

2022

Analysis of clinical and radiomic factors associated with intermediately-categorized pulmonary nodule lung-rads risk progression

<https://hdl.handle.net/2144/45215>

"Downloaded from OpenBU. Boston University's institutional repository."

BOSTON UNIVERSITY
SCHOOL OF MEDICINE

Thesis

**ANALYSIS OF CLINICAL AND RADIOMIC FACTORS ASSOCIATED WITH
INTERMEDIATELY-CATEGORIZED PULMONARY NODULE LUNG-RADS
RISK PROGRESSION**

by

CAMERON G. HURLBURT

B.S. Campbell University, 2020

Submitted in partial fulfillment of the
requirements for the degree of
Master of Science

2022

© 2022 by
CAMERON G. HURLBURT
All rights reserved

Approved by

First Reader

Kei Suzuki, M.D.
Assistant Professor of Surgery
Boston University School of Medicine

Second Reader

Kitt Shaffer, M.D., Ph.D.
Professor of Radiology and Anatomy and Neurobiology
Boston University School of Medicine

DEDICATION

I would like to acknowledge the primary personnel that have helped me get to this point in my academic career on this page. First off, I would like to thank my Mom, Dad and my stepfather Donnie. Without you, I would not have been raised to be the person I am today. School has never come naturally easy to me – in high school I floated by with C's and only was really focused on athletics. In my first semester of college, I received all B's and was so proud of myself. I accomplished that semester because of the trust and faith you all had in me as a student and as a person. The next semester I got straight A's and ended up switching my major from kinesiology to a more physical science-based Biology field where I felt I would be challenged slightly more. You all helped me realize that if I worked hard enough, asked for help and believed in myself, that I could be academically successful.

I got into the medical school at Boston University to complete my Master's Degree in a field I have always been interested in – anatomy and neuroscience. While there have been struggles, financially and academically, I know that everything will work itself out and be okay in the long run because I believe in myself, and I know I have an extraordinary support system at home.

I would also like to thank my grandmother, great Aunt Gogo, and my cousins Anne Marie and Jerry who have been consistently generous in helping fund my education my whole life. These gestures have never gone unnoticed, and I appreciate all your help endlessly.

**ANALYSIS OF CLINICAL AND RADIOMIC FACTORS ASSOCIATED WITH
INTERMEDIATELY-CATEGORIZED PULMONARY NODULE LUNG-RADS
RISK PROGRESSION**

CAMERON G. HURLBURT

ABSTRACT

Lung cancer currently has the greatest mortality rate of cancer patients of all sexes in the United States (Torre et al., 2016). Low-dose CT scans are utilized for lung cancer screening in patients who fall within the NLST entry criteria (Sanchez-Salcedo et al., 2015). The original criteria for screening were age over 55 and pack-year over 30, which were recently changed to age 50 and pack-year over 20 in 2021. The study population in this paper utilized the original criteria.

A system developed and copyrighted by the American College of Radiology (ACR) referred to as the Lung CT Screening Reporting and Data System (Lung-RADS) has implemented a standardized method of classifying and interpreting lung cancer chest CT screening results. Lung-RADS has a scoring system which is scaled 1 – 4x (Pinsky et al., 2015) The likelihood of malignancy based on nodule appearance, diameter, and presence of growth comprise the components of which score is given (Chung et al., 2017). Lung-RADS 2 scored nodules are benign nodules and patients follow up for another CT in a year. Lung-RADS 3 nodules are probably benign nodules; however, they do have a low-risk of malignancy. It is known that a select few of these relatively benign appearing nodules will

turn out to be malignant. Lung-RADS 4 nodules have a >5% chance of malignancy and can be confirmed through pathology.

In this project, a retrospective chart review analyzing patient demographics and pulmonary health history will be correlated to lung-RADS risk likelihood of malignancy. Machine learning will also be utilized to study and analyze radiographic factors associated with the sample. The CT scans of patients who previously scored in an intermediate category will be compiled and analyzed to determine potential common demographical, clinical and radiomic factors which will hopefully allow intermediately categorized nodule indicators to be used to detect cancers earlier and to more accurately classify lesions into benign or malignant categories.

In all, the goal of this research is to determine common clinical, demographic and radiomic factors of patients who were deemed intermediate risk and then progressed to a higher categorization. The importance of expanding current risk factors for discrimination of benign from malignant will also be analyzed, along with those specific risk-factors within Lung-RADS intermediately categorized nodules. The characteristics and baseline co-morbidities of RADS 2 and 3 lung cancer patients by follow-up CT results, progression to RADS-4 on follow-up CT and lung cancer diagnosis will be compiled and exemplified.

TABLE OF CONTENTS

DEDICATION.....iv

ABSTRACT.....v

TABLE OF CONTENTS.....vii

LIST OF TABLES.....viii

LIST OF FIGURES.....ix

LIST OF ABBREVIATIONS.....x

INTRODUCTION.....1

 United States Demographics of Lung Cancer.....1

 Global Prevalence of Lung Cancer.....4

 Lung Cancer Etiology.....5

 LDCT Screening Efficacy.....6

 Lung Nodules.....7

 Lung-RADS Scoring System.....8

METHODS

 Clinical Data via Retrospective Chart Review.....13

 Radiomics.....14

RESULTS.....16

DISCUSSION.....24

REFERENCES.....27

CURRICULUM VITAE.....31

LIST OF TABLES

Table 1: Characteristics of RADS 2 &3 lung cancer patients by follow-up CT results.....	16
Table 2: Characteristics of RADS 2 &3 lung cancer patients by progression to RADS-4 on follow-up CT.....	17
Table 3: Characteristics of RADS 2 &3 lung cancer patients by lung cancer diagnosis.....	18
Table 4: Baseline Co-Morbidities of RADS 2 &3 lung cancer patients by follow-up CT results.....	19
Table 5: Baseline Co-Morbidities of RADS 2 &3 lung cancer patients by progression to RADS-4 on follow-up CT.....	20
Table 6: Baseline Co-Morbidities of RADS 2 &3 lung cancer patients by lung cancer diagnosis.....	21
Table 7: Multivariate analysis - the effect of characteristics of RADS 2 & 3 lung cancer patients on progression to RADS-4 on follow-up CT (n=908)	23
Table 8: Multivariate analysis - the effect of characteristics of RADS 2 & 3 lung cancer patients on lung cancer diagnosis (n=2174).....	23
Table 9. AUC of Machine Learning Algorithms Trained with Radiomic Features.....	24

LIST OF FIGURES

Figure 1: Incidents and mortalities of lung cancer in the U.S. by race from 2014-2018....	2
Figure 2. Incident and mortality rate of lung cancer in the U.S. of race and sex from 2014-2018.....	3
Figure 3: Lung RADS Version 1.1 Assessment Categories	10
Figure 4: Consort diagram depicting the selection process for the retrospective chart review.....	13
Figure 5: Annotated image of a pulmonary nodule used in our research.....	15
Figure 6: Radiomic Analysis Graphs.....	23

LIST OF ABBREVIATIONS

ACR.....	American College of Radiology
BMC.....	Boston Medical Center
GLOBOCAN.....	Global Cancer Observatory
IHD.....	Ischemic Heart Disease
LDCT.....	Low dose computed tomography
Lung-RADS.....	Lung imaging reporting and data system

INTRODUCTION

United States Demographics of Lung Cancer

Lung cancer is the most abundant form of cancer-related mortality of both men and women in the United States (Dela Cruz et al., 2011). Per the United States Centers for Disease and Control and Prevention (CDC), from 2014-2018 the top three cancer types by evaluation of new cancer cases are female breast, prostate, and lung cancer, respectively (CDC 2021). On the other hand, lung cancer is substantially more fatal than both prostate and breast cancer, exemplifying a mortality rate in the United States of 38.5 per every 100,000 people compared to 20.1 and 19.0 per 100,000 for female breast and prostate cancer, respectively (CDC 2021).

Through a demographical race analysis of lung cancer history within the United States, it is shown that incidents are nearly identical amongst white and black individuals 58.2 and 58.1 cases per 100,000 individuals, respectively (CDC, 2021). The other analyzed races (American-Indian, Alaska native, Asian, and Pacific Islanders, and Hispanic populations) are much lower in the ranking as exemplified in figure 1 (below). Reasons for this could include genetic factors and/or smoking customs among different groups.

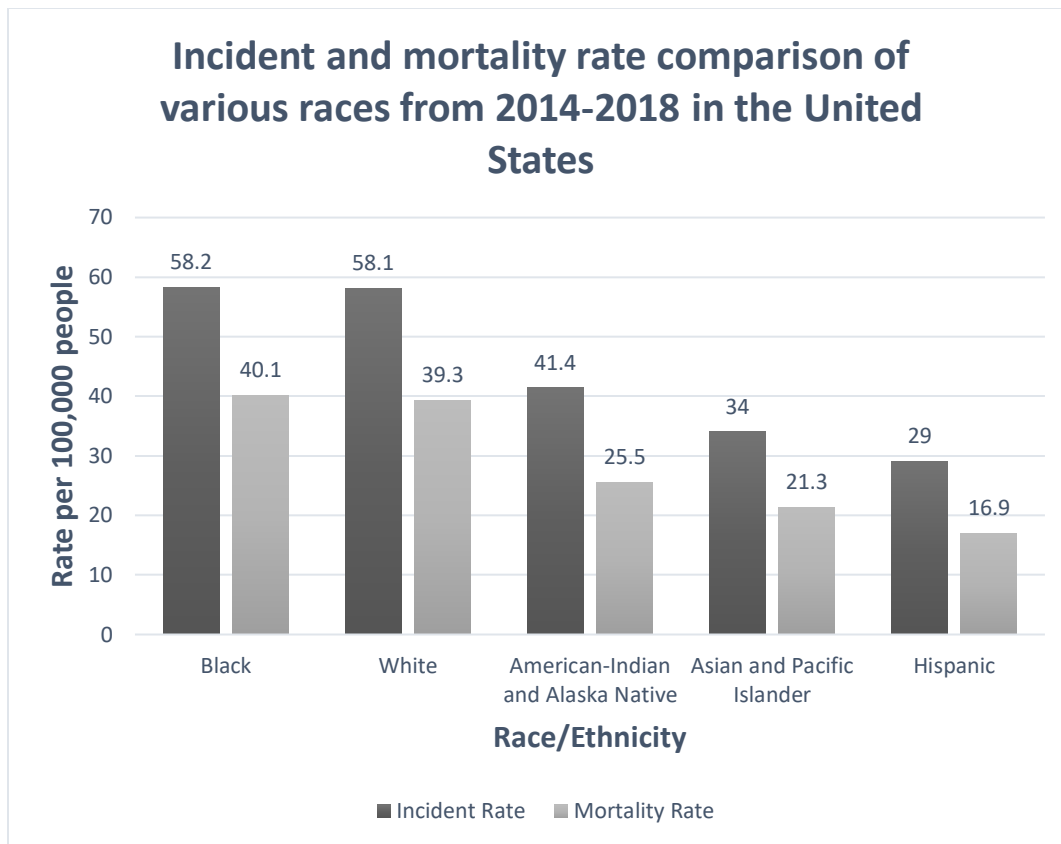


Figure 1: Incident and mortality rate of lung cancer in the U.S. by race from 2014-2018. The black color shows incident rate and the gray color shows mortality rate. 58.2 in 100,000 white individuals are diagnosed with lung cancer. 58.1 black individuals in every 100,000 people will be diagnosed with lung cancer. American Indian and Alaska-native, Asian, pacific islander and Hispanic races had a less incident rate from 2014-2018. Black individuals have the highest mortality rate followed by white individuals with a noticeably relevant decrease for other races. The data used to make this graph comes from the United States Center for Disease and Control. *U.S. Cancer Statistics Working Group. U.S. Cancer Statistics Data Visualizations Tool, based on 2020 submission data (1999-2018): U.S. Department of Health and Human Services, Centers for Disease Control and Prevention and National Cancer Institute; www.cdc.gov/cancer/dataviz, released in June 2021.*

Shown in figure 2. (below) is an incident and mortality rate chart by sex and race from 2014-2018 in the United States. Since Figure 1 indicated that the black and white races were the most prevalent races in the United States in terms of lung cancer incident

and mortality rate, those were focused in Figure 2 with the addition of gender data. In figure 2 data from the CDC was exported into excel and all possible combinations were mixed to examine the statistics. It is shown that black males have the highest incident and mortality rate essentially of any sex/race combination in the United States. Following are white males, white females, and black females, respectively.

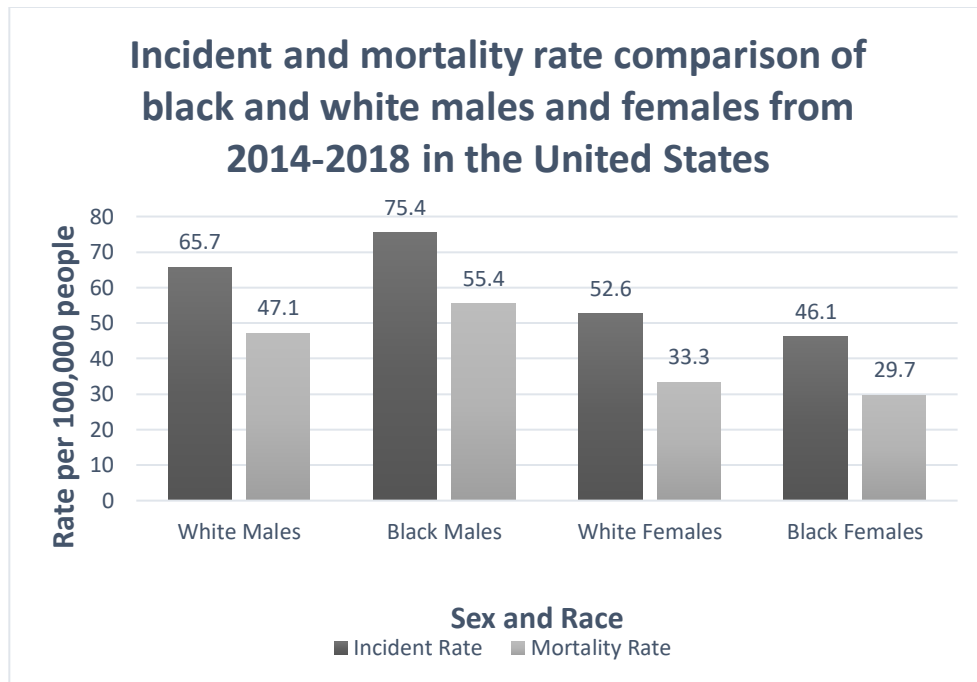


Figure 2: Incident and mortality rate of lung cancer in the U.S. by race and sex from 2014-2018. The black color shows incident rate and the gray color shows mortality rate. Black males have the highest incident and mortality rate compared to all other races in the U.S. White males follow black males in both incident and mortality rate. White males have the third highest incident and mortality rate while black females have the lowest rates. The data used to make this graph comes from the United States Center for Disease and Control. *U.S. Cancer Statistics Working Group. U.S. Cancer Statistics Data Visualizations Tool, based on 2020 submission data (1999-2018): U.S. Department of Health and Human Services, Centers for Disease Control and Prevention and National Cancer Institute; www.cdc.gov/cancer/dataviz, released in June 2021.*

Global Prevalence of Lung Cancer

Reported by the WHO Global Cancer Observatory (GLOBOCAN) in 2018, the most prevalent cancer of men worldwide in terms of *incidence* are cancers of the lung, trachea, and bronchi at 1.369 million, compared to 1.267 million cases of prostate cancer (Mattiuzzi and Lippi 2019). This is contrasted with the United States, where prostate cancer is more commonly diagnosed in men compared to lung cancer. For women, the most abundantly diagnosed cancer worldwide was breast cancer at 2.088 million cases, compared to cancers of the lung at 0.725 million cases (Mattiuzzi and Lippi 2019). Including all genders, there were a total of 2.094 million cases of cancers of the lung and 2.088 million cases of breast cancer (excluding cases of breast cancer in men) (Mattiuzzi and Lippi 2019).

Per the World Health Organization (WHO), in 2016 the most abundant source of mortality in the Western Pacific, Africa, and America's were cancer-related diagnoses (Mattiuzzi and Lippi 2019). In South-East Asia, Europe, and the Eastern Mediterranean the primary cause of mortality was ischemic heart disease (IHD) (Mattiuzzi and Lippi 2019). The leading cancer-related mortality worldwide are cancers of the lung, trachea and bronchi totaling 1,761 million mortalities in 2018 per GLOBOCAN (Mattiuzzi and Lippi 2019).

While the above data ranges from 2000-2018, new data suggests that the burden of lung cancer is not decreasing. Throughout the year of 2020, there was an astonishing 2,206,771 new cases of lung cancer worldwide, only second to breast cancer (2,261,419)

(Sung et al., 2021). The number of new deaths caused by lung cancer globally was reported an approximate value of 1,796,144 mortalities, second to none (Sung et al., 2021).

Lung Cancer Etiology

Smoking tobacco is the leading cause of lung cancer globally. Throughout the 1800s to the 1900's, the use of tobacco has substantially risen, inferentially due to the creation of cigarette wrapping machinery and the policy that troops in war were given cost-free cigarettes as a bonding technique. Over time, there was an increase in production of safety matches, electric lighters and more options for cigarette purchasing. All these factors positively influenced the smoking of tobacco. In the 1960's, there were an estimated 4400 cigarettes smoked per person over the course of a year (Warner et al., 2010). In 1964, a keystone report was published by the surgeon general informing the public on the health effects of smoking tobacco (U.S. Public Health Service 1972). In this report, it was confirmed that there was a strong causal relationship between tobacco smoking and the presence of lung cancer. It also confirmed that the chance of getting lung cancer rose with the duration of smoking history and the number of cigarettes smoked each day. The average mid-1900's smoker had 10x the likelihood of attaining lung cancer in comparison to a non-smoker, while heavy smokers were approximately 20x more likely. In conjunction with those negative affiliations, cigarette smoking was the primary cause of chronic bronchitis. Since 1964 when this publication came to life, yearly consumption of cigarettes has decreased (Warner et al., 2010).

While lung cancer is primarily a result of tobacco smoking - lifestyle characteristics, lung infections, lung health history including chronic lung diseases and

environmental and occupational hazards also can play a role in the causation of lung cancer (Bade and Dela Cruz 2020). The avoidance of tobacco smoking is the number one prevention method to lower the chances of getting lung cancer, while other measures include healthy diet consumption and living a physically active life.

LDCT Screening Efficacy

Since the National Lung Cancer Screening Trial (NLCST) in 2011, low-dose computed tomography (LDCT) has become the best method for screening of lung cancer. Lung screening is a vital prevention method for lung cancer as offers the opportunity to catch cancer at early stages which increases likelihood of survival, which is currently very low. Since the 2011 trial, LDCT screening has led to a 20% decrease in lung-cancer mortality (NLCST). That being said, LDCT annual screening has proved much more efficient at reducing mortality compared to conventional chest radiography. 57% of lung cancer diagnosis are made at an advanced stage (DeSantis et al., 2014). The primary cause for this is that early-stage cancer is generally asymptomatic. This leads to a problem as the survival rates are extremely low once lung cancers have reached an advanced stage. DeSantis estimates the 5-year survival rate of those with advanced lung cancers to be just 4% (DeSantis et al., 2014). Screening scans are then classified based on the Lung-RADS scoring system (ACR) which is further explained later in the introduction. Though there have been significant advances in screening for lung cancer, the cancer detection rate remains low (1.1% in NLST data), therefor enhancing the necessity of current guideline refinement (Teles et al., 2020). Established on the recommendations of the United States Preventive Service Task Force, lung cancer screening through LDCT scans is now offered

to asymptomatic individuals, aged 50-80 years, who have smoked at least 20 pack-years and currently smoke or have quit smoking within 15 years from screening (WV Department of Health and History 2018). Using this criterion, only an estimated 50% of those who will develop lung cancer will be deemed eligible for lung cancer screening (Wilson and Torres 2020). There are shown to be many issues in the form of excessive false-positive results, follow up studies, radiation exposure and increasing health care costs.

Lung Nodules

Per the Fleischner Society, a lung nodule is defined as a rounded opacity measuring under 3cm in diameter (Hansell et al., 2008). Lung nodules are graded on characteristics such as solid vs ground glass density, size, border structure, internal texture and the presence or absence of calcifications. Clinically, there are usually three categories of morphology of observed lung nodules. Category 1 nodules are solid and marked by homogenous soft-tissue attenuation (Loverdos et al., 2014). Category 2 nodules are ground-glass versus solid and not uniform in appearance (Loverdos et al., 2014). Category 3 nodules are part-solid and are made up of both solid and ground-glass attenuation parts (Loverdos et al., 2014).

Nodule characteristics that tend to be the most worrisome are size above 8mm in diameter with irregular shaped borders and a solid or part-solid (part-solid, part-ground glass) internal texture (Loverdos et al., 2014). Nodule size is associated with malignancy likelihood as nodules smaller than 6mm in diameter are considered low-risk nodules and above 8mm higher-risk (Loverdos et al., 2014). Another hallmark trait of a malignant lung

nodule is relative growth in size. To see this growth, there must be an initial scan to compare the most recent scan to for comparison.

There are currently multiple scientific organizations that have published recommendations on the management of lung nodules based on appearance and histology (Loverdos et al., 2014). On the other hand, these currently proposed recommendations are all rather vague, lack strong proof and are only followed by an estimated 40% of clinicians (Loverdos et al., 2014).

Lung-RADS Scoring System

In previous work completed by our team, we developed a clinic-demographic risk model for predicting the risk of lung cancer in those undergoing screening. The current work exemplified in this paper is an extension of that research. The study population used in this paper is different (RADS 2 and 3) as we are analyzing variables which could play a role in the risk progression of pulmonary nodules. It is not known whether 12-month follow-up and 6-month follow-up for RADS 2 and 3, respectively, is appropriate for all patients. It is known that some of these patients will have nodules that progress to RADS 4. If there were more specific guidelines and characteristics for increasing accuracy of classification in these intermediately-categorized nodules, it would be known which patients needed to be followed more closely than the 6 or 12 month periods. This is the aim of this project. In addition to the clinical factors, a radiomic analysis is included in this paper, which was not included in our teams last paper as their sample included patients who were already screened, and ours contained patients who were being screened.

The Lung CT Screening Reporting and Data System (Lung-RADS) is a 4-point classification system used to diagnose lung cancer nodules through the analysis of CT scans developed by the American College of Radiology (ACR). This system stratifies patients into lung cancer risk groups and provides associated management recommendations. The current protocol/guidelines are shown below.



Lung-RADS™ Version 1.1

Assessment Categories Release date: 2019

Category Descriptor	Lung-RADS Score	Findings	Management	Risk of Malignancy	Est. Population Prevalence
Incomplete	0	Prior chest CT examination(s) being located for comparison Part or all of lungs cannot be evaluated	Additional lung cancer screening CT images and/or comparison to prior chest CT examinations is needed	n/a	1%
Negative No nodules and definitely benign nodules	1	No lung nodules Nodule(s) with specific calcifications: complete, central, popcorn, concentric rings and fat containing nodules	Continue annual screening with LDCT in 12 months	< 1%	90%
Benign Appearance or Behavior Nodules with a very low likelihood of becoming a clinically active cancer due to size or lack of growth	2	Solid nodule(s): < 6 mm new < 4 mm			
		Part solid nodule(s): < 6 mm total diameter on baseline screening Non solid nodule(s) (GGN): < 30 mm OR ≥ 30 mm and unchanged or slowly growing Category 3 or 4 nodules unchanged for ≥ 3 months			
Probably Benign Probably benign finding(s) - short term follow up suggested; includes nodules with a low likelihood of becoming a clinically active cancer	3	Solid nodule(s): ≥ 6 to < 8 mm at baseline OR new 4 mm to < 6 mm Part solid nodule(s) ≥ 6 mm total diameter with solid component < 6 mm OR new < 6 mm total diameter Non solid nodule(s) (GGN) ≥ 30 mm on baseline CT or new	6 month LDCT	1-2%	5%
Probably Suspicious Findings for which additional diagnostic testing is recommended	4A	Solid nodule(s): ≥ 8 to < 15 mm at baseline OR growing < 8 mm OR new 6 to < 8 mm Part solid nodule(s): ≥ 6 mm with solid component ≥ 6 mm to < 8 mm OR with a new or growing < 4 mm solid component Endobronchial nodule	3 month LDCT; PET/CT may be used when there is a ≥ 8 mm solid component	5-15%	2%
Suspicious Findings for which additional diagnostic testing and/or tissue sampling is recommended	4B	Solid nodule(s) ≥ 15 mm OR new or growing, and ≥ 8 mm Part solid nodule(s) with: a solid component ≥ 8 mm OR a new or growing ≥ 4 mm solid component	Chest CT with or without contrast, PET/CT and/or tissue sampling depending on the "probability of malignancy and comorbidities. PET/CT may be used when there is a ≥ 8 mm solid component. For new large nodules that develop on an annual repeat screening CT, a 1 month LDCT may be recommended to address potentially infectious or inflammatory conditions	> 15%	2%
	4X	Category 3 or 4 nodules with additional features or imaging findings that increases the suspicion of malignancy			
Other Clinically Significant or Potentially Clinically Significant Findings (non lung cancer)	S	Modifier - may add on to category 0-4 coding	As appropriate to the specific finding	n/a	10%
Volumetric measurements		1.5 mm = 1.8 mm ³ 4 mm = 33.5 mm ³ 6 mm = 113.1 mm ³ 8 mm = 268.1 mm ³	10 mm = 523.6 mm ³ 15 mm = 1767.1 mm ³ 20 mm = 4188.8 mm ³ 30 mm = 14137.2 mm ³		

Figure 3: Lung RADS Version 1.1 Assessment Categories. This figure shows the assessment categories utilized in diagnosing pulmonary nodules. Adapted from: <https://www.acr.org/-/media/ACR/Files/RADS/Lung-RADS/LungRADSAssessmentCategoriesv1-1.pdf>

Lung-RADS 1 and 4 are the opposite ends of the diagnosis spectrum. Lung-RADS 1 is the category used for nodules that are definitely benign based on specific patterns of calcification or when there are no nodules present. In a study completed by Pinsky et al., 2016, a total of 26,722 participants were screened and noted as baseline screens. In this study, the prevalence rate of Lung-RADS 1 was 55.6% and the malignancy rate was less than 1% (Pinsky et al., 2015). Lung-RADS 4 is the category used when highly suspicious nodules are present. This is further investigated via a chest CT, PET/CT or biopsy. The prevalence rate is 6.2% and the likelihood of malignancy for a Lung-RADS 4 score is >5% (Pinsky et al., 2015).

Lung-RADS 2 and 3 are scores considered to be intermediate in terms of risk of malignancy. The nodules are probably benign; however, they still have a low risk of being malignant. A nodule with a relatively benign appearance and behavior is scored as Lung-RADS 2 and calls for a follow-up LDCT in 12 months. Per the Pinsky study, the prevalence rate of Lung-RADS 2 is 30.8% and the malignancy rate is 1% (Pinsky et al., 2015). A nodule that is thought to have a relatively low-risk of malignancy warranting closer observation is scored Lung-RADS 3 and the follow-up protocol consists of LDCT in 6 months. The prevalence rate is 6.4% and the likelihood of malignancy is 1-2% (Pinsky et al., 2015).

One of the questions we are asking is how we can better tailor our surveillance strategy for low to intermediate risk nodules. There is not a clear differentiation in the scoring protocol between these two scores currently. We decided to target the rads 2 and 3 population to better define the up to 2% of cases in these groups who actually have lung

cancer, in order to treat them as quickly as possible and in return, significantly improve their outcomes. They represent a large percentage (~75%) of our study sample. In this project, we are analyzing both clinical and radiologic risk factors associated with follow-up RADS score behavior, progression to RADS 4, and progression to lung cancer in a low-intermediate lung cancer risk patient population.

In the endeavor to further specify the current guidelines, current research studies are investigating the inclusion of additional risk factors into validated lung cancer screening prediction models. The goal of expanding the screening selection criteria to better identify the population at risk yielding prediction models with better sensitivity than the NLST findings. In addition, other studies successfully published similar models with the goal of predicting lung nodule malignancy risk in previously screened-detected pulmonary nodules with the goal of further refining the present clinical guidelines.

METHODS

Clinical Data acquired through Retrospective Chart Review

The first goal of this research was to perform a retrospective chart review from patients at Boston Medical Center (BMC) to look at demographics and pulmonary health history.

The population was selected by identifying first time screened patients at BMC between March 1st. 2015, and December 31st. 2019 (n=2847). The inclusion criteria were comprised of patients over the age of 55 with a 20+ pack-year smoking history. From this population, the research sample was narrowed down to patients with low-intermediate risk (RADS 2-3) (n=2174, 76.7%) scores. From there, only patients with a follow up LDCT screen based on Lung-RADS guidelines were selected (n=908, 38.8 %).

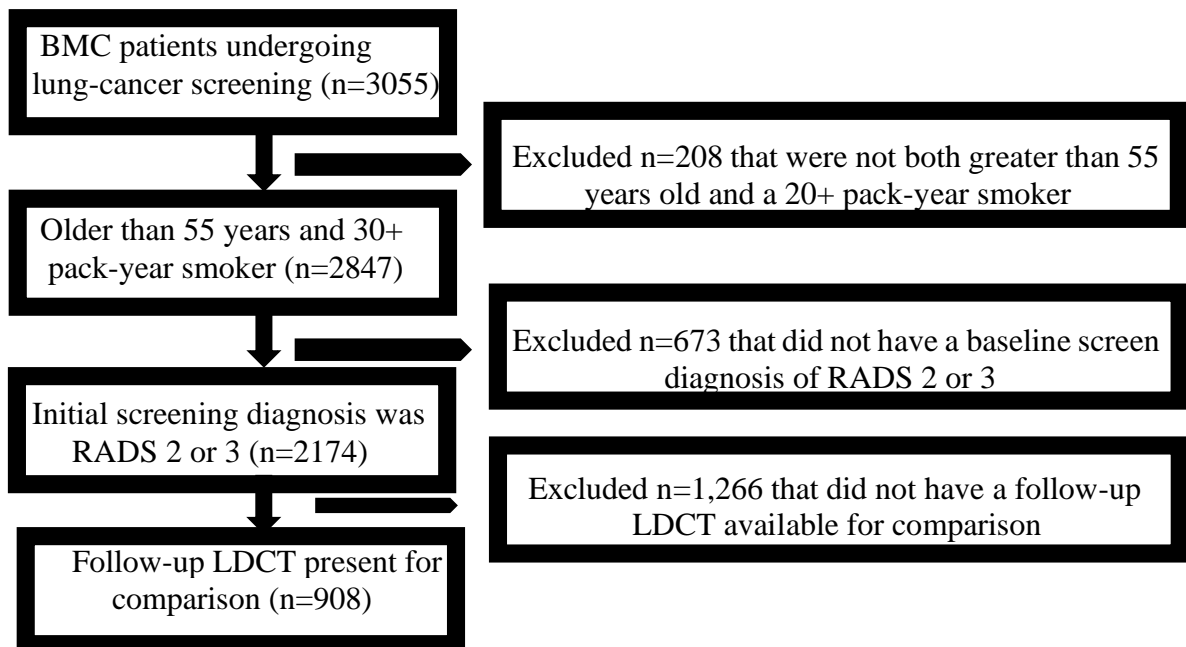


Figure 4: Consort diagram depicting the sample selection process for the retrospective chart review.

The personal history portion included age, gender, race, education, body-mass index, and pack-years, smoking history while the pulmonary health history included all reports of pneumonia, emphysema (if emphysema was present, including severity), COPD and the presence/absence of family history of lung cancer.

The first goal was to determine a significant difference between the compiled factors and the following end points through a chi-square analysis: Lung-RADS follow up score (improved, worsened, unchanged), progression to Lung-RADS 4 and progression of lung cancer. The second goal was to perform a logistic regression analysis to determine the strength of the significant associations that we found and to evaluate predictors of progression to Lung-RADS 4 and to Lung Cancer (tables 7 and 8).

Radiomic Analysis

The same n=908 population from the clinical chart review was narrowed down to 132 cases which mirrored the majority of clinical characteristics and variables of the general population. Regarding our goal in this research of trying to get the incidence of mortality in these groups even lower, the radiomic analysis was a technique utilized to search for new features that would better stratify benign from malignant nodules.

We selected a subset of 132 cases with available 1.25 mm LDCT studies and downloaded raw DICOM images from our institution's Picture Archiving and Communication System (PACS). For each scan, trained reviewers under the guidance of an attending thoracic surgeon annotated the locations of suspicious nodules as originally reported within radiology reports. Only the largest cross-sectional slice of each nodule was selected for annotation. We thereafter extracted 86 unique radiomic features from each

annotated nodule using the PyRadiomics package in Python 2.7.1 (Van Griethuysen, Joost JM, et al. "Computational radiomics system to decode the radiographic phenotype." *Cancer research* 77.21 (2017): e104-e107.).

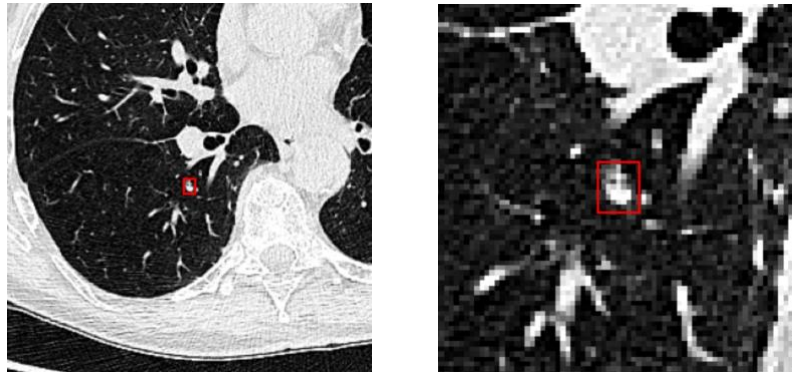


Figure 5: An annotated image of a pulmonary nodule used in our research.

Radiomic variables were then used to train Lasso regression, random forest, and K-nearest neighbors (KNN) machine learning algorithms with Synthetic Minority Oversampling Technique (SMOTE) to correct for imbalances in observed outcomes. Each algorithm was trained in a supervised fashion for three separate outcomes: (i) lung cancer diagnosis (ii) overall worsening in Lung-RADS progression and (iii) progression to RADS 4. The performance of each algorithm was assessed using the area under receiver operating characteristic curve (AUC).

Statistical Analysis

For the clinical retrospective chart review, we used SPSS statistics and ran multiple variables against worsened, unchanged, and improved population samples. A logistic regression analysis was also utilized. Significance was deemed to be $p=0.05$ or lower for

all statistics. For the radiomics, the PyRadiomics package was used in Python to analyze the nodules under multiple algorithms. Significant values are highlighted in the results section.

RESULTS

Table 1. (below) is a demographics table exemplifying the characteristics of RADS 2 and 3 lung cancer patients by follow-up CT results. There were no significant difference between groups in this table.

Table 1. Characteristics of RADS 2 & 3 lung cancer patients by follow-up CT results.

Variable	Improved (N=168)	Unchanged (N=656)	Worsened (N=84)	p-value
Age (years)	63.0 ± 5.9	Mean ± SD 63.5 ± 5.6	63.6 ± 6.0	0.621
Gender	N (column percent)			0.231
Male	108 (20.0%)	386 (71.6%)	45 (8.4%)	
Female	60 (16.3%)	270 (73.2%)	39 (10.6%)	
Race				0.651
White	72 (19.2%)	265 (70.5%)	39 (10.4%)	
Black	60 (18.2%)	236 (71.5%)	34 (10.3%)	
Other	6 (11.5%)	42 (80.8%)	4 (7.7%)	
Education				0.792
Did not attend school	9 (16.7%)	42 (77.8%)	3 (5.6%)	
Some high school or less	52 (17.6%)	215 (72.9%)	28 (9.5%)	
Graduated high school	72 (18.3%)	279 (71.0%)	42 (10.7%)	
Graduated college	19 (21.6%)	63 (71.6%)	6 (6.8%)	
Body-Mass Index				0.382
Underweight	8 (15.7%)	40 (78.4%)	3 (5.9%)	
Normal	44 (18.1%)	168 (69.1%)	31 (12.8%)	
Overweight	54 (17.8%)	223 (73.4%)	27 (8.9%)	
Obese	62 (20.0%)	225 (72.6%)	23 (7.4%)	
Pack-Years				0.536
>0-30	46 (17.8%)	192 (74.1%)	21 (8.1%)	
>30-60	109 (19.5%)	400 (71.6%)	50 (8.9%)	
>60-90	8 (12.5%)	47 (73.4%)	9 (14.1%)	
>90	5 (19.2%)	17 (65.4%)	4 (15.4%)	
Family History of Lung Cancer				0.320
Yes	22 (19.0%)	79 (68.1%)	15 (12.9%)	
No	146 (18.4%)	577 (72.9%)	69 (8.7%)	

Table 2. Characteristics of RADS 2 & 3 lung cancer patients by progression to RADS-4 on follow-up CT.

Variable	Progression to RADS-4 (N=39)	No Progression (N=869)	p-value
Age (years)	Mean ± SD		
	64.2 ± 5.8	63.4 ± 5.7	0.349
Gender	N (column percent)		
Male	17 (3.2%)	522 (96.9%)	0.040
Female	22 (6.0%)	347 (94.0%)	
Race			0.802
White	16 (4.3%)	360 (95.7%)	
Black	17 (5.2%)	313 (94.9%)	
Other	3 (5.8%)	49 (94.2%)	
Education			0.781
Did not attend school	2 (3.7%)	52 (96.3%)	
Some high school or less	16 (5.4%)	279 (94.6%)	
Graduated high school	16 (4.1%)	377 (95.9%)	
Graduated college	3 (3.4%)	85 (96.6%)	
Body-Mass Index			0.856
Underweight	2 (3.9%)	49 (96.1%)	
Normal	11 (4.5%)	232 (95.5%)	
Overweight	15 (4.9%)	289 (95.1%)	
Obese	11 (3.6%)	299 (96.5%)	
Pack-Years			0.589
>0-30	10 (3.9%)	249 (96.1%)	
>30-60	25 (4.5%)	534 (95.5%)	
>60-90	4 (6.3%)	60 (93.8%)	
>90	0 (0.0%)	26 (100.0%)	
Family History of Lung Cancer			0.014
Yes	10 (8.6%)	106 (91.4)	
No	29 (3.7%)	763 (96.4%)	

Table 2. (above) also shows demographical factors but differs from Table 1. because it is specifically comparing those who did not progress to RADS-4 to those who did. There was a significantly higher percentage (p=0.04) of females who progressed to RADS-4 (6.0%) as compared to males (3.2%). This table also shows the significance (p=0.014) of a family history of lung cancer. 8.6% of the cases with a family history of lung cancer progressed to RADS-4 and only 3.7% of the cases without that family history progressed to RADS-4.

Table 3. (below) illustrates the demographical factors in comparison with the presence or absence of a lung cancer diagnosis. There were no significant differences determined by p-value; however, there were very strong results in the age characteristics and might be an interesting factor to explore further in the future with a larger sample.

Table 3. Characteristics of RADS 2 & 3 lung cancer patients by lung cancer diagnosis.

Variable	Lung Cancer Diagnosis (N=23)	No Diagnosis (N=2151)	p-value
Age (years)	Mean ± SD		
	63.0 ± 5.9	65.4 ± 7.1	0.060
Gender	N (column percent)		
Male	12 (0.9%)	1320 (99.1%)	0.368
Female	11 (1.3%)	831 (98.7%)	
Race			0.067
White	7 (0.8%)	883 (99.2%)	
Black	9 (1.1%)	793 (98.9%)	
Other	4 (3.1%)	127 (97.0%)	
Education			0.282
Did not attend school	3 (2.5%)	115 (97.5%)	
Some high school or less	5 (0.8%)	653 (99.2%)	
Graduated high school	11 (1.1%)	959 (98.9%)	
Graduated college	1 (0.5%)	203 (99.5%)	
Body-Mass Index			0.794
Underweight	1 (0.6%)	163 (99.4%)	
Normal	7 (1.2%)	579 (98.8%)	
Overweight	9 (1.3%)	702 (98.7%)	
Obese	6 (0.8%)	707 (99.2%)	
Pack-Years			0.445
>0-30	5 (0.7%)	671 (99.3%)	
>30-60	14 (1.1%)	1293 (98.9%)	
>60-90	3 (2.2%)	133 (97.8%)	
>90	1 (1.8%)	54 (98.2%)	
Family History of Lung Cancer			0.092
Yes	5 (2.1%)	231 (97.9%)	
No	18 (0.9%)	1920 (99.1%)	

Table 4. Baseline co-morbidities of RADS 2 & 3 lung cancer patients by follow-up CT results.

Variable	Improved (N=168)	Unchanged (N=656)	Worsened (N=84)	p-value
Pneumonia	N (column percent)			0.028
Yes	47 (24.4%)	125 (64.8%)	21 (10.9%)	
No	121 (16.9%)	531 (74.3%)	63 (8.8%)	
Emphysema				0.002
Yes	127 (20.3%)	432 (68.9%)	68 (10.9%)	
No	41 (14.6%)	224 (79.7%)	16 (5.7%)	
Emphysema Grade				<0.001
None	41 (14.6%)	224 (79.7%)	16 (5.7%)	
Mild	112 (21.5%)	362 (69.4%)	48 (9.2%)	
Moderate	12 (15.6%)	53 (68.8%)	12 (15.6%)	
Severe	3 (10.7%)	17 (60.7%)	8 (28.6%)	
COPD				0.237
Yes	85 (18.4%)	326 (73.8%)	34 (7.6%)	
No	83 (18.6%)	330 (73.8%)	34 (7.6%)	
History of Lung Cancer				0.110
Yes	28 (17.3%)	112 (69.1%)	22 (13.6%)	
No	140 (18.8%)	544 (72.9%)	62 (8.3%)	

Table 4. (above) is a pulmonary health history table showing the characteristics of Lung-RADS 2 and 3 lung cancer patients by follow-up CT results. A patient history of pneumonia was significant at $p=0.028$ as nearly 11% of this sample worsened on their follow-up CT scans. It was also discovered that a patient history of emphysema and severity of emphysemas play a vastly significant role in RADS 2 and 3 patient outlooks. The sample of patients having a history of emphysema were 5.2% more likely to have worsened results on their follow-up CT scans compared to those without a history of emphysema. Very strong results were also acquired in this graph regarding the severity of emphysema noted in the patient records. From no grade to severe emphysema there was a 22.9% increase in the worsening of follow-up CT scans.

Table 5. Baseline co-morbidities of RADS 2 & 3 lung cancer patients by progression to RADS-4 on follow-up CT.

Variable	Progression to RADS-4 (N=39)	No Progression (N=869)	p-value
	N (column percent)		
Pneumonia			0.494
Yes	10 (5.2%)	183 (94.8%)	
No	29 (4.1%)	686 (95.9%)	
Emphysema			0.073
Yes	32 (5.1%)	595 (94.9%)	
No	7 (2.5%)	274 (97.5%)	
Emphysema Grade			0.010
None	7 (2.5%)	274 (97.5%)	
Mild	22 (4.2%)	500 (95.8%)	
Moderate	6 (7.8%)	71 (92.2%)	
Severe	4 (14.3%)	24 (2.8%)	
COPD			0.472
Yes	22 (4.8%)	439 (95.2%)	
No	17 (3.8%)	430 (96.2%)	
History of Lung Cancer			0.383
Yes	9 (5.6%)	153 (94.4%)	
No	30 (4.0%)	716 (96.0%)	

Table 5. (above) illustrates the baseline co-morbidities of RADS 2 and 3 lung cancer patients by progression to RADS-4 on follow-up CT. Out of the sample that had emphysema, there was a steady increase in percentage from those with no grade (2.5%) to those with severe emphysema (14.3%) creating a significance value of $p=0.010$. For this table, just the presence of emphysema while not significant, still showed promising results.

Table 6. (below) had several significant findings when examining the baseline co-morbidities of RADS 2 and 3 lung cancer patients by lung cancer diagnosis. The presence of emphysema was significant at $p=0.002$ while 1.6% of those with emphysema were diagnosed with lung cancer opposed to 0.1% of those without emphysema. Furthermore,

the grade of emphysema was shown to be significant yet again with a value of $p=0.002$. Out of the sample that had emphysema, there was an increase in percentage from those with no grade (0.1%) to those with severe emphysema (3.9%) creating a significance value of $p=0.002$. The only significant value present in the relationship between COPD and lung cancer likelihood outlook came from this table. It was observed that of those with a history with COPD, 1.7% progressed to a lung cancer diagnosis compared to 0.5% of those without a history of COPD.

Table 6. Baseline co-morbidities of RADS 2 & 3 lung cancer patients by lung cancer diagnosis.

Variable	Lung Cancer Diagnosis (N=23)	No Diagnosis (N=2151)	p-value
	N (column percent)		
Pneumonia			0.140
Yes	2 (0.4%)	459 (99.6%)	
No	21 (1.2%)	1692 (98.8%)	
Emphysema			0.002
Yes	22 (1.6%)	1395 (98.5%)	
No	1 (0.1%)	756 (99.9%)	
Emphysema Grade			0.002
None	1 (0.1%)	756 (99.9%)	
Mild	18 (1.5%)	1180 (98.5%)	
Moderate	1 (0.7%)	141 (99.3%)	
Severe	3 (3.9%)	74 (96.1%)	
COPD			0.006
Yes	17 (1.7%)	976 (98.3%)	
No	6 (0.5%)	1175 (99.5%)	
History of Lung Cancer			0.681
Yes	4 (1.3%)	309 (98.7%)	
No	19 (1.0%)	1842 (99.0%)	

Tables 7. and 8. (below) demonstrate the results of a logistic regression analysis which hoped to determine the strength of the significant associations that were found and to evaluate predictors to progression to Lung-RADS 4 and to lung cancer. Both analyses included a 95% confidence interval. Table 7. is a multivariate analysis looking at the effect of characteristics of RADS 2 and 3 lung cancer patients on progression to RADS-4 on follow-up CT. The sample for this analysis was n=908, as that was the number subjects we had with a follow-up LDCT screening for comparison. Table 7. illustrated significant results of both a family history of lung cancer and severe emphysema, at $p=0.050$ and $p=0.024$ respectively. Table 8. is a multivariate analysis looking at the effect of characteristics of RADS 2 and 3 lung cancer patients on lung cancer diagnosis. The sample for this analysis was n=2174, as that was the number subjects we had that had an initial diagnosis of RADS 2 or 3. COPD, mild emphysema, moderate emphysema and severe emphysema v. none were analyzed in this table. The presence of severe emphysema was shown to have an impact on lung cancer diagnosis with a significance of $p=0.021$.

Table 7. Multivariate analysis - the effect of characteristics of RADS 2 & 3 lung cancer patients on progression to RADS-4 on follow-up CT (n=908).

	Progression to RADS-4	
	Odds Ratio (95% CI)	p-value
Male (vs. Female)	0.6 (0.3-1.1)	0.094
Family History of Lung Cancer	2.2 (1.0-4.6)	0.050
Mild Emphysema (vs. None)	1.8 (0.8-4.3)	0.249
Moderate Emphysema (vs. None)	3.0 (1.0-9.4)	0.570
Severe Emphysema (vs. None)	6.6 (1.8-24.6)	0.024

Table 8. Multivariate analysis - the effect of characteristics of RADS 2 & 3 lung cancer patients on lung cancer diagnosis (n=2174).

	Progression to RADS-4	
	Odds Ratio (95% CI)	p-value
COPD	2.3 (0.9-5.9)	0.095
Mild Emphysema (vs. None)	9.1 (1.2-69.7)	0.177
Moderate Emphysema (vs. None)	3.9 (0.2-64.5)	0.737
Severe Emphysema (vs. None)	19.7 (1.9-200.7)	0.021

Radiomic Analysis

Radiomic features were generally poor predictors of future changes in RADS classification. While KNN predicted progression of lung nodules to RADS 4 with an AUC of 0.69, performance of all models was 0.50 or less in the prediction of overall RADS progression. Conversely, radiomics were a relatively improved predictor of an ultimate lung cancer diagnosis, with Lasso regression achieving an AUC of 0.70.

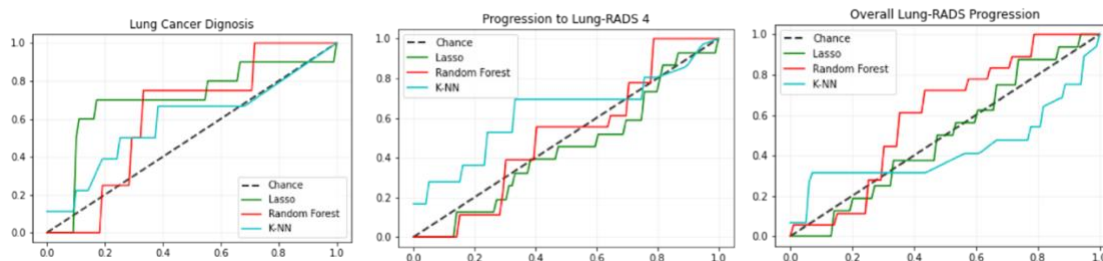


Figure 6: Radiomic Analysis Graphs (above) and Table 9: Radiomic Analysis Quantitative Values (below). In figure 6., the green line indicates lasso regression, the red line indicates random forest, and the blue line indicates K-Nearest Neighbors (KNN). Consider the X-axis as component 1 and the Y axis as Component 2 for the unsupervised clustering methods. The graph to the left compares chance to these three algorithms for lung cancer diagnosis. The graph in the middle compares chance to the variables regarding progression to Lung-RADS 4. The graph to the right compares chance to the variables regarding overall Lung-RADS progression. The quantitative values of the graphs are shown below in **Table 9**.

Clustering Method	Lung Cancer Diagnosis	Progression to RADS 4	Overall RADS Progression
Lasso Regression	0.7	0.47	0.5
Random Forest	0.62	0.44	0.49
K-Nearest Neighbors	0.59	0.69	0.43

Table 9. (above) AUC of Machine Learning Algorithms Trained with Radiomic Features

DISCUSSION

All in all, this research led to significant findings throughout the retrospective chart review portion of the project, while the radiomic analysis simply led to good predictive value. After calculation of AUC values which calculate area under the ROC curve, it was determined that numbers above 0.5 indicate good predictive value for the radiomic analysis.

Referencing table 2, it was expected that males would be slightly more vulnerable to the acquisition of lung cancer; however, our data suggests that females are slightly more vulnerable to progression of RADS 4 on follow-up LDCT screening. Clinically, it was

observed that those with a family history of lung cancer had a significantly higher likelihood of progressing to RADS 4 from RADS 2/3 ($p=0.014$). The presence of emphysema in patient history was significant in increasing the likelihood of developing lung cancer, while those with a history of severe emphysema were both significantly more likely to progress to RADS 4 from RADS 2/3 and therefor develop lung cancer.

Further exploration should be guided towards enhanced analysis of patients with a history of emphysema (specifically moderate to severe emphysema) and analyze whether a similar significant value is present throughout a larger study sample. Perhaps females should also be targeted as they were found to be significantly more likely to progress to RADS 4. Utilizing these pulmonary health history factors would likely help gather the most appropriate collection of high-risk patients to get earlier screening. If these high-risk patients that meet the criteria analyzed in this project can get LDCT screening done at an earlier age, there will likely be an increase at catching lung cancer at an earlier stage thus decreasing the mortality rate overall.

Referencing the current LDCT screening requirements in place (55+y/o, 20+pack-year smoker), it seems they are too limited. Based on our evidence in this research, LDCT screening requirements should be specified to include exceptions to these requirements for those with a patient history of emphysema, pneumonia and/or COPD. If a patient is a 20+ pack-year smoker, male, has a family history of lung cancer, has a history of pneumonia, COPD or emphysema, and is initially diagnosed with RADS 2/3, this is for example, a patient who should be classified high-risk for progression from RADS 2/3 to RADS 4 and

thus be able to receive LDCT screening earlier than those who do not present with these histories.

REFERENCES

- Brett C. Bade, Charles S. Dela Cruz, Lung Cancer (2020): Epidemiology, Etiology, and Prevention, *Clinics in Chest Medicine*, Volume 41, Issue 1, Pages 1-24, ISSN 0272-5231, ISBN 9780323791403, <https://doi.org/10.1016/j.ccm.2019.10.001>. (<https://www.sciencedirect.com/science/article/pii/S0272523119300802>)
- Dela Cruz, C. S., Tanoue, L. T., & Matthay, R. A. (2011). Lung cancer: epidemiology, etiology, and prevention. *Clinics in Chest Medicine*, 32(4), 605–644. <https://doi.org/10.1016/j.ccm.2011.09.001>
- DeSantis CE, Lin CC, Mariotto AB, Siegel RL, Stein KD, Kramer JL, Alteri R, Robbins AS, Jemal A. (2014) Cancer treatment and survivorship statistics. *CA: A Cancer Journal for Clinicians*. Jul-Aug;64(4):252-71. doi: 10.3322/caac.21235. Epub 2014 Jun 1. PMID: 24890451.
- Chudgar, N. P., Bucciarelli, P. R., Jeffries, E. M., Rizk, N. P., Park, B. J., Adusumilli, P. S., & Jones, D. R. (2015). Results of the national lung cancer screening trial: where are we now?. *Thoracic Surgery Clinics*, 25(2), 145–153. <https://doi.org/10.1016/j.thorsurg.2014.11.002>
- Chung K, Jacobs C, Scholten ET, Goo JM, Prosch H, Sverzellati N, Ciompi F, Mets OM, Gerke PK, Prokop M, van Ginneken B, Schaefer-Prokop CM. (2017) Lung-RADS Category 4X: Does It Improve Prediction of Malignancy in Subsolid Nodules? *Radiology*. Jul;284(1):264-271. doi: 10.1148/radiol.2017161624. Epub Mar 24. PMID: 28339311.

- Hansell DM, Bankier AA, MacMahon H, McLoud TC, Müller NL, Remy J. Fleischner Society: (2008) Glossary of terms for thoracic imaging. *Radiology*. Mar;246(3):697-722. doi: 10.1148/radiol.2462070712. Epub 2008 Jan 14. PMID: 18195376.
- Loverdos, K., Fotiadis, A., Kontogianni, C., Iliopoulou, M., & Gaga, M. (2019). Lung nodules: A comprehensive review on current approach and management. *Annals of Thoracic Medicine*, 14(4), 226–238. https://doi.org/10.4103/atm.ATM_110_19
- Mattiuzzi, C., & Lippi, G. (2019) Current Cancer Epidemiology. *Journal of Epidemiology and Global Health*, 9(4), 217–222. <https://doi.org/10.2991/jegh.k.191008.001>
- Pedregosa, Fabian, et al. (2011) "Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825-2830
- Pinsky PF, Berg CD. (2012) Applying the National Lung Screening Trial eligibility criteria to the US population: what percent of the population and of incident lung cancers would be covered? *Journal of Medical Screening*. Sep;19(3):154-6. doi: 10.1258/jms.2012.012010. Epub 2012 Oct 11. PMID: 23060474.
- Pinsky PF, Gierada DS, Black W, et al. (2015) Performance of Lung-RADS in the National Lung Screening Trial: A Retrospective Assessment. *Annals of Internal Medicine* ;162(7):485-491. doi:10.7326/M14-2086
- Sanchez-Salcedo P, Wilson DO, de-Torres JP, et al. (2015) Improving selection criteria for lung cancer screening. The potential role of emphysema. *American Journal of Respiratory and Critical Care Medicine*.191(8):924-931. doi:10.1164/rccm.201410-1848OC

Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global Cancer Statistics 2020: (2021) GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*. May;71(3):209-249. doi: 10.3322/caac.21660. Epub 33538338. PMID:

Teles, G., Macedo, A., Chate, R. C., Valente, V., Funari, M., & Szarf, G. (2020). LDCT lung cancer screening in populations at different risk for lung cancer. *BMJ Open Respiratory Research*, 7(1), e000455. <https://doi.org/10.1136/bmjresp-2019-000455>

Torre LA, Siegel RL, Jemal A. (2016) Lung Cancer Statistics. *Advances in Experimental Medicine and Biology*. 893:1-19. doi: 10.1007/978-3-319-24223-1_1. PMID: 26667336.

U.S. Cancer Statistics Working Group. U.S. Cancer Statistics Data Visualizations Tool, based on 2020 submission data (1999-2018): (2021) U.S. Department of Health and Human Services, Centers for Disease Control and Prevention and National Cancer Institute; www.cdc.gov/cancer/dataviz, released in June.

U.S. Public Health Service, office of the Surgeon General: (1972) The health consequences of smoking. *National Clearinghouse for Smoking Health*.

Van Griethuysen, Joost JM, et al. (2017) "Computational radiomics system to decode the radiographic phenotype." *Cancer Research*. e104-e107.

Warner KE, Mendez D. (2010) Tobacco control policy in developed countries: yesterday, today, and tomorrow. *Nicotine and Tobacco Research*. Sep;12(9):876-87. doi: 10.1093/ntr/ntq125. Epub 2010 Aug 11. PMID: 20702814.

West Virginia Cancer Burden Report (2018). WV Cancer Registry, West Virginia Department of Health and Human Resources and West Virginia University Cancer Institute, Morgantown, West Virginia

Yuki Tomonaga, Kevin ten Haaf, et al. (2018) “Cost-effectiveness of low-dose CT screening for lung cancer in a European country with high prevalence of smoking—A modeling study” *Lung Cancer*, Volume 121, 2018, Pages 61-69, ISSN 0169-5002, <https://doi.org/10.1016/j.lungcan.2018.05.008>.

Curriculum Vitae

