. Added benefits of using radiomic analyses include the ability to analyze readily available routine CT scan images and the non-invasive nature of the risk assessment without the need for additional biopsy specimens.

Further rigorous validation in independent cohorts of patients and radiomic–histopathological correlative analyses would further strengthen the argument for using radiomic analyses in routine clinical practice.

Chapter 5

Combining Radiomic and Pathomic Features for Predicting Recurrence and Treatment Response in Non-Small Cell Lung Cancer

5.1 Overview

In non-small cell lung cancer space (NSCLC), current treatment guidelines for early-stage patients suggests complete surgical resection, followed by chemotherapy for the subset of patients, even though there's no definite guideline on which patients would benefit from it. There have been controversies within a few clinical trials, and the recurrence rate for these patients still remains around 55%.[15] In the late-stage NSCLC setting, immunotherapy has changed the treatment regime. Even though it has significantly improved the overall survival rate of these patients, not every late-stage NSCLC patient responds to it. In this work, we have specifically focused on these two clinical cases and combined the corresponding radiology and pathology information to build a novel multiscale AI model.[54]

Within the Radiology space, the Radiomic Features are known to capture the subvisual imaging attributes by looking at the voxel level information of the radiographic scans.[13] The radiomic information extracted from the annotated tumor region as well as from the area surrounding the nodule on the CT scans has shown diagnostic, prognostic, and predictive performance within the NSCLC space[15][90].[46] The various radiomic features contributing to the disease stratifications include textural patterns extracted using various filters (e.g., wavelet, Laws, etc.) as well as various patterns extracted using co-occurrence matrix-based feature patterns[91][92].[93]

With the advancements in high-speed, high-resolution whole slide image scanning hardware, the histological tissue slides can be digitized and analyzed efficiently. Pathomics or quantitative histomorphometric analysis refers to the process of extraction and mining of computer-derived measurements from digitized histopathology images. While the visual reading of routine histopathology slides of tumors by pathologists can help predict cancer behavior to a certain degree, sophisticated pathomics has the potential to "unlock" more revealing sub-visual attributes about tumors.[94] The research community has developed approaches quantifying nuclear arrangement, texture, and orientation for disease presence, risk, aggressiveness, progression, and survival, thus potentially providing a comprehensive portrait of the tumor's morphologic heterogeneity[95][96].[52]

Even though there have been significant efforts within Radiomics and Pathomics within the NSCLC, there have been relatively few attempts to combine radiology and pathology[95].[97] Several recent studies show that the multimodal datasets carry complementary information and could improve the disease-specific AI models[95].[97] Specifically, within the lung cancer domain, for each individual patient, the radiology and pathology data are usually readily available. It stands to reason that one could potentially combine these two modalities and create an integrated, accurate model.

For early-stage cases, surgical specimens are available for each patient, as well as they have the preoperative CT scans, which are usually used for the initial diagnosis. These two

modalities, presenting information on the different scales, could be potentially combined to predict the recurrence of these patients and, subsequently, the added benefit of adjuvant chemotherapy.[15] Similarly, for late-stage lung cancer cases, the availability of both histology and radiology scans could improve the prediction of patient responses to the immune checkpoint blockade.[54]

In this study, we have combined Radiomic from CT scans, and Pathomics from whole-slide tissue scans to predict the added benefit of adjuvant-chemotherapy for early-stage NSCLC patients and predict response to immune checkpoint inhibitors for late-stage NSCLC patients, both clinically relevant question with respect to lung cancer. The radiomic features were extracted from the tumor region and from the region immediately surrounding the nodule. The pathomic features included features explaining the shape, texture, and orientation of nuclei, as well as features explaining the interplay between cancerous and non-cancerous clusters on the whole-slide tissue scans. We specifically focused our analysis on interpretable features. The combined radio-pathomic model was constructed by combining the best radiomic and pathomic feature sets. This is one of the largest radio-pathomic studies in the field of lung cancer, spanning 250 patients across multiple NSCLC stages. Our results show that the combination of radiomic and pathomic provides us with an opportunity to look at the complementary information across multiscale and provides better risk stratification and response prediction, and improves treatment predictions. To compare our fusion approach, we implemented individual Radiomic and Pathomic-based models for these two applications and compared the AUC, accuracies, sensitivities, and specificities. Finally, we have also calculated Hazard Ratios for all models and compared them against each other.

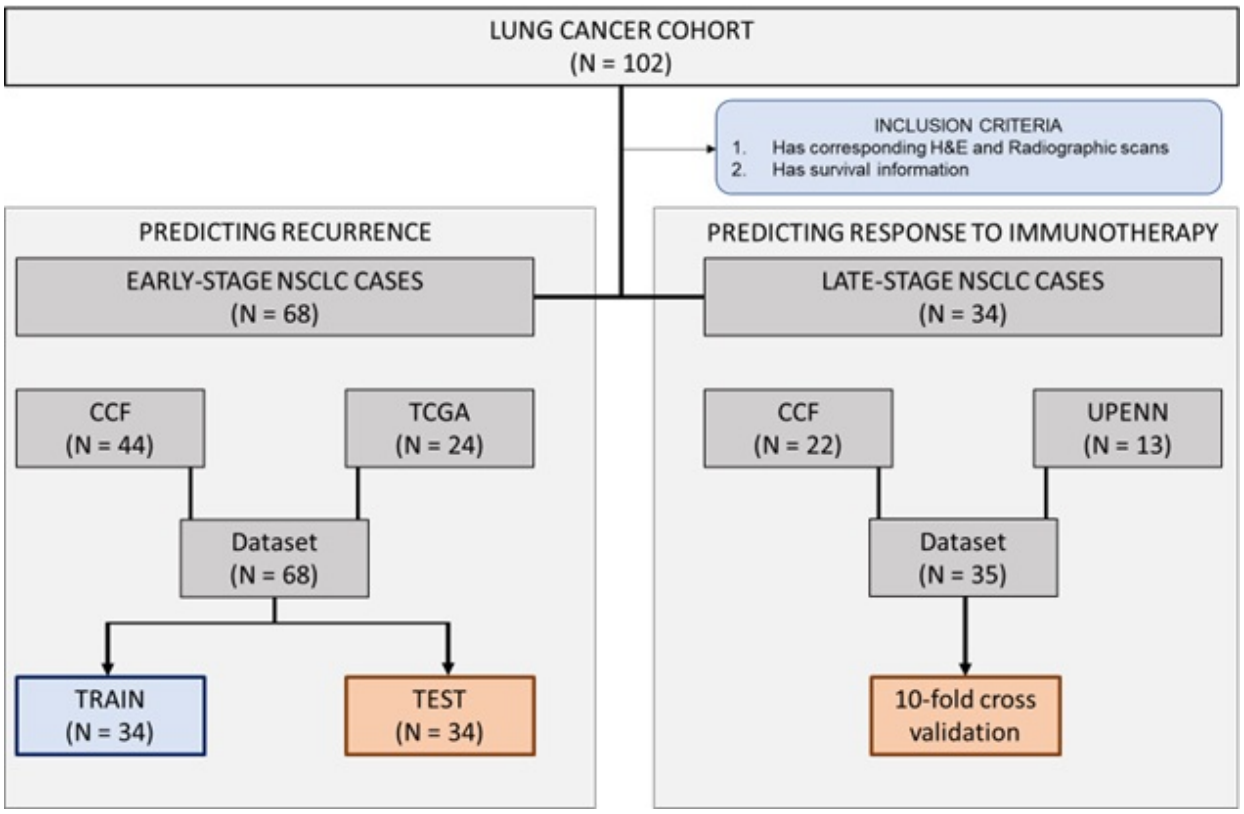

Figure 5.1: Dataset distribution for the lung cancer cohorts. All the above cases had CT scans and corresponding histopathological scans. The cases were used within two problem statements- predicting recurrence and predicting response to immunotherapy.

5.2 Methods

5.2.1 Dataset

This study included CT scans and HE-stained whole slide images for a total of N=228 patients. Figure 1 explains the patient distribution diagram. Of these patients, 193 were from stage I and stage II, whereas 35 were from stage III and IV. All these patients had CT scans, and corresponding HE-stained whole slide tissue scans along with the required clinical information. The cases from stage I and stage II were used for predicting recurrence and the added benefit of adjuvant chemotherapy. The cases from stages III and IV, a total of 35 patients, were used to predict patients' responses to immunotherapy.

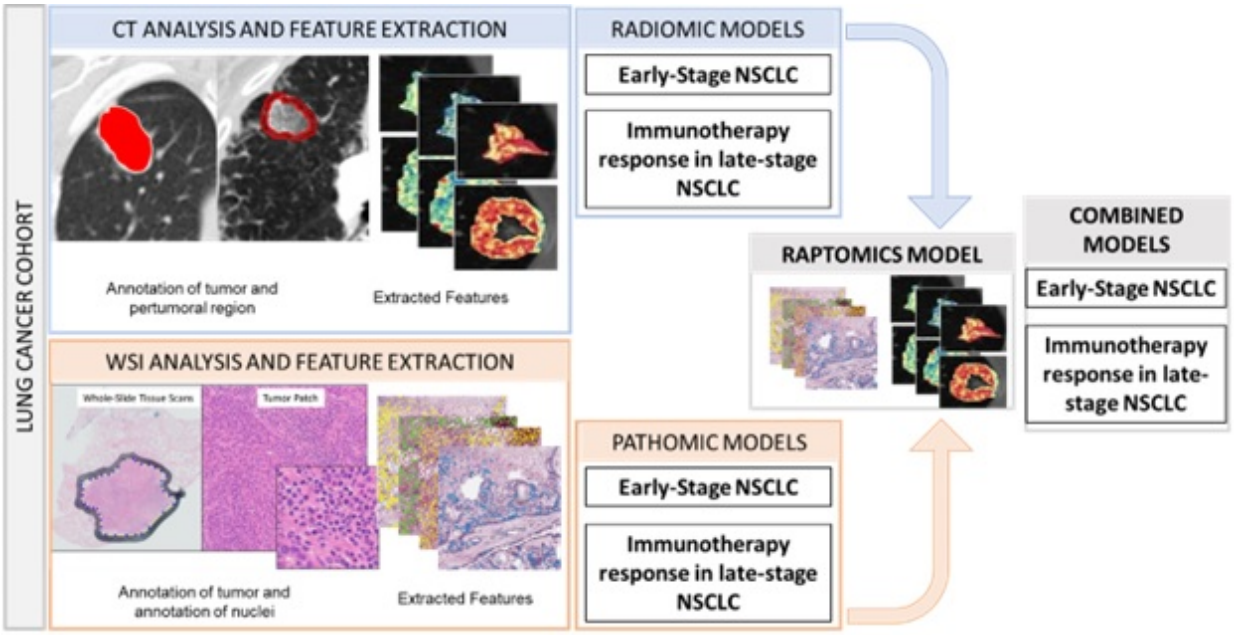

Figure 5.2: The pipeline for the Raptomic Models. For CT scans, Radiomic features were extracted from intra and peritumoral regions for building models to predict recurrence in early-stage NSCLC (prognostic model) and predict response to immunotherapy (predictive model) in late-stage NSCLC. HE stained scans were used for pathomic analysis for extracting pathomic features and constructing prognostic and predictive models. For RaPtomic analysis, top radiomic and pathomic features were combined together, and the best combination was selected for prognostic and predictive models.

5.2.2 Radio-Pathomic Analysis

The study's primary goal was to validate whether radiomic and pathomic feature fusion improve lung cancer clinical decision-making outcomes. We used two test cases to validate the hypothesis. First, we used the combined radio-pathomic feature set to predict response to adjuvant chemotherapy for early-stage NSCLC, and second, the combined feature set was used for predicting response to immunotherapy response for late-stage NSCLC. The following diagram (figure-2) explains the overall workflow of the analysis. The Radiomic features were extracted from the CT scans, and simultaneously, pathomic features were extracted from the corresponding whole-slide tissue scans. The top features from radiomic and pathomic were combined for combined radio-pathomic analysis in the machine learning classifier. The performance was compared against individual radiomic and pathomic models.

5.2.3 Individual Radiomic Analysis

CT acquisition and segmentation

The lung nodules from CT scans were segmented using an expert radiologist using 3D slicer software, where a freehand tool was used to segment the lesion on each 2D section manually. Further, the area around the tumor region was segmented using MATLAB software15 via morphological erosion and dilation operations. Five annular rings peritumorally were analyzed, each with a 3 mm increment leading up to a maximum radius of 15 mm from the nodule boundary. Radiomic features were extracted from the tumor and peritumoral regions from the 3 slices with the largest tumor areas.

Extracted Radiomic Features

Quantitative features were extracted from within and outside the annotated lung tumor regions. The extracted features included features from Gabor, Haralick, laws, and collage feature families. The first-order statistics (mean, median, SD, skewness, kurtosis, and variance) of each of the features were computed.

Reproducibility of Radiomic Features

Radiomic feature reproducibility and correlations were considered before selecting the subset of Radiomic Features to work on. First, the most correlated features (correlation ¿0.9) were removed from the analysis. Next, the reproducibility of radiomic features was selected using the RIDER test re-test dataset17. This dataset contains two CT scans of the same patient taken 20 mins difference apart. The intra-class correlation coefficient (ICC) was calculated for each feature and features having an ICC greater than 0.8 were retained for the analysis. The final uncorrelated, reproducible feature set was used for analysis.

5.2.4 Individual Pathomic Analysis

WSI, Tile Construction and Segmentations

Whole slide tissue images of patients were used for pathomic analysis. The analysis was done on surgical specimens that had been digitally scanned at 20x magnification. Pathomic features were extracted patches of 1000 pixel×1000 pixels within tumor annotated regions on the whole slide images. A U-Net-based deep learning-based model convolutional neural network (CNN) approach was used to identify tumor regions on the whole-slide tissue scans. Next, another deep learning CNN-based automated nuclei segmentation approach was used to accurately segment nuclei on each whole slide image. The approach ultimately resulted in identifying each nucleus, along with its corresponding boundary. The detail regarding each model is explained in appendix 1. Next, each nuclei on the scans were divided into cancerous and non-cancerous nuclei using a machine learning-based model. These annotations and sets were used for pathomic analysis.

Extracted Pathomic Features

Various pathomic features related to nuclei shape, orientation, and texture were extracted for pathomic analysis. The segmented nuclei were also used for extracting various local and global graph-based features. The cancerous and non-cancerous nuclei clusters were used for extracting features explaining spatial arrangements between these two groups.

5.2.5 Statistical Analysis

For individual Radiomic, Pathomic models, the total radiomic feature set and a total of pathomic feature sets were used respectively.

The early-stage cohort was divided into training and validation sets, where the training cohort was used for selecting the top features and training the model, and the performance was validated on an independent validation cohort. The feature selection pipeline followed

the integrated mRMR feature selection method followed by LASSO analysis18.

The best features, along with their LASSO coefficients, were used to build a Radiomic Risk Score (RRS) and Pathomic Risk Scores (PRS) for each patient. Finally, RRS and PRS were combined and used for the combined Radio-Pathomic analysis. The improved model performance was validated using hazard ratios (HRs), and C-indexes were used to compare the improved model's performance.

For the late-stage NSCLC cohort, all patients had received immunotherapy, and the patients who had a complete response (CR) or partial response (PR) were considered immunotherapy responders, and patients with stable disease (SD) or (PD) were considered non-responders19. For individual radiomic and pathomic analysis, the features were selected using the mRMR feature selection algorithm followed by linear discriminant analysis (LDA) classifier construction and simultaneously validated on an independent fold. The process was repeated 100 times to get the more accurate estimations, and the results were calculated with mean +/- standard deviation.

For the combined radio-pathomic analysis, the best Radiomic and Pathomic features were combined and used within a 10-fold cross-validation setting to select the best radio-pathomic feature set, and the performance was validated with an LDA classifier.

5.3 Results

5.3.1 Study Population and Characteristics

Table 1 below shows the details regarding the patient characteristics for the prognostic early-stage cohort and predictive late-stage cohort. Within the early-stage cohort, 39.7% of patients had a recurrence, whereas 60.3% of patients did not recur. Within the late-stage lung cancer cases, 38.2% had stable disease, and 61.8% did not. Within the late-stage NSCLC cohort, 38.2 responded to immunotherapy, and 61.8% did not.

5.3.2 Experiment-1- Predicting recurrence and added benefit of adjuvant chemotherapy in Early-Stage Non-Small Cell Lung Cancer

Within the radiomics model, the top selected features included haralick and collage features from inside the nodule and 0-3mm surrounding the tumor region. The details of the top features, along with their LASSO coefficients, are reported in appendix 1. The Radiomic Risk Score (RRS) was constructed by using the weighted combination of the Radiomic feature value with their corresponding coefficient. The constructed RRS had a c-index of 0.625 (se 0.062; p = 0.006) on an independent validation cohort of D1V1, which included patients who only received surgery and no additional treatment. The threshold of 0.0401 was selected using the training cohort D1T, and on the validation cohort D1V1, the same threshold gave HR of 3.48 and a statistically significant difference between the two groups (p=0.005). For the validation cohort D1V2 there was no difference between high-risk and low-risk groups and the c-index was 0.45 (p=0.56), potentially suggesting that this groups which had patients who received adjuvant-chemotherapy following surgery potentially had a subset of patients who benefitted from the treatment and better survival.

Within the pathomics models, the top features included features were from the SPATIL feature family. The constructed pathomic risk score (PRS), constructed using weighted combinations of LASSO coefficients, had a C-index of 0.657 (p=0.03) on an independent validation set D1V1. The threshold of 0.0142 was selected using the training cohort D1T, and on the validation cohort D1V1, the same threshold gave HR of 4.39 and a statistically significant difference between the two groups (p=0.0089). On D1V2 we observed the same performance as that of Radiomics Model. There was no statistically significant difference between Kaplen-Meier curves and the c-index was 0.55 (p=0.345).

Finally, these two signatures were combined. Individually, these two signatures have no correlation between them, and while combining, the Radiomics signature contributed

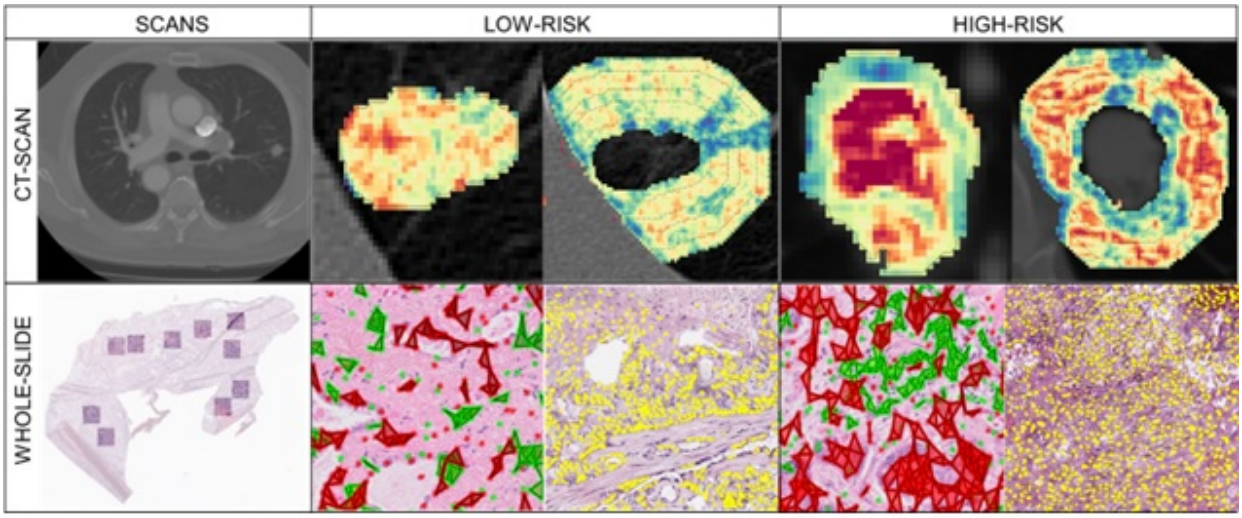

Figure 5.3: RaPtomic Feature maps for early-stage NSCLC patients. The best features from Pathomics included features from SPATIL feature families, and from Radiomics, the best features included features from Haralick Feature Families. The best features seem to have higher expressions in recurrent cases compared to non-recurrent ones

40%, whereas the Pathomic signature contributed 60% into the final RaPtomics model. The C-index on an independent validation cohort improved to 0.715 on D1V1 showing a total improvement of 3% compared to radiomics and 7% compared to pathomics based models. The hazard ratio was 4.8 for the combined RaPtomics model (p=0.0016).

Figure-3 shows the feature maps for radiomics an pathomics for low and high-risk patients. Figure-3 shows the Kaplan-Meier survival curves for Radiomics, Pathomics, and RaPtomics-based models for D_1V1 and D_1V2.

5.3.3 Experiment-2- Predicting Immunotherapy response in late-stage non-small cell lung cancer

For Radiomics Model, the top selected features included features from Gabor and Laws feature families. The top features are reported in appendix 1. The AUC of the model was 0.73, and accuracy, sensitivity, and specificity were 0.82, 0.87, and 0.73, respectively.

Within the Pathomics model, the top selected features included features from the shape-based feature families and SPATIL feature families and had an AUC of 0.69 within the

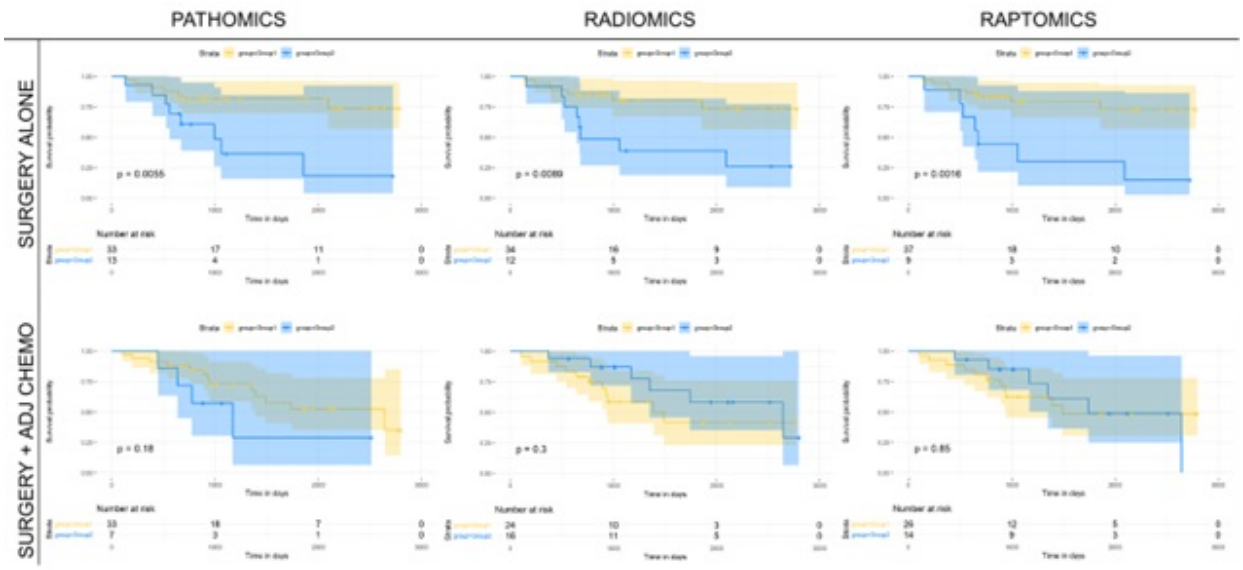

Figure 5.4: Kaplan-Meier survival curves for the independent validation cohorts D_1V1 and D_1V2 , The constructed signature showed good separation for surgery alone cohort, but within patients who received adjuvant-chemotherapy following the surgery, there was no difference between the two groups, potentially suggesting that this group probably had a subset of patients who benefitted from chemotherapy and had a better survival.

Table 5.1: Classifier comparison between Radiomic, Pathomic, and RaPtomics models for predicting recurrence in early-stage NSCLC and predicting response to immunotherapy in late-stage NSCLC.

	AUC	**ACCURACY**	**Sensitivity (%)**	**SPECIFICITY**
D_{Pathomic}	0.69	0.73	0.98	0.33
D_{Radiomic}	0.73	0.82	0.87	0.73
$D_{\text{Combined Radiomic-Pathomic}}$	0.77	0.85	0.90	0.76

10-fold cross-validation setting. The accuracy, sensitivity, and specificity of the model were 0.73, 0.98, and 0.33, respectively.

For radio-pathomic analysis, top radiomic and top pathomic features were combined together to be used as a RaPtomics feature set. Within this combined feature set, after performing 10-fold cross-validation, top features included both combined Radiomic and pathomic Features, and the overall performance of the model improved to 0.77, which was 4% higher than Radiomics and 8% higher than individual pathomics based models.

5.4 Discussion

In this work, we showed that combining Radiomics and Pathomics features improve the prediction accuracy for lung cancer patients over models that uses monomodal data alone. We showed two test cases within the non-small cell lung cancer (NSCLC) – first, within the early-stage domain, we showed that combining radiomics and pathomics improves the recurrence predictions, and within the late-stage domain, we showed that combining radiomic and pathomic improves the prediction of which patients would benefit from immunotherapy.

Within the radiology and pathology fusion analysis, few deep-learning-based models have been proposed within the lung space 20, but the interpretability of these models remains limited due to the black-box approach used within the deep learning-based analysis. Compared to deep learning approaches, the hand-crafted features extracted from histology images and radiology images provide better explainability since the features were pre-defined, either in a domain agnostic[98][99] or domain inspired[100] way. Radiomics refers to quantitative measurements of texture and shape attributes extracted using advanced image processing and computer vision techniques from imaging modalities.[101] Whereas Pathomics or quantitative histomorphometric analysis refers to the process of extraction and mining of computer-derived measurements from digitized histopathology images. Within the hand-crafted feature fusion approach, Feng et al.[102] and wan et al.[103] have shown the ability to combine radiographic features from MRI scans and pathology-based features from HE-stained biopsies for advanced rectal cancer for predicting complete pathological response and response to neoadjuvant chemoradiotherapy. They noticed an improved performance within AUC at each grade of pathological response with integration when compared against individual models. To the best of our knowledge, within the lung space, even though there is the easy accessibility of radiographic and histology scans for each patient, none of the previous studies has explored its combined effect.

In the first experiment, we have combined Radiomics and Pathomics for early-stage NSCLC patients for whom the surgical resection is currently the standard of care.[36] Adjuvant

cisplatin-based chemotherapy is currently recommended for patients with stage II NSCLC cases, whereas patients with stage I NSCLC continue to be treated with surgery alone[3837] even though about 30–40% of patients with stage I NSCLC recur with post-operation observation alone. The inconsistencies between various clinical trials suggest the need for a biomarker that could identify early-stage patients who would potentially recur and benefit from the added benefit of adjuvant chemotherapy[42?41].[40] In our work, we show that individually Radiomics and Pathomics models are able to predict high-risk populations who would benefit from adjuvant chemotherapy (Pathomic: C-index 0.647 (se=0.083); Radiomic: C-index 0.622 (se=0.084)). After combining these two signatures together, the model performance improved to 0.69 (se 0.07) within the independent validation cohort DV1. For the population cohort from DV2 where patients received adjuvant chemotherapy following surgery, we did not notice a significant difference between low and high-risk groups, potentially suggesting that this population had patients who benefitted from chemotherapy and had better survival as observed in the previous studies (Pathomic 0.546 (se 0.095) Radiomic 0.587 (se 0.07) RaPtomic 0.483 (se 0.084)).

The top radiomic features for early-stage NSCLC analysis included features from Haralick and CoLIAGe feature families from the tumor region and the region immediately surrounding the nodule. Our results align with those previously published within the Radiomics and early-stage domain, where authors noticed the top features observed from co-occurrence-based feature families. Within the pathomic analysis, the top features were noticed from SPATIL feature family, where the features characterizing the spatial arrangement33,34 of tumor infiltrated lymphocytes (TILs) and interplays between lymphocytes and cancer cells[96].[104] The recent works suggest the utility of these SPATIL features not only within the lung but also within various other cancers as well.[105]

Stage III and stage IV accounts for late-stage NSCLC cases. The addition of immune checkpoint inhibitors (ICIs) to the armamentarium of cancer therapies has resulted in unprecedented improvement in survival outcomes for late-stage patients37. However, the re-

sponse rates to these drugs remain modest (27% in PD-L1 NSCLC in the first-line setting38 and 45% in the PD-L1high subgroup39, and 19% in the second-line setting) . In addition, optimal patient selection for therapy remains a challenge because most available biomarkers, such as PD-L1 expression, are insufficient to classify patients accurately, generating the need for a consensus biomarker to identify patients who would benefit from these therapies. In our analysis, we noticed that Radiomics and Pathomics individually were able to predict a subset of the population who would benefit from ICIs. After combining both the feature sets together, the performance of the model improved to 8% over pathomic and 4% over radiomics. The combined model had 2 features from Radiomics and 2 from Pathomics.The Radiomic featrues were notices from the Gabor and Haralick Family from the peritumoral region and from the Pathomic feature set, the top features were noticed from the nuclei-shape based features. The chaotic and disturbed microarchitecture for immunotherapy non-responders were noticed both on radiomic and pathomic feature sets.

Our analysis has some limitations. First, we had limited datasets. In the future, we plan to validate the models on extensive on multiple external validation cohorts along with validating radiomic features on multiple different scanners, slice thicknesses and reader segmentations. Second, there are multiple ways in which Radiomics and Pathomics could be fused together. We adapted a single fusion method, but we did not compare our fusion strategy with other methods. In the future, we plan to check multiple fusion strategies to combine radiology and pathology datasets. Future directions for this work also include linking radiomic features from corresponding digital pathology surgical tissue sections and genomic measurements to more comprehensively characterize the tumor.

Even after the limitations, we believe our analysis shows promising results for combining radiology and pathology to improve the prediction accuracies within the non-small cell lung cancer domain problems.

Chapter 6

Integrated Nomogram for COVID-19 Prognosis

6.1 Overview

The coronavirus disease 2019 (COVID-19), caused by severe acute respiratory syndrome 2 (SARS-CoV-2), is an ongoing global pandemic with over 3.93 million deaths and 181 million total diagnosed cases worldwide so far[106].[107] The new COVID-19 delta variant, recently diagnosed and spreading across the world, has the ability to cause very dense outbreaks[108].[109]

The majority of COVID-19 patients present with mild disease to an outpatient clinic or via telehealth with minor clinical symptoms. A lesser proportion of the patients develop moderate to severe disease with significant pulmonary dysfunction or damage as evidenced by signs of hypoxemia and moderate to severe dyspnea.[107] According to one study, 20% of diagnosed COVID-19 cases have severe or critical diseases, and about 8% of them require intensive care management with or without mechanical ventilation.[110] If we can diagnose this high-risk population at the earliest stages, it will likely allow for optimal resource management and individualized treatment planning[111].[112]

Imaging plays an essential role in the management of COVID-19 patients, with chest CT

being the preferred modality for these patients.[113] However, despite the high sensitivity of chest CT, the reported specificity is quite low at about 25–33%, which is due to considerable overlap in CT imaging features of COVID-19 and other viral types of pneumonia.[114] This, coupled with other challenges, such as transmission risk to uninfected health care workers and other patients, consumption of PPE, and need for cleaning and downtime of radiology equipment in resource-constrained environments, has led to the recommendation by multiple professional societies against usage of CT as a routine screening test for COVID-19 but reserved for only selected clinical scenarios.

Furthermore, a variety of prediction models have been reported for diagnosing and prognosticating COVID-19, including a combination of clinical and lab data as well as imaging features[115][116][117][118].[119] According to a systematic review, flu-like symptoms and neutrophil count are more predictive in diagnostic models, while comorbidities, sex, C reactive protein, and serum creatinine levels are the frequently reported prognostic factors.[120] Most of the AI analysis has focused on chest x-rays (CXRs)[106],[107] though more recently, more and more works on AI for CT scans have also been published. In this work, our focus has been solely on CT scans, and especially machine learning-based models. However, many of the proposed models are poorly reported and are at high risk of bias, and at present, it is not recommended to use any of the reported prediction models for use in clinical practice.[120]

Therefore, there is an unmet need to develop non-invasive tools, preferably based on existing imaging techniques and available clinical parameters, that can help prospectively identify patients at higher risk for developing severe disease phenotype. The ability to identify these patients who will probably need mechanical ventilation and develop severe symptoms will allow us for optimal use of existing precious resources.

In the past few years, high-throughput computer extracted features from the radiographic images (radiomics) has been useful for a variety of diagnostic, prognostic, and predictive applications across several cancers as well as other diseases[13][121].[14] These features are known to capture the underlying tissue morphology and characteristics, which are not visually

apparent to the naked eyes[15].[16] Within the COVID-19 space, radiomics has been used for various applications. Radiomics has been successful in differentiating COVID-19 patients from other pneumonia cases (diagnostic), as well as has shown application for predicting the severity of COVID-19 patients (prognostic).

In this work, we aim to combine the clinical and laboratory parameters with imaging data to build an accurate and easy-to-use nomogram to predict the need for mechanical ventilation for COVID-19 patients. The imaging data includes radiomic features extracted from the regions corresponding to COVID consolidation on CT scans; these regions of consolidation were automatically segmented using the U-Net-based model, making the whole end-to-end pipeline completely automated. Our model has been validated on roughly 1,000 patients from two different institutions making this one of the largest radiomic-based prognosis predictions for COVID-19 studies to date.

6.2 Methods

6.2.1 Study Population

The Institutional Review Board Committee approved the retrospective chart review study of record at the University Hospitals, Cleveland (STUDY20200463), and the Renmin Hospital of Wuhan University (ethics number: V1.0; IRB number 2020KS02010). The need for written consent was waived. Following the inclusion and exclusion criteria, the study included D_1 (N = 787) patients from the hospital of Wuhan University, Hubei General Hospital, and D_2 (N = 110) patients from University Hospitals, Cleveland. The details regarding the inclusion–exclusion criteria and patient flowchart are mentioned in Figure 1.

Stratified random sampling was performed to split the data from institution-1 into 60% training ${D_{\mathrm{T}}}^1$ (N = 473) and 40% testing ${D_{\mathrm{V}}}^1$ (N = 314). While randomly dividing the data, the COVID patients being on the ventilator were kept approximately similar within training and testing cohorts (The training cohort had 64% of the COVID patients being

on ventilator, whereas 55% of the COVID patients did not use the ventilator. Similarly, the testing set had 36% of the COVID patients who used the ventilator, and 45% of the COVID patients did not use the ventilator).

The patients from institution-2 were used for independent external validation $D_{\mathrm{V}}{}^{2}$ (N = 110). The patients were acquired by following the chart review for patients who were seen between January and September 2020.

6.2.2 Radiomic Feature Analysis

Detection and Segmentation of Lung Lesions

An expert radiologist with 14 years of experience delineated ground-glass (GGO) and consolidation regions on a subset of $D_{\mathrm{T}}{}^{1}$ [$D_{\mathrm{T}}{}^{UNET}$ N = 88 (training cohort) and $D_{\mathrm{V}}{}^{UNET}$ N = 96 (validation cohort)]. The UNET-based model to segment the COVID consolidations on CT scans was trained within a threefold cross-validation setting using DTUNET, and the performance was validated on $D_{\mathrm{V}}{}^{UNET}$. A CNN with U-Net architecture was employed to segment out ground-glass opacities (GGOs) and consolidations in the lung region on the baseline chest CT scans (29). An automatic lung segmentation method utilizing a watershed transform was used to segment out and crop the CT volume around the regions of the lung (30). Each 2D slice of the cropped volume was resized to a size of 256 by 320. Furthermore, the 2D slice was vertically divided into two parts dividing the right and left lung regions (input size: 256 by160), and parts of the lung region (right, left) were given as separate inputs (input size: 256 by160). The two vertical slices from each 2D input were used as inputs to the UNET model to segment COVID consolidations.

Following figure 6.2 explains the architectural diagram of the 2D U-Net used for segmentation of GGOs and consolidations.

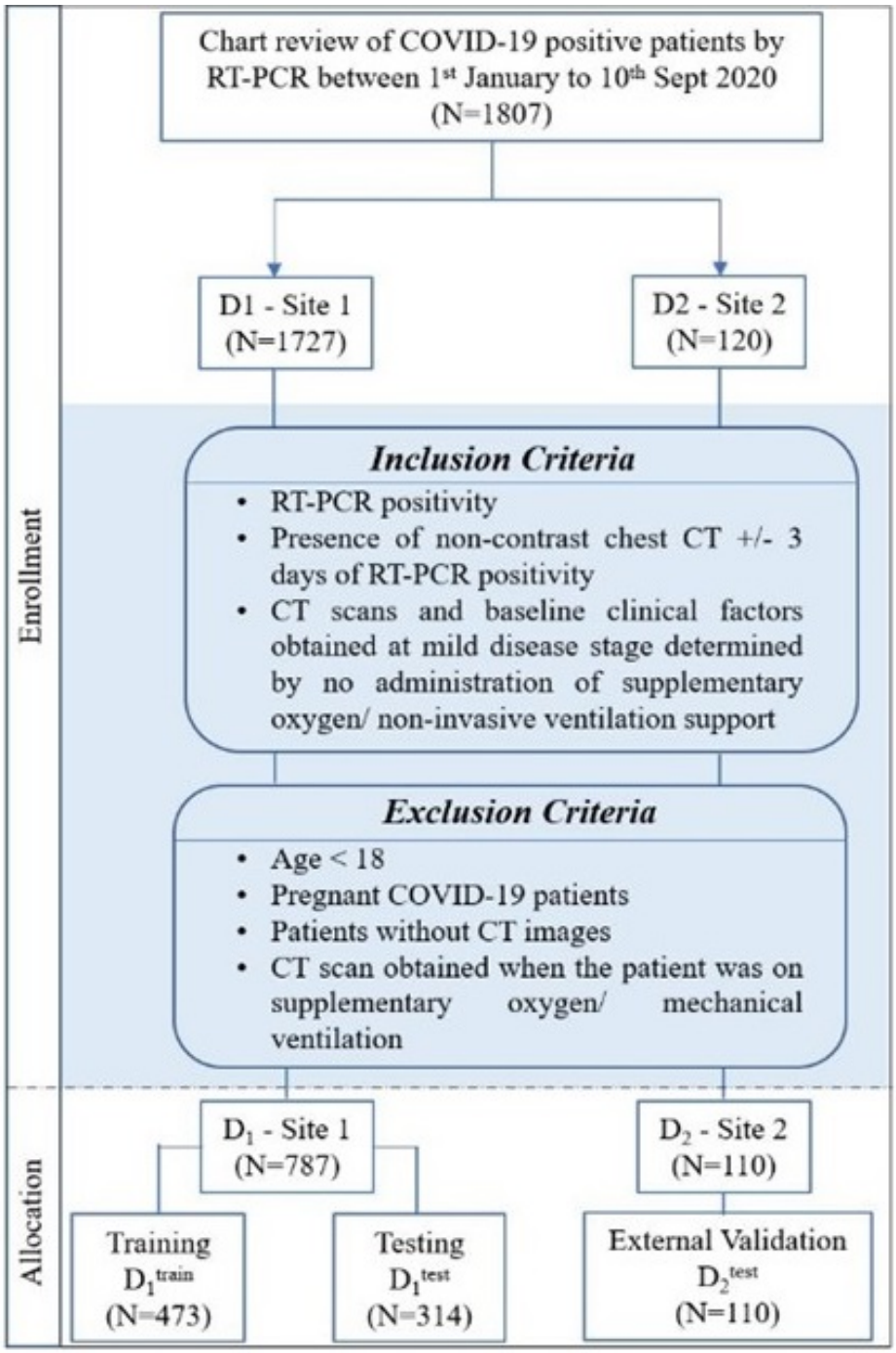

Figure 6.1: Patient selection criteria and dataset distribution..

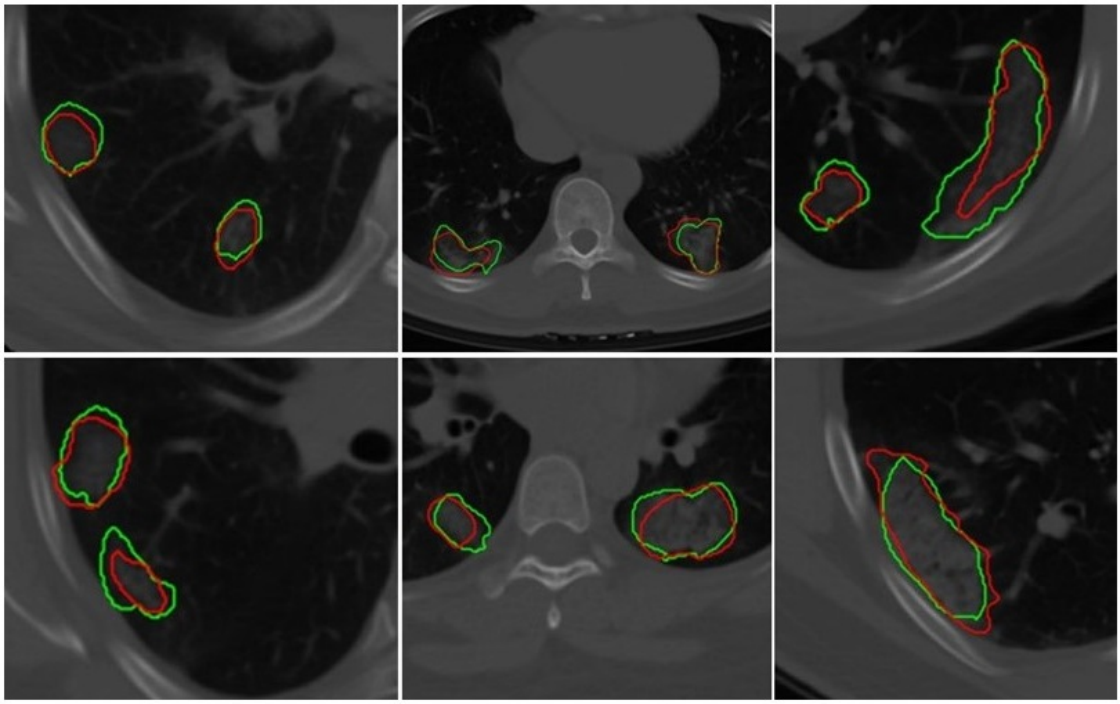

Figure 6.2: Architectural diagram of a 2D U-Net used for segmentation of ground-glass opacities and consolidations. U-Net consists of an encoder block and a decoder block. Both encoder and decoder have 5 convolutional blocks with two convolutional layers in each block

Radiomic Feature Extraction

After automatic segmentation of lung volume, all the scans were resampled to 0.75 mm in the x- and y-directions and simultaneously added a uniform slice thickness of 5 mm to reduce the impact of different equipment and scanning parameters. The total infection size was calculated by calculating the volume of the COVID consolidations annotated using the U-Net model. These consolidations were termed as COVID regions. Next, a total of 187 radiomic features were extracted from annotated CT scans. These features included 37 first-order features and 150 higher-order textural features. The textural features included the gray-level co-occurrence matrix (GLCM), gray-level size zone matrix (GLSZM), gray-level run length matrix (GLRLM), neighboring gray tone difference matrix (NGTDM), and gray level dependence matrix (GLDM).

These radiomic features capture textural patterns of COVID consolidations that are not apparent with the naked eye and could potentially help describe the heterogeneity of these

regions.

The top predictive radiomic features from the training cohort $D_T{}^1$ were selected using the least absolute shrinkage and selection operator (LASSO) feature selection algorithm.[122] These features were further used for constructing a continuous radiomic risk score using the weighted sum of their LASSO coefficients. The radiomics model (M_{RM}) was constructed using this developed radiomic risk score.

6.2.3 Clinical Feature Analysis

A total of 20 clinical variables and laboratory parameters were included in the analysis, as explained in table-2. Specifically, these features included patients' age and laboratory parameters, such as albumin (ALB), lymphocytes, WBCs, etc. Previous studies show a high correlation of these clinical variables with the patient being on the ventilator when admitted to a hospital[123].[124]

A total of 545 cases out of 897 had all the clinical variables available. The total missing rate of the clinical variables was 39.34%. To make use of all the available data, the missing clinical values were imputed by the mean values of available clinical entities from $D_T{}^1$. For an external validation set, the missing values were replaced by the mean obtained from the complete cases of the same cohort.

Similar to radiomics analysis, the most prognostic clinical variables were selected from the training cohort DT1 using LASSO analysis[122] and used within the logistic regression model for predicting the need for ventilators in COVID-19 patients (clinical model: M_{CM}).

6.2.4 Statistical Analysis

The primary endpoint of the study was predicting the severity of the COVID-19 disease, specifically, predicting patients who would require an invasive mechanical ventilator vs. those who would not. Figure 2 explains the entire experimental design pipeline.

First, to validate the automatic CNN-based segmentation model's performance, the Dice

similarity coefficient (DSC) was used. The DSC was evaluated on the voxel-wise segmentation performance and compared against an expert radiologist reader.

For building the prediction models, the top features were selected from the entire feature pool using the LASSO algorithm on $D_{\mathrm{T}}{}^{1}$ to constrict $\mathrm{M_{RM}}$ and $\mathrm{M_{CM}}$. LASSO provides a principled way to reduce the number of features in a model. LASSO penalizes the L1 norm of the weights, which induces sparsity in the solution (many weights are forced to zero). This performs variable selection (the "relevant" variables are allowed to have non-zero weights). The degree of sparsity is controlled by the penality term, which was selected within a 10-fold cross-validation setting. The $\mathrm{M_{RM}}$ model had top Radiomic features in the form of "radiomic score" constructed using the weighted sum of these features with their corresponding LASSO coefficients. The $\mathrm{M_{CM}}$ model consisted of top clinical features, and the final model, $\mathrm{M_{RCM}}$, was constructed using the top clinical features integrated with "radiomic score" in the form of nomogram analysis.

All three models were constructed with logistic regression (LR) classifiers. The receiver operating characteristic (ROC) and precision-recall (PR) analysis, along with sensitivity, specificity, and area under the curve (AUC), were used as performance metrics to evaluate the accuracy of the $\mathrm{M_{RM}}$, $\mathrm{M_{CM}}$, and $\mathrm{M_{RCM}}$. DeLong test was used to compare the statistical significance of differences between the models (34). Odds ratio (OR) and 95% confidence intervals (CI) were calculated to estimate the effect size of important clinical factors and image features. For $D_{\mathrm{T}}{}^{1}$, cross-validation results were reported as mean $\pm$ standard deviation.

The final $\mathrm{M_{RCM}}$ model was represented as a clinico-radiomic nomogram.[51] The patients were divided into high-risk (ventilator) groups and low-risk (non-ventilator) groups using the optimal cutoff point obtained from the LR model. The decision curve was plotted and evaluated to see the added improvement of the nomogram over the individual models. The net benefit was calculated by summing the benefits (true-positive results) and subtracting the harms (false-positive results), weighting the latter by a factor related to the relative harm of undetected disease severity with the harm of unnecessary ventilator treatment.[125] In this

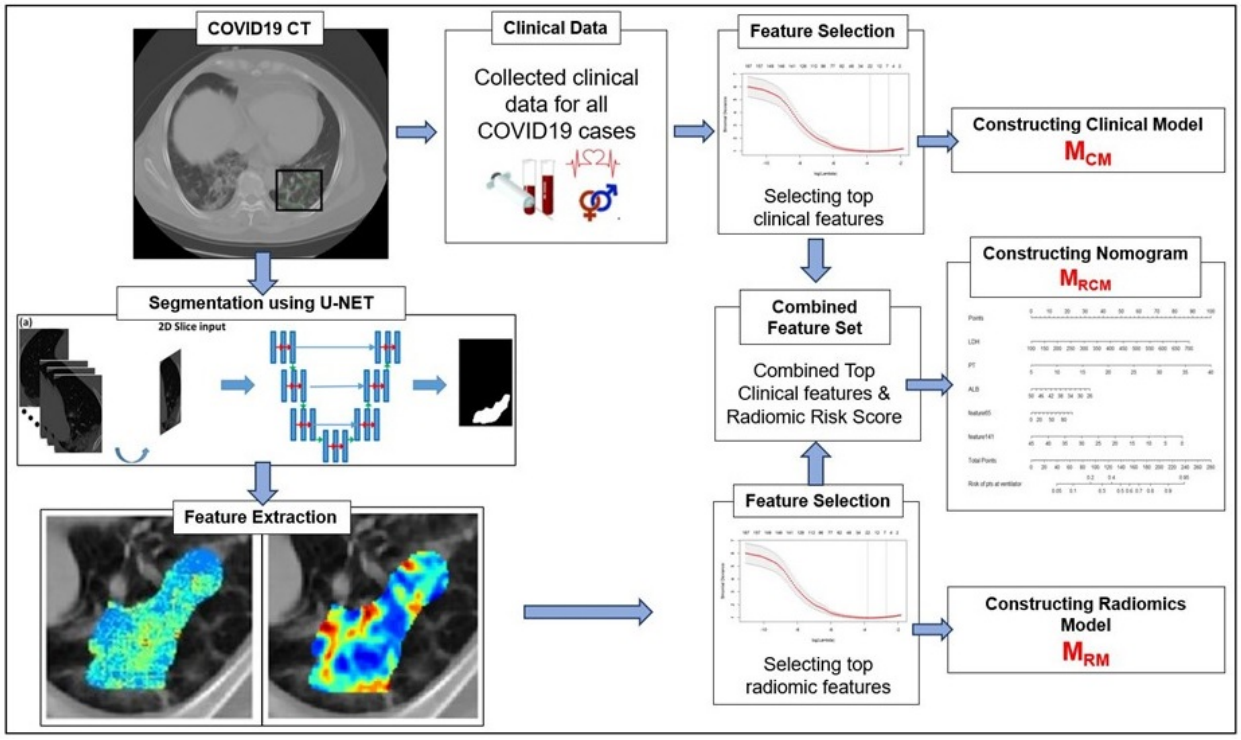

Figure 6.3: Workflow of the experiment. The first step involves segmentation of coronavirus disease (COVID) consolidations, which were further used for radiomic feature extractions. Next, the top clinical and radiomic features were selected using (LASSO) analysis and further used for constructing radiomic model (M_{RM}), clinical model (M_{CM}), and combined combined clinical–radiomic (M_{RCM}) nomogram.

analysis, the added improvement of the M_{RCM} model was shown over M_{CM} and M_{RM}.

6.3 Results

6.3.1 Study Population Characteristics

Table 1 lists the study population characteristics for the two institutions D^1 and D^2. The median age of the patients was 59 in D^1 and 60 in D^2. In D^1 and D^2, 41.9, 55.3% had a mild disease, whereas 58.1, 44.7% had a severe disease having ended up requiring invasive mechanical ventilation.

	Renmin Hospital (Wuhan, China)—D_1		University Hospitals (Cleveland, US)—D_2
	Train	**Test**	
Age median (IQR)	59 (46–67)	60 (48–69)	62 (20–94)
Ventilator			
Yes (%)	267 (62.09%)	154	28
No (%)	163	137	85
Laboratory findings			
Lactate dehydrogenase (U/L)	280.7 (126–936)	271.8 (108–1,039)	348.5 (127–919)
Albumin (g/L)	37.29 (22.7–49)	38.48 (24.6–50.6)	36 (25.0–44.0)
Radiomics			
Radiomic score median (IQR)	2.99 (2.57–3.47)	2.98 (2.46–3.52)	3.51 (3.184.02)

Figure 6.4: Patient Characteristics

Table 6.1: Patient characteristics.

Variables	$D_1 - Train$	
	D_2	
Age Median (IQR)	59 (49 - 69)	60 (51.25 – 68.75)
Sex		
Male (%)	385 (46.8%)	24 (51%)
Female (%)	437 (53.2%)	23 (49%)
Ventilator		
Yes (%)	478 (58.1%)	21 (44.7%)
No (%)	344 (41.9%)	26 (55.3%)
Laboratory findings Median (IQR)		
Lactate Dehydrogenase (units/liter)	230 (187 - 310.75)	250 (230 – 368.5)
Prothrombin Time (seconds)	11.6 (11 – 12.2)	12.7 (11.3 – 14.1)
Lymphocytes %	23.45 (14.6 – 31.8)	16.6 (12.35 – 23.55)
Albumin (grams/liter)	38.1 (34.6 – 41.37)	37 (33 - 39)
Aspartate Aminotransferase (units/liter)	25 (19-37)	27 (22-46)

6.3.2 Segmentation Model

The U-Net network detected 1,017 of 1,260 COVID regions (3D connected components) annotated by the radiologist with 449 false positives on ${D_\mathrm{T}}^{UNET}$. The corresponding sensitivity and positive predictive value (PPV) were found to be 80.71 and 69.3%, respectively. The output segmentation by the network had an overlap of DSC = 0.60 ± 0.02 with ground-truth delineations for the detected regions. On the validation set of N = 96 (${D_\mathrm{V}}^{UNET}$), 1,071 of 1,353 annotated regions were detected with 470 false positives, which resulted in a sensitivity of 79.15% and PPV of 69.5%. The corresponding DSC of the segmentation on ${D_\mathrm{V}}^{UNET}$ was 0.59. The corresponding DSC of the segmentation on ${D_\mathrm{V}}^{UNET}$ was 0.59 (Table 6.2).

Table 6.2: U-Net-based model analysis.

	Detected	False Positives	Sensitivity (%)	PPV(%)	DSC
${D_{\mathrm{T}}}^{UNET}$	1017/1260	449	80.71%	69.3%	0.60 +/- 0.02
${D_{\mathrm{V}}}^{UNET}$	1071/1353	470	79.15%	69.5%	0.59

Table 6.3: Selected top features.

Feature Family		Feature Name	LASSO Coefficient
Texture Feature	GLCM	Inverse Variance	1.65 e(0)
	GLRLM	High Gray Level Run Emphasis	1.96e(-2)
	GLSZM	Small Area Low Gray Level Emphases	8.95 e(-5)
First Order		90th percentile pixel value	-3.48 e(-2)
Absolute Infection Size			2.38 e(-6)

6.3.3 Individual Radiomic- and Clinical-Based Machine Learning Models for Predicting Patients Being on the Ventilator for COVID-19 Patients

The top five features selected within the radiomic model using the LASSO analysis are listed in Table 3. Figure 4 shows the difference between feature maps for ventilator and non-ventilator cases. These features were statistically significant between the ventilator and non-ventilator groups, with higher feature values potentially representing patients at higher risk of disease. The violin plots of the top features are represented in figure 3.

The constructed logistic regression model with radiomic score ($\mathrm{M_{RM}}$) had an AUC of 0.754, 95% CI (0.709–0.799) on ${D_{\mathrm{T}}}^{1}$. The same model gave an AUC of 0.836, 0.758, and 0.719 on ${D_{\mathrm{V}}}^{1}$, ${D_{\mathrm{V}}}^{2}$, and combined test set (${D_{\mathrm{V}}}^{1}$ + ${D_{\mathrm{V}}}^{2}$). For the clinical model, $\mathrm{M_{CM}}$, the LASSO method selected albumin (ALB), lactate dehydrogenase (LDH), and age as the most predictive parameters. Using the most discriminating clinical factors, the model trained yielded an AUC of 0.784, 95% CI: (0.743–0.825) on ${D_{\mathrm{T}}}^{1}$ and 0.813, 0.688, and 0.703 on ${D_{\mathrm{V}}}^{1}$, ${D_{\mathrm{V}}}^{2}$, and combined test set (${D_{\mathrm{V}}}^{1}$ + ${D_{\mathrm{V}}}^{2}$), respectively.

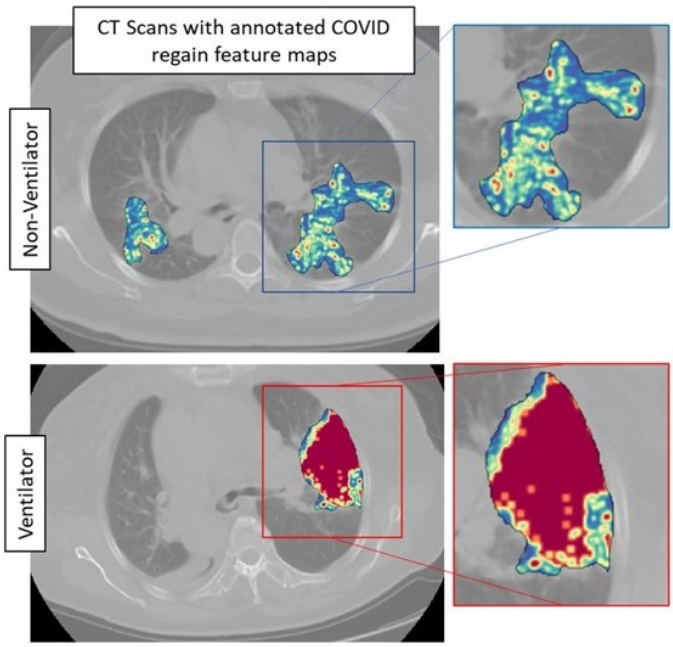

Figure 6.5: Feature maps of top selected features for ventilated (lower row) and non-ventilated (upper row) cases.

6.3.4 An Integrated Clinical and Imaging Nomogram to Predict the Need for Mechanical Ventilation in COVID-19 Patients

The integrated radiomic–clinical nomogram, MRCM, included the radiomic score and three clinical parameters—age, albumin, and lactate dehydrogenase. Table 4 shows the effect size and odds ratio for these variables.

The M_{RCM} model outperformed both M_{CM} and M_{RM}, resulting in an AUC of 0.847 and 0.771, and 0.735 on ${D_V}^1$, ${D_V}^2$, and combined ${D_V}^1 + {D_V}^2$ test set, respectively.

The multivariate logistic regression analysis of the M_{RCM} nomogram showed that the radiomic score was found to add independent prognostic value to the M_{RCM} model. The predicted score of 0.54 or greater [an optimal cutoff point on the receiver operating characteristic (ROC) curve] suggested the need for mechanical ventilation, while scores ≤ 0.54 could be managed conservatively (Figure 5). Additionally, the AUC comparison within the three models showed that the increase in AUC in M_{RCM} was statistically significant when

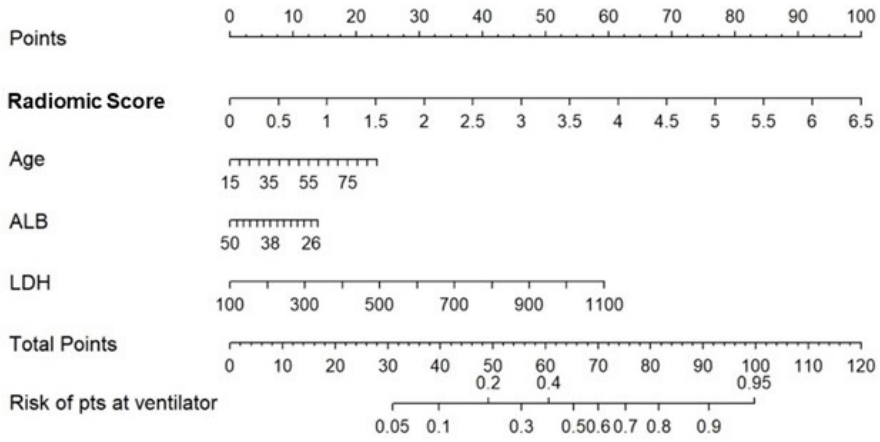

Figure 6.6: Constructed nomogram, M_{RCM}, which included radiomic score, age, albumin (ALB), and lactate dehydrogenase (LDH). The nomogram calculates the probability of the patient being on the ventilator.

compared against the clinical model M_{CM}.

The decision curve analysis indicated an added net benefit using the integrated model M_{RCM} over M_{CM} and M_{RM} (Figure 6). The combined M_{RCM} model had the highest net benefit compared with M_{CM}, M_{RM}, and simple strategies, such as treating all patients (light vertical curve line) or treating no patients (horizontal black line) across the full range of threshold probabilities.

6.4 Discussion

In this study, we presented an integrated radiomic and clinical nomogram (M_{RCM}) to predict at baseline patients with a severe phenotype of COVID-19 and who would end up needing mechanical ventilation and intubation. We explicitly used patients with baseline CT scans and laboratory parameters observed within the milder stage of the disease to reduce the bias. M_{RCM} comprised a radiomic score constructed using the annotated GGO and consolidation regions on lung CT scans along with age, albumin (ALB), and lactate dehydrogenase (LDH). Meanwhile, the radiomic model (M_{RM}) incorporated the radiomic score constructed using

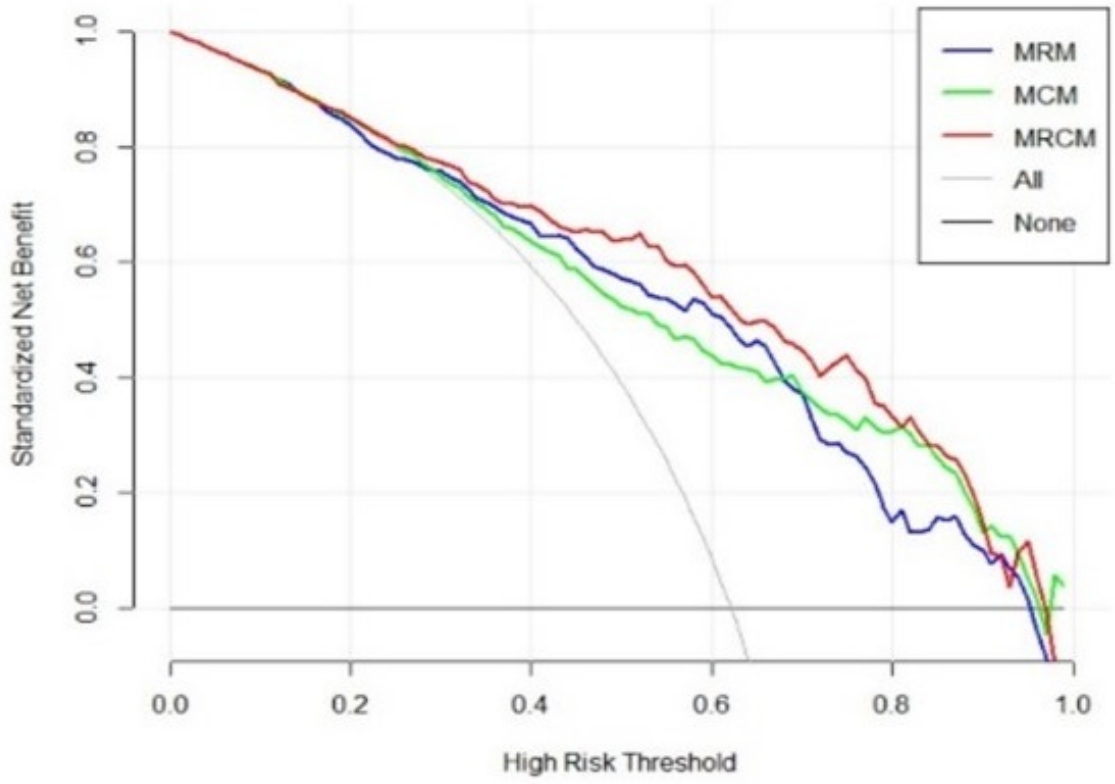

Figure 6.7: Decision curve analysis of M_{RCM} (clinical and imaging integrated nomogram) constructed using developed radiomic score, age, and three laboratory parameters (LDH and ALB). The other (bottom) were clinical (M_{CM}) and radiomic model (M_{RM}).

five radiomic features. The clinical model (M_{CM}) was built using age, albumin, and lactate dehydrogenase out of routine clinical laboratory parameters. We constructed a U-NET-based segmentation algorithm to segment COVID-19 regions from the baseline CT scans to completely automate the whole process. The three models were trained and independently validated on a large multi-institutional dataset making this the most extensive study to date involving AI and radiomics for the prognosis of COVID-19 patients.

Our radiomic model, M_{RM}, incorporated radiomic score constructed using top features observed from within the gray-level matrix-based feature families explaining textural patterns of COVID regions. These features had higher expression in potentially high-risk cases, suggesting a more chaotic and disturbed microarchitecture in patients at a higher risk of disease (Figure 6.5). Our results are in line with results presented by Wu et al.,[117] where four features out of five were observed from gray-level matrix-based feature family. The higher textural value from the gray-level co-occurrence matrix indicates the more abnormal lung tissues, which further seemed to be associated with the worse outcome. This is consistent with previous findings that show that peripheral, diffuse distributions and paving patterns are associated with poor survival in COVID-19 cases.[126] Compared with the usual imaging CT model features, radiomics offer superior performance in the COVID-19 space. Simply looking at radiomic models for predicting the severity of COVID-19 patients, the signatures constructed using SVM by Fu et al.[127] achieved an AUC of 0.83 on N = 64, and Wei (25) achieved an AUC of 0.93 on N = 81. Our results show a better performance considering that we had larger datasets with completely independent multi-institutional validation sets.

The most prognostic clinical variables observed within the clinical model were age, ALB, and LDH selected using the LASSO. A low level of ALB was associated with poorer outcomes, i.e., the patient being on the ventilator.[124] In contrast, low levels of LDH were associated with better outcomes[124].[123] The boxplots of these features are depicted in Appendix 1. ALB and LDH are considered biomarkers for predicting the COVID-19 severity in the previously published findings.[124] We observed the third important clinical feature to be the patient's

age, where an advanced age was associated with a worse outcome for COVID19 patients.[128]

The integrated M_{RCM} model outperformed M_{RM} and M_{CM} models in predicting which COVID-19 patients would ultimately need invasive mechanical ventilation on both internal and external validation sets ${D_V}^1$ and ${D_V}^2$. M_{RCM} improved performance by over 2.5% over M_{RM} and 3.77% over M_{CM} in terms of AUC, with the performance increase statistically significant by DeLong's test. The M_{RCM} model was used to individualize risk assessments. The predicted score of 0.54 or greater [an optimal cutoff point on the receiver operating characteristic (ROC) curve that had an optimal balance between sensitivity and specificity] suggested the need for mechanical ventilation, while scores ≤ 0.54 could be managed conservatively. We only noticed one nomogram approach developed by Yu et al.,[129] which used age, density, perfusion signs, and severity score of lungs constructed by assessing each lobe of the lung for predicting the severity of COVID-19. The nomogram achieved an AUC of 0.929 (95% CI, 0.889–0.969) on training (N = 152) and 0.936 (95% CI, 0.867–1.000) on the validation set (N = 65), but their analysis did not involve radiomics. Our developed nomogram was completely automated, had minimal involvement of a radiologist, and achieved almost comparable results within larger datasets.

The previous work on combining radiomics with clinical variables shows promising results for predicting disease severity. For the combined clinical and radiomic model, in the work by Chao et al.,[116] the authors integrated the L/W ratio, lymphocyte count, WBC, and age into whole lung radiomics to achieve the highest AUC of 0.88 in predicting the need for ICU admission. The advantage in our approach compared with previous ones includes a higher number of cases and a nomogram representation.

In the recent study by Roberts et al.,[130] the authors point out that many recent AI/machine learning studies on diagnosis and prognosis of COVID-19 from radiographic scans are not reproducible and would not be clinically deployable. Furthermore, they point out that many studies within this space have not been stress tested or validated on independent external test sets. Many of these models have not assessed model sensitivity or robustness and have

methodological flaws and/or underlying biases. In our work, we have attempted to deliberately and purposefully develop, validate, and analyze our approach in a more rigorous manner, including validating this model on one of the largest external test sets reported to date.

Despite the favorable prognostic efficacy of the clinico-radiomic nomogram, we acknowledge that our approach does have its limitations. First, our study was retrospective, and the two cohorts were not homogeneously defined. To ensure the clinical usefulness of M_{RCM}, we need to validate the tool in a prospective setting by following up with patients until discharge. Second, the study's retrospective nature also precluded us from standardizing the time between RT-PCR and CT scans across the cohort. Finally, we did not explicitly compare segmentation and prediction performances between the AI model and expert radiologist interpretations. We will attempt to address these limitations in future work.

6.5 Conclusions

We presented an integrated radiomic and clinical parameter-based prognostic model using routinely available blood parameters and standard-of-care CT scans at baseline in SARS-CoV2-positive patients at the milder stage of the disease. We showed in a multi-institutional cohort that our integrated model had a good performance in identifying which of these patients would decline in severe respiratory distress with need for intubation and mechanical ventilation. Further multisite prospective validation would allow for the clinical deployment of M_{RCM}, especially to triage patients for ventilator usage, in the face of worldwide shortages in the availability of mechanical ventilators. The developed tool, once prospectively validated, could provide an objective way to risk stratifying patients immediately following diagnosis with COVID-19.

Chapter 7

Concluding remarks and future work

7.1 Concluding remarks

In this dissertation, we explored lung cancer related novel prognostic and treatment prediction strategies using imaging scans. First, we used novel Radiomic-Based approached and extracted hand-crafted features from tumor region and form the region outside the nodule and constructed AI models for targeting four specific clinical problems- a) predicting recurrence and subsequently added benefit of adjuvant chemotherapy for early-stage non-small cell lung cancer b) predicting the minimally invasive lung cancer from invasive lung cancer cases using c) predicting immunotherapy response for late-stage lung cancer cases d) predicting severe COVID-19 cases looking at the CT scans alone. We further explored various, unique fusion strategies for each of these clinical problem statements, spanning from combining multimodal data as well as non-imaging data with the imaging models. These strategies include - a) including clinical variables within the nomogram fashion b) integrating radiology and pathology features for creating multimodal analyses, c) integrating input from the radiologist for creating human-machine integrated models. Lastly, to understand the biological interpretability of the hand-crafted CT-scan based features, we performed a) radiology-pathology correlation, b) radiogenomic analysis on a small subset of patients as a

preliminary step towards an interpretable AI.

- We developed a CT-based Radiomics model for Early-Stage non-small cell lung cancer cases to predict recurrence and specifically, the added benefit of adjuvant-chemotherapy for these patients. For Early-Stage cases, currently there's no consensus on which patients should receive adjuvant-chemotherapy, especially stage I to IIB. Various clinical trials have an opposite results on which patients should be receiving it. Our developed imaging model was able to identify specific population who would benefit from adjuvant-chemotherapy following surgery. Further, this model was integrated with clinical variables and we showed that overall performance of the model increases. The whole analysis was performed on 500 cases from three institutions and the overall model was validated on independent validation cohorts to ensure the accuracies.

- We developed a CT-based Radiomics model to predict the invasive adenocarcinoma nodules from the minimally invasive ones. Currently, the gold standard for predicting the level of invasion of lung nodules depends on the surgically resected tissue scans. Predicting minimally invasive nodules using CT scans would help in the surgical planning and treatment of these patients. We fused the developed imaging based model with the radiologists interpretation and created a human-machine integrated model which not only improved our performance, but also showed a way ho wto actually implement these models into clinical settings.

- We constructed a model to predict the response to immunotherapy for late-stage NSCLC cases. Specifically, within this late-stage cohort, there's a subset of population, known as hyperprogressors, who if received an immunotherapy, would have complete adverse effect of it and would reduce their life-span. We used CT scans and Radiomics analysis to specifically find out this population from the baseline scans that could help avoiding immunothreay for these patients.

- Our work within the COVID-19 space shows that Radiomics features observed on the

CT scans can be useful even within other diseases. We could stratify severe patients within the COVID-19 space using the CT scans alone. Further, we integrated these radiomic features with various labpratory tests currently used in a practice in a novel nomogram fashion and showed an overall improved performance of the model.

- We created a model using multi-scale imaging information. We specially looked at radiology-based and pathology-based features and demonstrated that these two modalities present complementary information which could be fused together to improve the model prediction, specifically within the lung cancer space. We looked at two specific use cases- for predicting added benefit of adjuvant-chemotherapy for early-stage cases as well as combined these features for predicting immunotherapy response for late-stage NSCLC cases.

- We showed that Radiomics features observed on CT scans could potentially representing underlying biology by exploring pathology whole-slide tissue scans as well as radiogenomic analysis.

Together, the work described here helps in advancing the translation of clinically useful artificial intelligence tools that can help in improving patient quality of life. The integration methods used here may provide a useful insight of novel techniques for combining imaging , non-imaging, multiscale biomarkers within the non-small cell lung cancer domain.

7.2 Future work

While we have successfully shown the application of radiomics, and integrated Radiomics-Pathomics Models for lung cancer cases and COVID-19 disease, we do acknowledge that there are several limitations which we hope to conduct as part of the future study.

Firstly, successfully transitioning these technologies into clinical practices would require extensive external validation from multiple institutions and larger cohorts. First, we plan

to start validating these models on clinical trial datasets, and in the future plan to validate them on prospective datasets.

In addition, we also plan to include an extensive analysis of the raw parameters before building the models. As we have noticed so far, radiomic features get affected by the raw parameters of the CT scans (including kernel size, tube voltage, slice thickness etc.). In the future, we plan to develop the standardization procedure to help increase the reproducibility of the Radiomics Models. We also plan to include an automated tumor and peritumoral area segmentation pipeline in the model, which would reduce the burden on the radiologists, and they could focus on the big picture, i.e., on identifying the patient population at risk.

Similarly, with Pathomics, we plan to have a robust pipeline that would be generalized across multiple institutions.

Another important step would be to understand the biological interpretability of these hand-crafted radiology and pathology features for the successful transition of these methods into clinical practices. Even though we have performed an elementary experiment to understand the biological meaning of radiomic features, we plan to do that extensively in the future. Additionally, the work presented in this dissertation is primarily focused on applications related to the lung. However, exploring whether these approaches can be translated to risk-stratification and characterization of other diseases are considered part of future work.

Bibliography

[1] W. D. Travis, E. Brambilla, M. Noguchi, A. G. Nicholson, K. R. Geisinger, Y. Yatabe, D. G. Beer, C. A. Powell, G. J. Riely, P. E. Van Schil, K. Garg, J. H. M. Austin, H. Asamura, V. W. Rusch, F. R. Hirsch, G. Scagliotti, T. Mitsudomi, R. M. Huber, Y. Ishikawa, J. Jett, M. Sanchez-Cespedes, J.-P. Sculier, T. Takahashi, M. Tsuboi, J. Vansteenkiste, I. Wistuba, P.-C. Yang, D. Aberle, C. Brambilla, D. Flieder, W. Franklin, A. Gazdar, M. Gould, P. Hasleton, D. Henderson, B. Johnson, D. Johnson, K. Kerr, K. Kuriyama, J. S. Lee, V. A. Miller, I. Petersen, V. Roggli, R. Rosell, N. Saijo, E. Thunnissen, M. Tsao, and D. Yankelewitz, "International association for the study of lung cancer/american thoracic society/european respiratory society international multidisciplinary classification of lung adenocarcinoma," *Journal of Thoracic Oncology: Official Publication of the International Association for the Study of Lung Cancer*, vol. 6, pp. 244–285, Feb. 2011.

[2] C. Rampinelli, D. Origgi, and M. Bellomi, "Low-dose CT: technique, reading methods and image interpretation," *Cancer Imaging*, vol. 12, pp. 548–556, Feb. 2013.

[3] M. B. Amin, F. L. Greene, S. B. Edge, C. C. Compton, J. E. Gershenwald, R. K. Brookland, L. Meyer, D. M. Gress, D. R. Byrd, and D. P. Winchester, "The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging," *CA: a cancer journal for clinicians*, vol. 67, pp. 93–99, Mar. 2017.

[4] A. G. Nicholson, M. S. Tsao, M. B. Beasley, A. C. Borczuk, E. Brambilla, W. A. Cooper, S. Dacic, D. Jain, K. M. Kerr, S. Lantuejoul, M. Noguchi, M. Papotti, N. Rekhtman, G. Scagliotti, P. v. Schil, L. Sholl, Y. Yatabe, A. Yoshida, and W. D. Travis, "The 2021 WHO Classification of Lung Tumors: Impact of Advances Since 2015," *Journal of Thoracic Oncology*, vol. 17, pp. 362–387, Mar. 2022. Publisher: Elsevier.

[5] J. H. M. Austin, K. Garg, D. Aberle, D. Yankelevitz, K. Kuriyama, H.-J. Lee, E. Brambilla, and W. D. Travis, "Radiologic Implications of the 2011 Classification of Adenocarcinoma of the Lung," *Radiology*, vol. 266, pp. 62–71, Jan. 2013. Publisher: Radiological Society of North America.

[6] N. Gardiner, S. Jogai, and A. Wallis, "The revised lung adenocarcinoma classification—an imaging guide," *Journal of Thoracic Disease*, vol. 6, pp. S537–S546, Oct. 2014.

[7] W. Weichert and A. Warth, "Early lung cancer with lepidic pattern: adenocarcinoma in situ, minimally invasive adenocarcinoma, and lepidic predominant adenocarcinoma," *Current Opinion in Pulmonary Medicine*, vol. 20, pp. 309–316, July 2014.

[8] L. Xu, F. Tavora, R. Battafarano, and A. Burke, "Adenocarcinomas with prominent lepidic spread: retrospective review applying new classification of the American Thoracic Society," *The American Journal of Surgical Pathology*, vol. 36, pp. 273–282, Feb. 2012.

[9] M. Noguchi, A. Morikawa, M. Kawasaki, Y. Matsuno, T. Yamada, S. Hirohashi, H. Kondo, and Y. Shimosato, "Small adenocarcinoma of the lung. Histologic characteristics and prognosis," *Cancer*, vol. 75, pp. 2844–2852, June 1995.

[10] Y. Tsutani, Y. Miyata, H. Nakayama, S. Okumura, S. Adachi, M. Yoshimura, and M. Okada, "Appropriate Sublobar Resection Choice for Ground Glass Opacity-Dominant Clinical Stage IA Lung Adenocarcinoma: Wedge Resection or Segmentectomy," *Chest*, vol. 145, pp. 66–71, Jan. 2014.

[11] A. P. Brady, "Error and discrepancy in radiology: inevitable or avoidable?," *Insights into Imaging*, vol. 8, pp. 171–182, Dec. 2016.

[12] K. Loverdos, A. Fotiadis, C. Kontogianni, M. Iliopoulou, and M. Gaga, "Lung nodules: A comprehensive review on current approach and management," *Annals of Thoracic Medicine*, vol. 14, no. 4, pp. 226–238, 2019.

[13] R. J. Gillies, P. E. Kinahan, and H. Hricak, "Radiomics: Images Are More than Pictures, They Are Data," *Radiology*, vol. 278, pp. 563–577, Feb. 2016.

[14] K. Bera, V. Velcheti, and A. Madabhushi, "Novel Quantitative Imaging for Predicting Response to Therapy: Techniques and Clinical Applications," *American Society of Clinical Oncology educational book. American Society of Clinical Oncology. Annual Meeting*, pp. 1008–1018, May 2018.

[15] P. Vaidya, K. Bera, A. Gupta, X. Wang, G. Corredor, P. Fu, N. Beig, P. Prasanna, P. Patil, P. Velu, P. Rajiah, R. Gilkeson, M. Feldman, H. Choi, V. Velcheti, and A. Madabhushi, "CT derived radiomic score for predicting the added benefit of adjuvant chemotherapy following surgery in Stage I, II resectable Non-Small Cell Lung Cancer: a retrospective multi-cohort study for outcome prediction," *The Lancet. Digital Health*, vol. 2, pp. e116–e128, Mar. 2020.

[16] P. Grossmann, O. Stringfield, N. El-Hachem, M. M. Bui, E. Rios Velazquez, C. Parmar, R. T. Leijenaar, B. Haibe-Kains, P. Lambin, R. J. Gillies, and H. J. Aerts, "Defining the biological basis of radiomic phenotypes in lung cancer," *eLife*, vol. 6, p. e23421, 2017.

[17] Q. Weng, L. Zhou, H. Wang, J. Hui, M. Chen, P. Pang, L. Zheng, M. Xu, Z. Wang, and J. Ji, "A radiomics model for determining the invasiveness of solitary pulmonary nodules that manifest as part-solid nodules," *Clinical Radiology*, vol. 74, pp. 933–943, Dec. 2019.

[18] Y. She, L. Zhang, H. Zhu, C. Dai, D. Xie, H. Xie, W. Zhang, L. Zhao, L. Zou, K. Fei, X. Sun, and C. Chen, "The predictive value of CT-based radiomics in differentiating indolent from invasive lung adenocarcinoma in patients with pulmonary nodules," *European Radiology*, June 2018.

[19] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural Features for Image Classification," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-3, pp. 610–621, Nov. 1973.

[20] P. Prasanna, P. Tiwari, and A. Madabhushi, "Co-occurrence of Local Anisotropic Gradient Orientations (CoLlAGe): A new radiomics descriptor," *Scientific Reports*, vol. 6, p. 37241, Nov. 2016.

[21] K. Laws, "Textured image segmentation," IPI Report 940, University of Southern California, 1980.

[22] J.-K. Kamarainen, "Gabor features in image analysis," in *2012 3rd International Conference on Image Processing Theory, Tools and Applications (IPTA)*, (Istanbul, Turkey), pp. 13–14, IEEE, Oct. 2012.

[23] Y. Balagurunathan, Y. Gu, H. Wang, V. Kumar, O. Grove, S. Hawkins, J. Kim, D. B. Goldgof, L. O. Hall, R. A. Gatenby, and others, "Reproducibility and prognosis of quantitative features extracted from CT images," *Translational oncology*, vol. 7, no. 1, pp. 72–87, 2014.

[24] C. Ding and H. Peng, "MINIMUM REDUNDANCY FEATURE SELECTION FROM MICROARRAY GENE EXPRESSION DATA," *Journal of Bioinformatics and Computational Biology*, vol. 03, pp. 185–205, Apr. 2005.

[25] E. R. DeLong, D. M. DeLong, and D. L. Clarke-Pearson, "Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach," *Biometrics*, vol. 44, pp. 837–845, Sept. 1988.

[26] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53–65, Nov. 1987.

[27] M. Takahashi, Y. Shigematsu, M. Ohta, H. Tokumasu, T. Matsukura, and T. Hirai, "Tumor invasiveness as defined by the newly proposed IASLC/ATS/ERS classification has prognostic significance for pathologic stage IA lung adenocarcinoma and can be predicted by radiologic parameters," *The Journal of Thoracic and Cardiovascular Surgery*, vol. 147, pp. 54–59, Jan. 2014.

[28] Y. Moon, K. Y. Lee, and J. K. Park, "The prognosis of invasive adenocarcinoma presenting as ground-glass opacity on chest computed tomography after sublobar resection," *Journal of Thoracic Disease*, vol. 9, Oct. 2017. Publisher: AME Publishing Company.

[29] M. Behera, T. K. Owonikoko, A. A. Gal, C. E. Steuer, S. Kim, R. N. Pillai, F. R. Khuri, S. S. Ramalingam, and G. L. Sica, "Lung Adenocarcinoma Staging Using the 2011 IASLC/ATS/ERS Classification: A Pooled Analysis of Adenocarcinoma In Situ and Minimally Invasive Adenocarcinoma," *Clinical Lung Cancer*, vol. 17, pp. e57–e64, Sept. 2016.

[30] F. Xu, W. Zhu, Y. Shen, J. Wang, R. Xu, C. Qutesh, L. Song, Y. Gan, C. Pu, and H. Hu, "Radiomic-Based Quantitative CT Analysis of Pure Ground-Glass Nodules to Predict the Invasiveness of Lung Adenocarcinoma," *Frontiers in Oncology*, vol. 10, Aug. 2020.

[31] X. Xue, Y. Yang, Q. Huang, F. Cui, Y. Lian, S. Zhang, L. Yao, W. Peng, X. Li, P. Pang, J. Yan, and F. Chen, "Use of a Radiomics Model to Predict Tumor Invasiveness of Pulmonary Adenocarcinomas Appearing as Pulmonary Ground-Glass Nodules," June 2018. ISSN: 2314-6133 Publisher: Hindawi Volume: 2018.

[32] G. Wu, H. C. Woodruff, J. Shen, T. Refaee, S. Sanduleanu, A. Ibrahim, R. T. H. Leijenaar, R. Wang, J. Xiong, J. Bian, J. Wu, and P. Lambin, "Diagnosis of Invasive Lung Adenocarcinoma Based on Chest CT Radiomic Features of Part-Solid Pulmonary Nodules: A Multicenter Study," *Radiology*, vol. 297, pp. 451–458, Nov. 2020.

[33] T. Luo, K. Xu, Z. Zhang, L. Zhang, and S. Wu, "Radiomic features from computed tomography to differentiate invasive pulmonary adenocarcinomas from non-invasive pulmonary adenocarcinomas appearing as part-solid ground-glass nodules," *Chinese Journal of Cancer Research*, vol. 31, pp. 329–338, Apr. 2019.

[34] L. Wu, C. Gao, J. Ye, J. Tao, N. Wang, P. Pang, P. Xiang, and M. Xu, "The value of various peritumoral radiomic features in differentiating the invasiveness of adenocarcinoma manifesting as ground-glass nodules," *European Radiology*, vol. 31, pp. 9030–9037, Dec. 2021.

[35] P. Goldstraw, K. Chansky, J. Crowley, R. Rami-Porta, H. Asamura, W. E. Eberhardt, A. G. Nicholson, P. Groome, A. Mitchell, V. Bolejack, P. Goldstraw, R. Rami-Porta, H. Asamura, D. Ball, D. G. Beer, R. Beyruti, V. Bolejack, K. Chansky, J. Crowley, F. Detterbeck, W. E. Erich Eberhardt, J. Edwards, F. Galateau-Sallé, D. Giroux, F. Gleeson, P. Groome, J. Huang, C. Kennedy, J. Kim, Y. T. Kim, L. Kingsbury, H. Kondo, M. Krasnik, K. Kubota, A. Lerut, G. Lyons, M. Marino, E. M. Marom, J. van Meerbeeck, A. Mitchell, T. Nakano, A. G. Nicholson, A. Nowak, M. Peake, T. Rice, K. Rosenzweig, E. Ruffini, V. Rusch, N. Saijo, P. Van Schil, J.-P. Sculier, L. Shemanski, K. Stratton, K. Suzuki, Y. Tachimori, C. F. Thomas, W. Travis, M. S. Tsao, A. Turrisi, J. Vansteenkiste, H. Watanabe, Y.-L. Wu, P. Baas, J. Erasmus, S. Hasegawa, K. Inai, K. Kernstine, H. Kindler, L. Krug, K. Nackaerts, H. Pass, D. Rice, C. Falkson, P. L. Filosso, G. Giaccone, K. Kondo, M. Lucchi, M. Okumura, E. Blackstone, F. Abad Cavaco, E. Ansótegui Barrera, J. Abal Arca, I. Parente Lamelas, A. Arnau Obrer, R. Guijarro Jorge,

D. Ball, G. Bascom, A. Blanco Orozco, M. González Castro, M. Blum, D. Chimondeguy, V. Cvijanovic, S. Defranchi, B. de Olaiz Navarro, I. Escobar Campuzano, I. Macía Vidueira, E. Fernández Araujo, F. Andreo García, K. Fong, G. Francisco Corral, S. Cerezo González, J. Freixinet Gilart, L. García Arangüena, S. García Barajas, P. Girard, T. Goksel, M. González Budiño, G. González Casaurrán, J. Gullón Blanco, J. Hernández Hernández, H. Hernández Rodríguez, J. Herrero Collantes, M. Iglesias Heras, J. Izquierdo Elena, E. Jakobsen, S. Kostas, P. León Atance, A. Núñez Ares, M. Liao, M. Losanovscky, G. Lyons, R. Magaroles, L. De Esteban Júlvez, M. Mariñán Gorospe, B. McCaughan, C. Kennedy, R. Melchor Íñiguez, L. Miravet Sorribes, S. Naranjo Gozalo, C. Álvarez de Arriba, M. Núñez Delgado, J. Padilla Alarcón, J. Peñalver Cuesta, J. Park, H. Pass, M. Pavón Fernández, M. Rosenberg, E. Ruffini, V. Rusch, J. Sánchez de Cos Escuín, A. Saura Vinuesa, M. Serra Mitjans, T. Strand, D. Subotic, S. Swisher, R. Terra, C. Thomas, K. Tournoy, P. Van Schil, M. Velasquez, Y. Wu, and K. Yokoi, "The IASLC Lung Cancer Staging Project: Proposals for Revision of the TNM Stage Groupings in the Forthcoming (Eighth) Edition of the TNM Classification for Lung Cancer," *Journal of Thoracic Oncology*, vol. 11, pp. 39–51, Jan. 2016.

[36] W. J. Scott, J. Howington, S. Feigenberg, B. Movsas, and K. Pisters, "Treatment of Non-small Cell Lung Cancer Stage I and Stage II: ACCP Evidence-Based Clinical Practice Guidelines (2nd Edition)," *Chest*, vol. 132, pp. 234S–242S, Sept. 2007.

[37] R. Arriagada, A. Dunant, J.-P. Pignon, B. Bergman, M. Chabowski, D. Grunenwald, M. Kozlowski, C. Le Péchoux, R. Pirker, M.-I. S. Pinel, M. Tarayre, and T. Le Chevalier, "Long-term results of the international adjuvant lung cancer trial evaluating adjuvant Cisplatin-based chemotherapy in resected lung cancer," *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, vol. 28, pp. 35–42, Jan. 2010.

[38] R. Arriagada, B. Bergman, A. Dunant, T. Le Chevalier, J.-P. Pignon, J. Vansteenkiste, and International Adjuvant Lung Cancer Trial Collaborative Group, "Cisplatin-based adjuvant chemotherapy in patients with completely resected non-small-cell lung cancer," *The New England Journal of Medicine*, vol. 350, pp. 351–360, Jan. 2004.

[39] T. Winton, R. Livingston, D. Johnson, J. Rigas, M. Johnston, C. Butts, Y. Cormier, G. Goss, R. Inculet, E. Vallieres, W. Fry, D. Bethune, J. Ayoub, K. Ding, L. Seymour, B. Graham, M.-S. Tsao, D. Gandara, K. Kesler, T. Demmy, F. Shepherd, National Cancer Institute of Canada Clinical Trials Group, and National Cancer Institute of the United States Intergroup JBR.10 Trial Investigators, "Vinorelbine plus cisplatin vs. observation in resected non-small-cell lung cancer," *The New England Journal of Medicine*, vol. 352, pp. 2589–2597, June 2005.

[40] J.-P. Pignon, H. Tribodet, G. V. Scagliotti, J.-Y. Douillard, F. A. Shepherd, R. J. Stephens, A. Dunant, V. Torri, R. Rosell, L. Seymour, S. G. Spiro, E. Rolland, R. Fossati, D. Aubert, K. Ding, D. Waller, T. Le Chevalier, and LACE Collaborative Group, "Lung adjuvant cisplatin evaluation: a pooled analysis by the LACE Collaborative

Group," *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, vol. 26, pp. 3552–3559, July 2008.

[41] D. Waller, M. D. Peake, R. J. Stephens, N. H. Gower, R. Milroy, M. K. B. Parmar, R. M. Rudd, and S. G. Spiro, "Chemotherapy for patients with non-small cell lung cancer: the surgical setting of the Big Lung Trial," *European Journal of Cardio-Thoracic Surgery*, vol. 26, pp. 173–182, July 2004.

[42] G. M. Strauss, J. E. Herndon, M. A. Maddaus, D. W. Johnstone, E. A. Johnson, D. H. Harpole, H. H. Gillenwater, D. M. Watson, D. J. Sugarbaker, R. L. Schilsky, E. E. Vokes, and M. R. Green, "Adjuvant Paclitaxel Plus Carboplatin Compared With Observation in Stage IB Non–Small-Cell Lung Cancer: CALGB 9633 With the Cancer and Leukemia Group B, Radiation Therapy Oncology Group, and North Central Cancer Treatment Group Study Groups," *Journal of Clinical Oncology*, vol. 26, pp. 5043–5051, Nov. 2008.

[43] C.-Q. Zhu, K. Ding, D. Strumpf, B. A. Weir, M. Meyerson, N. Pennell, R. K. Thomas, K. Naoki, C. Ladd-Acosta, N. Liu, M. Pintilie, S. Der, L. Seymour, I. Jurisica, F. A. Shepherd, and M.-S. Tsao, "Prognostic and predictive gene signature for adjuvant chemotherapy in resected non-small-cell lung cancer," *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, vol. 28, pp. 4417–4424, Oct. 2010.

[44] P. Roepman, J. Jassem, E. F. Smit, T. Muley, J. Niklinski, T. van de Velde, A. T. Witteveen, W. Rzyman, A. Floore, S. Burgers, G. Giaccone, M. Meister, H. Dienemann, M. Skrzypski, M. Kozlowski, W. J. Mooi, and N. van Zandwijk, "An immune response enriched 72-gene prognostic profile for early-stage non-small-cell lung cancer," *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, vol. 15, pp. 284–290, Jan. 2009.

[45] L. He, Y. Huang, L. Yan, J. Zheng, C. Liang, and Z. Liu, "Radiomics-based predictive risk score: A scoring system for preoperatively predicting risk of lymph node metastasis in patients with resectable non-small cell lung cancer," *Chin J Cancer Res*, vol. 31, pp. 641–652, Aug. 2019.

[46] Y. Huang, Z. Liu, L. He, X. Chen, D. Pan, Z. Ma, C. Liang, J. Tian, and C. Liang, "Radiomics Signature: A Potential Biomarker for the Prediction of Disease-Free Survival in Early-Stage (I or II) Non—Small Cell Lung Cancer," *Radiology*, vol. 281, pp. 947–957, Dec. 2016.

[47] A. Hosny, C. Parmar, T. P. Coroller, P. Grossmann, R. Zeleznik, A. Kumar, J. Bussink, R. J. Gillies, R. H. Mak, and H. J. W. L. Aerts, "Deep learning for lung cancer prognostication: A retrospective multi-cohort radiomics study," *PLoS Medicine*, vol. 15, Nov. 2018.

[48] T. Hoang, R. Xu, J. H. Schiller, P. Bonomi, and D. H. Johnson, "Clinical Model to Predict Survival in Chemonaive Patients With Advanced Non–Small-Cell Lung

Cancer Treated With Third-Generation Chemotherapy Regimens Based on Eastern Cooperative Oncology Group Data," *Journal of Clinical Oncology*, vol. 23, pp. 175–183, Jan. 2005.

[49] S.-Q. Yuan, W.-J. Wu, M.-Z. Qiu, Z.-X. Wang, L.-P. Yang, Y. Jin, J.-P. Yun, Y.-H. Gao, Y.-H. Li, Z.-W. Zhou, F. Wang, and R.-H. Xu, "Development and Validation of a Nomogram to Predict the Benefit of Adjuvant Radiotherapy for Patients with Resected Gastric Cancer," *Journal of Cancer*, vol. 8, pp. 3498–3505, Sept. 2017.

[50] A. Mauguen, J.-P. Pignon, S. Burdett, C. Domerg, D. Fisher, R. Paulus, S. J. Mandrekar, C. P. Belani, F. A. Shepherd, T. Eisen, H. Pang, L. Collette, W. T. Sause, S. E. Dahlberg, J. Crawford, M. O'Brien, S. E. Schild, M. Parmar, J. F. Tierney, C. L. Pechoux, and S. Michiels, "Surrogate endpoints for overall survival in chemotherapy and radiotherapy trials in operable and locally advanced lung cancer: a re-analysis of meta-analyses of individual patients' data," *The Lancet Oncology*, vol. 14, pp. 619–626, June 2013.

[51] V. P. Balachandran, M. Gonen, J. J. Smith, and R. P. DeMatteo, "Nomograms in oncology: more than meets the eye," *The Lancet Oncology*, vol. 16, pp. e173–e180, Apr. 2015.

[52] G. Corredor, X. Wang, Y. Zhou, C. Lu, P. Fu, K. Syrigos, D. L. Rimm, M. Yang, E. Romero, K. A. Schalper, V. Velcheti, and A. Madabhushi, "Spatial Architecture and Arrangement of Tumor-Infiltrating Lymphocytes for Predicting Likelihood of Recurrence in Early-Stage Non-Small Cell Lung Cancer," *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, vol. 25, pp. 1526–1534, Mar. 2019.

[53] J. R. Kratz, J. He, S. K. Van Den Eeden, Z.-H. Zhu, W. Gao, P. T. Pham, M. S. Mulvihill, F. Ziaei, H. Zhang, B. Su, X. Zhi, C. P. Quesenberry, L. A. Habel, Q. Deng, Z. Wang, J. Zhou, H. Li, M.-C. Huang, C.-C. Yeh, M. R. Segal, M. R. Ray, K. D. Jones, D. J. Raz, Z. Xu, T. M. Jahan, D. Berryman, B. He, M. J. Mann, and D. M. Jablons, "A practical molecular assay to predict survival in resected non-squamous, non-small-cell lung cancer: development and international validation studies," *Lancet (London, England)*, vol. 379, pp. 823–832, Mar. 2012.

[54] M. Khorrami, P. Prasanna, A. Gupta, P. Patil, P. D. Velu, R. Thawani, G. Corredor, M. Alilou, K. Bera, P. Fu, M. Feldman, V. Velcheti, and A. Madabhushi, "Changes in CT radiomic features associated with lymphocyte distribution predict overall survival and response to immunotherapy in non-small cell lung cancer," *Cancer Immunology Research*, Jan. 2019.

[55] N. Braman, P. Prasanna, J. Whitney, S. Singh, N. Beig, M. Etesami, D. D. B. Bates, K. Gallagher, B. N. Bloch, M. Vulchi, P. Turk, K. Bera, J. Abraham, W. M. Sikov, G. Somlo, L. N. Harris, H. Gilmore, D. Plecha, V. Varadan, and A. Madabhushi, "Association of Peritumoral Radiomics With Tumor Biology and Pathologic Response to Preoperative Targeted Therapy for HER2 (ERBB2)-Positive Breast Cancer," *JAMA network open*, vol. 2, p. e192561, Apr. 2019.

[56] J. Morimoto, T. Nakajima, H. Suzuki, K. Nagato, T. Iwata, S. Yoshida, M. Fukuyo, S. Ota, Y. Nakatani, and I. Yoshino, "Impact of free tumor clusters on prognosis after resection of pulmonary adenocarcinoma," *The Journal of Thoracic and Cardiovascular Surgery*, vol. 152, pp. 64–72.e1, July 2016.

[57] S. Shiono and N. Yanagawa, "Spread through air spaces is a predictive factor of recurrence and a prognostic factor in stage I lung adenocarcinoma," *Interactive Cardiovascular and Thoracic Surgery*, vol. 23, no. 4, pp. 567–572, 2016.

[58] B. He, W. Zhao, J.-Y. Pi, D. Han, Y.-M. Jiang, Z.-G. Zhang, and W. Zhao, "A biomarker basing on radiomics for the prediction of overall survival in non–small cell lung cancer patients," *Respiratory Research*, vol. 19, 2018.

[59] S. Champiat, L. Dercle, S. Ammari, C. Massard, A. Hollebecque, S. Postel-Vinay, N. Chaput, A. Eggermont, A. Marabelle, J.-C. Soria, and C. Ferté, "Hyperprogressive Disease Is a New Pattern of Progression in Cancer Patients Treated by Anti-PD-1/PD-L1," *Clinical Cancer Research*, vol. 23, pp. 1920–1928, Apr. 2017.

[60] S. Kato, A. Goodman, V. Walavalkar, D. A. Barkauskas, A. Sharabi, and R. Kurzrock, "Hyperprogressors after Immunotherapy: Analysis of Genomic Alterations Associated with Accelerated Growth Rate," *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, vol. 23, pp. 4242–4250, Aug. 2017.

[61] E. Saâda-Bouzid, C. Defaucheux, A. Karabajakian, V. P. Coloma, V. Servois, X. Paoletti, C. Even, J. Fayette, J. Guigay, D. Loirat, F. Peyrade, M. Alt, J. Gal, and C. Le Tourneau, "Hyperprogression during anti-PD-1/PD-L1 therapy in patients with recurrent and/or metastatic head and neck squamous cell carcinoma," *Annals of Oncology: Official Journal of the European Society for Medical Oncology*, vol. 28, pp. 1605–1611, July 2017.

[62] R. Ferrara, L. Mezquita, M. Texier, J. Lahmar, C. Audigier-Valette, L. Tessonnier, J. Mazieres, G. Zalcman, S. Brosseau, S. L. Moulec, L. Leroy, B. Duchemann, C. Lefebvre, R. Veillon, V. Westeel, S. Koscielny, S. Champiat, C. Ferté, D. Planchard, J. Remon, M.-E. Boucher, A. Gazzah, J. Adam, E. Bria, G. Tortora, J.-C. Soria, B. Besse, and C. Caramella, "Hyperprogressive Disease in Patients With Advanced Non–Small Cell Lung Cancer Treated With PD-1/PD-L1 Inhibitors or With Single-Agent Chemotherapy," *JAMA Oncology*, vol. 4, pp. 1543–1552, Nov. 2018.

[63] Y. Kim, C. H. Kim, H. S. Kim, J.-M. Sun, J. S. Ahn, M.-J. Ahn, S.-H. Lee, H. Y. Lee, and K. Park, "Hyperprogression after immunotherapy: Clinical implication and genomic alterations in advanced non-small cell lung cancer patients (NSCLC).," *Journal of Clinical Oncology*, vol. 36, pp. 9075–9075, May 2018.

[64] G. Lo Russo, M. Moro, M. Sommariva, V. Cancila, M. Boeri, G. Centonze, S. Ferro, M. Ganzinelli, P. Gasparini, V. Huber, M. Milione, L. Porcu, C. Proto, G. Pruneri, D. Signorelli, S. Sangaletti, L. Sfondrini, C. Storti, E. Tassi, A. Bardelli, S. Marsoni, V. Torri, C. Tripodo, M. P. Colombo, A. Anichini, L. Rivoltini, A. Balsari, G. Sozzi,

and M. C. Garassino, "Antibody-Fc/FcR Interaction on Macrophages as a Mechanism for Hyperprogressive Disease in Non-small Cell Lung Cancer Subsequent to PD-1/PD-L1 Blockade," *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, vol. 25, no. 3, pp. 989–999, 2019.

[65] S. C. M. Lau and N. B. Leighl, "Hyperprogressive disease with immunotherapy: new directions," *Journal of Thoracic Disease*, vol. 11, pp. S1877–S1880, Sept. 2019.

[66] Y. Kanjanapan, D. Day, L. Wang, A. R. Hansen, A. R. Abdul Razak, A. Spreafico, N. B. Leighl, A. M. Joshua, M. O. Butler, D. Hogg, L. L. Siu, and P. L. Bedard, "Hyperprogressive disease (HPD) in early-phase immunotherapy (IO) trials.," *Journal of Clinical Oncology*, vol. 36, pp. 3063–3063, May 2018. Publisher: American Society of Clinical Oncology.

[67] H. J. W. L. Aerts, P. Grossmann, Y. Tan, G. G. Oxnard, N. Rizvi, L. H. Schwartz, and B. Zhao, "Defining a Radiomic Response Phenotype: A Pilot Study using targeted therapy in NSCLC," *Scientific Reports*, vol. 6, p. 33860, Sept. 2016.

[68] T. P. Coroller, V. Agrawal, E. Huynh, V. Narayan, S. W. Lee, R. H. Mak, and H. J. W. L. Aerts, "Radiomic-Based Pathological Response Prediction from Primary Tumors and Lymph Nodes in NSCLC," *Journal of Thoracic Oncology: Official Publication of the International Association for the Study of Lung Cancer*, vol. 12, pp. 467–476, Mar. 2017.

[69] N. Beig, M. Khorrami, M. Alilou, P. Prasanna, N. Braman, M. Orooji, S. Rakshit, K. Bera, P. Rajiah, J. Ginsberg, C. Donatelli, R. Thawani, M. Yang, F. Jacono, P. Tiwari, V. Velcheti, R. Gilkeson, P. Linden, and A. Madabhushi, "Perinodular and Intranodular Radiomic Features on Lung CT Images Distinguish Adenocarcinomas from Granulomas," *Radiology*, vol. 290, pp. 783–792, Mar. 2019.

[70] M. Khorrami, P. Jain, K. Bera, M. Alilou, R. Thawani, P. Patil, U. Ahmad, S. Murthy, K. Stephans, P. Fu, V. Velcheti, and A. Madabhushi, "Predicting pathologic response to neoadjuvant chemoradiation in resectable stage III non-small cell lung cancer patients using computed tomography radiomic features," *Lung Cancer (Amsterdam, Netherlands)*, vol. 135, pp. 1–9, Sept. 2019.

[71] N. M. Braman, M. Etesami, P. Prasanna, C. Dubchuk, H. Gilmore, P. Tiwari, D. Pletcha, and A. Madabhushi, "Intratumoral and peritumoral radiomics for the pretreatment prediction of pathological complete response to neoadjuvant chemotherapy based on breast DCE-MRI," *Breast Cancer Research : BCR*, vol. 19, 2017.

[72] M. Guo, B. S. Rabin, J. T. Johnson, and I. L. Paradis, "Lymphocyte phenotypes at tumor margins in patients with head and neck cancer," *Head & Neck Surgery*, vol. 9, pp. 265–271, June 1987.

[73] A. Kotrotsou, P. O. Zinn, and R. R. Colen, "Radiomics in Brain Tumors: An Emerging Technique for Characterization of Tumor Environment," *Magnetic Resonance Imaging Clinics of North America*, vol. 24, pp. 719–729, Nov. 2016.

[74] N. G. Beig, V. Hill, R. Verma, V. Varadan, A. Madabhushi, P. Tiwari, and P. Prasanna, "Radiogenomic characterization of response to chemo-radiation therapy in glioblastoma is associated with PI3K/AKT/mTOR and apoptosis signaling pathways," in *Medical Imaging 2019: Computer-Aided Diagnosis* (H. K. Hahn and K. Mori, eds.), (San Diego, United States), p. 47, SPIE, Mar. 2019.

[75] N. Beig, K. Bera, P. Prasanna, J. Antunes, R. Correa, S. Singh, A. Saeed Bamashmos, M. Ismail, N. Braman, R. Verma, V. B. Hill, V. Statsevych, M. S. Ahluwalia, V. Varadan, A. Madabhushi, and P. Tiwari, "Radiogenomic-Based Survival Risk Stratification of Tumor Habitat on Gd-T1w MRI Is Associated with Biological Processes in Glioblastoma," *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, vol. 26, pp. 1866–1876, Apr. 2020.

[76] P. Kickingereder, M. Götz, J. Muschelli, A. Wick, U. Neuberger, R. T. Shinohara, M. Sill, M. Nowosielski, H.-P. Schlemmer, A. Radbruch, W. Wick, M. Bendszus, K. H. Maier-Hein, and D. Bonekamp, "Large-scale Radiomic Profiling of Recurrent Glioblastoma Identifies an Imaging Predictor for Stratifying Anti-Angiogenic Treatment Response," *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, vol. 22, pp. 5765–5771, Dec. 2016.

[77] J. Wang, C.-J. Wu, M.-L. Bao, J. Zhang, X.-N. Wang, and Y.-D. Zhang, "Machine learning-based analysis of MR radiomics can help to improve the diagnostic performance of PI-RADS v2 in clinically relevant prostate cancer," *European Radiology*, Apr. 2017.

[78] R. Stoyanova, M. Takhar, Y. Tschudi, J. C. Ford, G. Solórzano, N. Erho, Y. Balagurunathan, S. Punnen, E. Davicioni, R. J. Gillies, and A. Pollack, "Prostate cancer radiomics and the promise of radiogenomics," *Translational Cancer Research*, vol. 5, pp. 432–447, Aug. 2016.

[79] K. Bera, R. Thawani, J.-P. Achkar, C. Fiocchi, M. Kay, R. Gupta, J. Kurowski, S. Viswanath, and I. Barbur, "Automated segmentation and radiomic characterization of visceral fat on bowel MRIs for Crohn's disease," in *Medical Imaging 2018: Image-Guided Procedures, Robotic Interventions, and Modeling* (R. J. Webster and B. Fei, eds.), (Houston, United States), p. 32, SPIE, Mar. 2018.

[80] S. Trebeschi, I. Kurilova, A. M. Călin, D. M. J. Lambregts, E. F. Smit, H. Aerts, and R. G. Beets-Tan, "Radiomic biomarkers for the prediction of immunotherapy outcome in patients with metastatic non-small cell lung cancer.," *Journal of Clinical Oncology*, vol. 35, pp. e14520–e14520, May 2017.

[81] S. Trebeschi, S. Drago, N. Birkbak, I. Kurilova, A. Călin, A. Delli Pizzi, F. Lalezari, D. Lambregts, M. Rohaan, C. Parmar, E. Rozeman, K. Hartemink, C. Swanton, J. B. A. G. Haanen, C. Blank, E. Smit, R. Beets-Tan, and H. Aerts, "Predicting response to cancer immunotherapy using noninvasive radiomic biomarkers," *Annals of Oncology*, vol. 30, pp. 998–1004, June 2019.

[82] M. Alilou, M. Orooji, N. Beig, P. Prasanna, P. Rajiah, C. Donatelli, V. Velcheti, S. Rakshit, M. Yang, F. Jacono, R. Gilkeson, P. Linden, and A. Madabhushi, "Quantitative vessel tortuosity: A potential CT imaging biomarker for distinguishing lung granulomas from adenocarcinomas," *Scientific Reports*, vol. 8, p. 15290, Oct. 2018.

[83] M. Alilou, P. Vaidya, M. Khorrami, A. Zagouras, P. Patil, K. Bera, P. Fu, V. Velcheti, and A. Madabhushi, "Quantitative vessel tortuosity radiomics on baseline non-contrast lung CT predict response to immunotherapy and are prognostic of overall survival," in *Medical Imaging 2019: Computer-Aided Diagnosis*, vol. 10950, p. 109501F, International Society for Optics and Photonics, Mar. 2019.

[84] E. A. Eisenhauer, P. Therasse, J. Bogaerts, L. H. Schwartz, D. Sargent, R. Ford, J. Dancey, S. Arbuck, S. Gwyther, M. Mooney, L. Rubinstein, L. Shankar, L. Dodd, R. Kaplan, D. Lacombe, and J. Verweij, "New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1)," *European Journal of Cancer (Oxford, England: 1990)*, vol. 45, pp. 228–247, Jan. 2009.

[85] H. Borghaei, L. Paz-Ares, L. Horn, D. R. Spigel, M. Steins, N. E. Ready, L. Q. Chow, E. E. Vokes, E. Felip, E. Holgado, F. Barlesi, M. Kohlhäufl, O. Arrieta, M. A. Burgio, J. Fayette, H. Lena, E. Poddubskaya, D. E. Gerber, S. N. Gettinger, C. M. Rudin, N. Rizvi, L. Crinò, G. R. Blumenschein, S. J. Antonia, C. Dorange, C. T. Harbison, F. Graf Finckenstein, and J. R. Brahmer, "Nivolumab versus Docetaxel in Advanced Nonsquamous Non–Small-Cell Lung Cancer," *New England Journal of Medicine*, vol. 373, pp. 1627–1639, Oct. 2015.

[86] S. Champiat, R. Ferrara, C. Massard, B. Besse, A. Marabelle, J.-C. Soria, and C. Ferté, "Hyperprogressive disease: recognizing a novel pattern to improve patient management," *Nature Reviews Clinical Oncology*, vol. 15, pp. 748–762, Dec. 2018. Number: 12 Publisher: Nature Publishing Group.

[87] M. Khorrami, M. Khunger, A. Zagouras, P. Patil, R. Thawani, K. Bera, P. Rajiah, P. Fu, V. Velcheti, and A. Madabhushi, "Combination of Peri- and Intratumoral Radiomic Features on Baseline CT Scans Predicts Response to Chemotherapy in Lung Adenocarcinoma," *Radiology. Artificial Intelligence*, vol. 1, Mar. 2019.

[88] I. Tunali, J. E. Gray, J. Qi, M. Abdalah, D. K. Jeong, A. Guvenis, R. J. Gillies, and M. B. Schabath, "Novel clinical and radiomic predictors of rapid disease progression phenotypes among lung cancer patients treated with immunotherapy: An early report," *Lung Cancer (Amsterdam, Netherlands)*, vol. 129, pp. 75–79, 2019.

[89] Y. Sun, P. Hu, J. Wang, L. Shen, F. Xia, G. Qing, W. Hu, Z. Zhang, C. Xin, W. Peng, T. Tong, and Y. Gu, "Radiomic features of pretreatment MRI could identify T stage in patients with rectal cancer: Preliminary findings," *Journal of magnetic resonance imaging: JMRI*, Feb. 2018.

[90] Y. Zhang, A. Oikonomou, A. Wong, M. A. Haider, and F. Khalvati, "Radiomics-based Prognosis Analysis for Non-Small Cell Lung Cancer," *Scientific Reports*, vol. 7, Apr. 2017.

[91] I. Fogel and D. Sagi, "Gabor filters as texture discriminator," *Biological Cybernetics*, vol. 61, pp. 103–113, June 1989.

[92] P. Prasanna, P. Tiwari, and A. Madabhushi, "Co-occurrence of local anisotropic gradient orientations (CoLIAGe): distinguishing tumor confounders and molecular subtypes on MRI.," *Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention*, vol. 17, no. Pt 3, pp. 73–80, 2014.

[93] R. M. Haralick, "Statistical and structural approaches to texture," *Proceedings of the IEEE*, vol. 67, no. 5, pp. 786–804, 1979.

[94] H. K. Bhargava, P. Leo, R. Elliott, A. Janowczyk, J. Whitney, S. Gupta, P. Fu, K. Yamoah, F. Khani, B. D. Robinson, T. R. Rebbeck, M. Feldman, P. Lal, and A. Madabhushi, "Computationally Derived Image Signature of Stromal Morphology Is Prognostic of Prostate Cancer Recurrence Following Prostatectomy in African American Patients," *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, vol. 26, pp. 1915–1923, Apr. 2020.

[95] K. Bera, K. A. Schalper, D. L. Rimm, V. Velcheti, and A. Madabhushi, "Artificial intelligence in digital pathology - new tools for diagnosis and precision oncology," *Nature Reviews. Clinical Oncology*, vol. 16, no. 11, pp. 703–715, 2019.

[96] C. Lu, C. Koyuncu, G. Corredor, P. Prasanna, P. Leo, X. Wang, A. Janowczyk, K. Bera, J. Lewis Jr., V. Velcheti, and A. Madabhushi, "Feature-driven local cell graph (FLocK): New computational pathology-based descriptors for prognosis of lung cancer and HPV status of oropharyngeal cancers," *Medical Image Analysis*, vol. 68, p. 101903, Feb. 2021.

[97] L. Fournier, L. Costaridou, L. Bidaut, N. Michoux, F. E. Lecouvet, L.-F. de Geus-Oei, R. Boellaard, D. E. Oprea-Lager, N. A. Obuchowski, A. Caroli, W. G. Kunz, E. H. Oei, J. P. B. O'Connor, M. E. Mayerhoefer, M. Franca, A. Alberich-Bayarri, C. M. Deroose, C. Loewe, R. Manniesing, C. Caramella, E. Lopci, N. Lassau, A. Persson, R. Achten, K. Rosendahl, O. Clement, E. Kotter, X. Golay, M. Smits, M. Dewey, D. C. Sullivan, A. van der Lugt, N. M. deSouza, and E. S. of Radiology, "Incorporating radiomics into clinical trials: expert consensus endorsed by the European Society of Radiology on considerations for data-driven compared to biologically driven quantitative biomarkers," *European Radiology*, vol. 31, no. 8, pp. 6001–6012, 2021.

[98] M. Khorrami, K. Bera, P. Leo, P. Vaidya, P. Patil, R. Thawani, P. Velu, P. Rajiah, M. Alilou, H. Choi, M. D. Feldman, R. C. Gilkeson, P. Linden, P. Fu, H. Pass, V. Velcheti, and A. Madabhushi, "Stable and discriminating radiomic predictor of recurrence in early stage non-small cell lung cancer: Multi-site study," *Lung Cancer (Amsterdam, Netherlands)*, vol. 142, pp. 90–97, Apr. 2020.

[99] J. Whitney, G. Corredor, A. Janowczyk, S. Ganesan, S. Doyle, J. Tomaszewski, M. Feldman, H. Gilmore, and A. Madabhushi, "Quantitative nuclear histomorphom-

etry predicts oncotype DX risk categories for early stage ER+ breast cancer," *BMC cancer*, vol. 18, p. 610, May 2018.

[100] G. Corredor, C. Lu, C. Koyuncu, K. Bera, P. Toro, P. Fu, S. A. Koyfman, D. Chute, D. J. Adelstein, W. Thorstad, J. A. Bishop, F. Faraji, J. Lewis, and A. Madabhushi, "Computerized features of spatial interplay of tumor-infiltrating lymphocytes predict disease recurrence in p16+ oropharyngeal squamous cell carcinoma: A multisite validation study.," *Journal of Clinical Oncology*, vol. 38, pp. 6559–6559, May 2020.

[101] H. J. W. L. Aerts, E. R. Velazquez, R. T. H. Leijenaar, C. Parmar, P. Grossmann, S. Carvalho, S. Cavalho, J. Bussink, R. Monshouwer, B. Haibe-Kains, D. Rietveld, F. Hoebers, M. M. Rietbergen, C. R. Leemans, A. Dekker, J. Quackenbush, R. J. Gillies, and P. Lambin, "Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach," *Nature Communications*, vol. 5, p. 4006, 2014.

[102] M. Bilal, S. E. A. Raza, A. Azam, S. Graham, M. Ilyas, I. A. Cree, D. Snead, F. Minhas, and N. M. Rajpoot, "Development and validation of a weakly supervised deep learning framework to predict the status of molecular pathways and key mutations in colorectal cancer from routine histology images: a retrospective study," *The Lancet Digital Health*, vol. 3, pp. e763–e772, Dec. 2021. Publisher: Elsevier.

[103] L. Wan, Z. Sun, W. Peng, S. Wang, J. Li, Q. Zhao, S. Wang, H. Ouyang, X. Zhao, S. Zou, and H. Zhang, "Selecting Candidates for Organ-Preserving Strategies After Neoadjuvant Chemoradiotherapy for Rectal Cancer: Development and Validation of a Model Integrating MRI Radiomics and Pathomics," *Journal of magnetic resonance imaging: JMRI*, Feb. 2022.

[104] C. Lu, K. Bera, X. Wang, P. Prasanna, J. Xu, A. Janowczyk, N. Beig, M. Yang, P. Fu, J. Lewis, H. Choi, R. A. Schmid, S. Berezowska, K. Schalper, D. Rimm, V. Velcheti, and A. Madabhushi, "A prognostic model for overall survival of patients with early-stage non-small cell lung cancer: a multicentre, retrospective study," *The Lancet. Digital Health*, vol. 2, no. 11, pp. e594–e606, 2020.

[105] S. Azarianpour, G. Corredor, K. Bera, P. Leo, P. Fu, P. Toro, A. Joehlin-Price, M. Mokhtari, H. Mahdi, and A. Madabhushi, "Computational image features of immune architecture is associated with clinical benefit and survival in gynecological cancers across treatment modalities," *Journal for ImmunoTherapy of Cancer*, vol. 10, p. e003833, Feb. 2022.

[106] Y. Liu, B. Mao, S. Liang, J.-W. Yang, H.-W. Lu, Y.-H. Chai, L. Wang, L. Zhang, Q.-H. Li, L. Zhao, Y. He, X.-L. Gu, X.-B. Ji, L. Li, Z.-J. Jie, Q. Li, X.-Y. Li, H.-Z. Lu, W.-H. Zhang, Y.-L. Song, J.-M. Qu, and J.-F. Xu, "Association between age and clinical characteristics and outcomes of COVID-19," *The European Respiratory Journal*, vol. 55, May 2020.

[107] C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, L. Zhang, G. Fan, J. Xu, X. Gu, Z. Cheng, T. Yu, J. Xia, Y. Wei, W. Wu, X. Xie, W. Yin, H. Li, M. Liu, Y. Xiao,

H. Gao, L. Guo, J. Xie, G. Wang, R. Jiang, Z. Gao, Q. Jin, J. Wang, and B. Cao, "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China," *Lancet (London, England)*, vol. 395, no. 10223, pp. 497–506, 2020.

[108] A. O'Dowd, "Covid-19: Cases of delta variant rise by 79%, but rate of growth slows," *BMJ*, vol. 373, p. n1596, June 2021. Publisher: British Medical Journal Publishing Group Section: News.

[109] E. Anthes, "The Delta Variant: What Scientists Know," *The New York Times*, June 2021.

[110] "Coronavirus disease 2019 (COVID-19): Critical care and airway management issues - UpToDate."

[111] E. Callaway, "Delta coronavirus variant: scientists brace for impact," *Nature*, June 2021. Bandiera_abtest: a Cg_type: News Publisher: Nature Publishing Group Subject_term: SARS-CoV-2, Public health, Virology.

[112] CDC, "Healthcare Workers," Feb. 2020.

[113] G. D. Rubin, C. J. Ryerson, L. B. Haramati, N. Sverzellati, J. P. Kanne, S. Raoof, N. W. Schluger, A. Volpi, J.-J. Yim, I. B. Martin, D. J. Anderson, C. Kong, T. Altes, A. Bush, S. R. Desai, J. Goldin, J. M. Goo, M. Humbert, Y. Inoue, H.-U. Kauczor, F. Luo, P. J. Mazzone, M. Prokop, M. Remy-Jardin, L. Richeldi, C. M. Schaefer-Prokop, N. Tomiyama, A. U. Wells, and A. N. Leung, "The Role of Chest Imaging in Patient Management During the COVID-19 Pandemic," *Chest*, vol. 158, pp. 106–116, July 2020.

[114] B. Xu, Y. Xing, J. Peng, Z. Zheng, W. Tang, Y. Sun, C. Xu, and F. Peng, "Chest CT for detecting COVID-19: a systematic review and meta-analysis of diagnostic accuracy," *European Radiology*, pp. 1–8, May 2020.

[115] Y. Chen, A. Janowczyk, and A. Madabhushi, "Quantitative Assessment of the Effects of Compression on Deep Learning in Digital Pathology Image Analysis," *JCO clinical cancer informatics*, vol. 4, pp. 221–233, Mar. 2020.

[116] H. Chao, X. Fang, J. Zhang, F. Homayounieh, C. D. Arru, S. R. Digumarthy, R. Babaei, H. K. Mobin, I. Mohseni, L. Saba, A. Carriero, Z. Falaschi, A. Pasche, G. Wang, M. K. Kalra, and P. Yan, "Integrative analysis for COVID-19 patient outcome prediction," *Medical Image Analysis*, vol. 67, p. 101844, Jan. 2021.

[117] Q. Wu, S. Wang, L. Li, Q. Wu, W. Qian, Y. Hu, L. Li, X. Zhou, H. Ma, H. Li, M. Wang, X. Qiu, Y. Zha, and J. Tian, "Radiomics Analysis of Computed Tomography helps predict poor prognostic outcome in COVID-19," *Theranostics*, vol. 10, pp. 7231–7244, June 2020.

[118] F. Homayounieh, S. Ebrahimian, R. Babaei, H. Karimi Mobin, E. Zhang, B. C. Bizzo, I. Mohseni, S. R. Digumarthy, and M. K. Kalra, "CT Radiomics, Radiologists and

Clinical Information in Predicting Outcome of Patients with COVID-19 Pneumonia," *Radiology: Cardiothoracic Imaging*, vol. 2, p. e200322, July 2020. Publisher: Radiological Society of North America.

[119] "A model based on CT radiomic features for predicting RT-PCR becoming negative in coronavirus disease 2019 (COVID-19) patients."

[120] L. Wynants, B. Van Calster, M. M. J. Bonten, G. S. Collins, T. P. A. Debray, M. De Vos, M. C. Haller, G. Heinze, K. G. M. Moons, R. D. Riley, E. Schuit, L. J. M. Smits, K. I. E. Snell, E. W. Steyerberg, C. Wallisch, and M. van Smeden, "Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal," *The BMJ*, vol. 369, Apr. 2020.

[121] P. Lambin, E. Rios-Velazquez, R. Leijenaar, S. Carvalho, R. G. P. M. van Stiphout, P. Granton, C. M. L. Zegers, R. Gillies, R. Boellard, A. Dekker, and H. J. W. L. Aerts, "Radiomics: extracting more information from medical images using advanced feature analysis," *European Journal of Cancer (Oxford, England: 1990)*, vol. 48, pp. 441–446, Mar. 2012.

[122] R. Muthukrishnan and R. Rohini, "LASSO: A feature selection technique in predictive modeling for machine learning," in *2016 IEEE International Conference on Advances in Computer Applications (ICACA)*, pp. 18–20, Oct. 2016.

[123] J. Hu, J. Zhou, F. Dong, J. Tan, S. Wang, Z. Li, X. Zhang, H. Zhang, J. Ming, and T. Huang, "Combination of serum lactate dehydrogenase and sex is predictive of severe disease in patients with COVID-19," *Medicine*, vol. 99, Oct. 2020.

[124] M. Kermali, R. K. Khalsa, K. Pillai, Z. Ismail, and A. Harky, "The role of biomarkers in diagnosis of COVID-19 – A systematic review," *Life Sciences*, vol. 254, p. 117788, Aug. 2020.

[125] A. J. Vickers, B. van Calster, and E. W. Steyerberg, "A simple, step-by-step guide to interpreting decision curve analysis," *Diagnostic and Prognostic Research*, vol. 3, p. 18, Oct. 2019.

[126] M. A. Kazemi, H. Ghanaati, B. Moradi, M. Chavoshi, H. Hashemi, S. Hemmati, P. Rouzrokh, M. Gity, Z. Ahmadinejad, and H. Abdollahi, "Prognostic factors of chest CT findings for ICU admission and mortality in patients with COVID-19 pneumonia," *medRxiv*, p. 2020.10.30.20223024, Nov. 2020. Publisher: Cold Spring Harbor Laboratory Press.

[127] L. Fu, Y. Li, A. Cheng, P. Pang, and Z. Shu, "A Novel Machine Learning-derived Radiomic Signature of the Whole Lung Differentiates Stable From Progressive COVID-19 Infection," *Journal of Thoracic Imaging*, vol. 35, pp. 361–368, Nov. 2020.

[128] N. D. Yanez, N. S. Weiss, J.-A. Romand, and M. M. Treggiari, "COVID-19 mortality risk for older men and women," *BMC Public Health*, vol. 20, p. 1742, Nov. 2020.

[129] Y. Yu, X. Wang, M. Li, L. Gu, Z. Xie, W. Gu, F. Xu, Y. Bao, R. Liu, S. Hu, M. Hu, and C. Hu, "Nomogram to identify severe coronavirus disease 2019 (COVID-19) based on initial clinical and CT characteristics: a multi-center study," *BMC Medical Imaging*, vol. 20, p. 111, Oct. 2020.

[130] M. Roberts, D. Driggs, M. Thorpe, J. Gilbey, M. Yeung, S. Ursprung, A. I. Aviles-Rivero, C. Etmann, C. McCague, L. Beer, J. R. Weir-McCall, Z. Teng, E. Gkrania-Klotsas, J. H. F. Rudd, E. Sala, and C.-B. Schönlieb, "Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans," *Nature Machine Intelligence*, vol. 3, pp. 199–217, Mar. 2021. Number: 3 Publisher: Nature Publishing Group.