

**IDENTIFYING CT IMAGE RADIOMIC BIOMARKERS
FOR PREDICTING IMMUNOTHERAPY RESPONSE OF
NON-SMALL CELL LUNG CANCER PATIENTS**

by

İlke Tunalı

B.S., in Biomedical Engineering, Başkent University, 2008

M.S., in Computer Engineering, Ondokuz Mayıs University, 2013

Submitted to the Institute of Biomedical Engineering

in partial fulfillment of the requirements

for the degree of

Doctor

of

Philosophy

Boğaziçi University

2020

**IDENTIFYING CT IMAGE RADIOMIC BIOMARKERS
FOR PREDICTING IMMUNOTHERAPY RESPONSE OF
NON-SMALL CELL LUNG CANCER PATIENTS**

APPROVED BY:

Assoc. Prof. Dr. Albert Güveniř
(Thesis Advisor)

Prof. Dr. Robert J. Gillies
(Thesis Co-advisor)

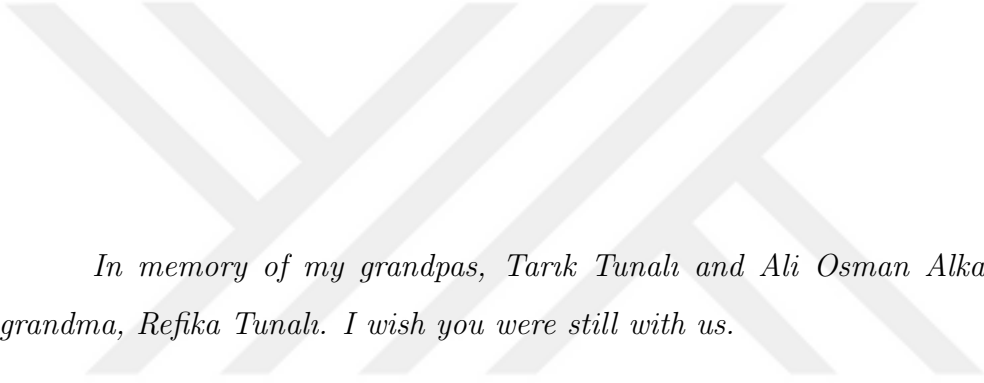
Prof. Dr. Cengizhan Öztürk

Assist. Prof. Dr. Bora Garipcan

Assoc. Prof. Dr. Matthew B. Schabath

Assist. Prof. Dr. İmran Göker

DATE OF APPROVAL: 17 December 2019



*In memory of my grandpas, Tarık Tunah and Ali Osman Alkanlar, and my
grandma, Refika Tunah. I wish you were still with us.*

ACKNOWLEDGMENTS

Completion of this doctoral dissertation was possible with the support of a group of great, intelligent and humble people. I would like to express my sincere gratitude to all of them.

First, I want to thank my advisor, Assoc. Prof. Albert Güveniş for believing in me, being open-minded, supporting me throughout the long years of my PhD study and being a guide for doing good science. Thank you.

I owe a sincere thank you to Prof. Robert J. Gillies for changing my career and giving me the opportunity to work with the one of the best group of scientists in cancer research. Your support, valuable guidance, and encouragement not only made possible of this study but also has shifted my perception on how to be a better scientist. Thank you.

I am deeply grateful to Prof. Matthew B. Schabath for his enormous support and excellent mentoring, guiding me patiently throughout my dissertation study. I truly admired your approach and passion to science and will try to use it as a model for my future studies. Thank you.

Many thanks to my committee members Prof. Dr. Cengizhan Öztürk, Assist. Prof. Bora Garipcan and Prof. Mahmut Gümüş for their time and scientific advices.

I am also grateful to all of those with whom I have had the pleasure to work with through the years of this study. I was lucky enough to be part of this group of excellent scientist at H. Lee Moffitt Cancer Center that were kind enough to extend their help at various phases of this research, whenever I approached them. As such, I thank Drs. Jhanelle E. Gray, Olya Stringfield, Mahmoud A. Abdalla, Yoganand Balagurunathan, Jin Qi, Yan Tan, Steven A. Eschrich, Daniel K. Jeong, Theresa A. Boyle, and James

J. Saller.

I would like to extend my sincere word of thanks to Rachel Snayd, Connie Schmitt and Fiorella Lanatta for all their helps with the administrative processes and creating a much more efficient working environment.

I owe a lot to my family. They encouraged and helped me at every stage of my personal and academic life, and longed to see this achievement come true. Especially my mom, Dr. Berna Tunalı, for inspiring me as role model of a hardworking scientist and being the most caregiving mother that someone can have. I love you. I would also like to express my feelings and thank to my loving and supporting wife, Nazlı, as being with there whenever I need and being generously supportive, even when going through your very own tough times. You are a gift for me. I love you. I also wish to give thanks to my dad, Ahmet Tunalı and my brother Utku for supporting me and cheering me up as always. I am gifted by your presence. Also, thank you grandma, Melahat Özkan, for taking care of me not just as a baby but also as a young college student for years. You gave me endless love and care. I would also like to thank my uncle Namık Kemal Alkanlar for being a great supporter as well as a great friend during my early years as a PhD student. Finally, I would like to thank Alberto L. Garcia for being the best friend through my years in US and easing my transition to another country and culture.

Special thanks to H. Lee Moffitt Cancer Center & Research Institute for supporting me and providing me valuable resources and environment for conducting my research. I was privileged to be part of a team that contributes to the prevention and cure of cancer.

ACADEMIC ETHICS AND INTEGRITY STATEMENT

I, İlke Tunalı, hereby certify that I am aware of the Academic Ethics and Integrity Policy issued by the Council of Higher Education (YÖK) and I fully acknowledge all the consequences due to its violation by plagiarism or any other way.

Name :

Signature:

Date:

ABSTRACT

IDENTIFYING CT IMAGE RADIOMIC BIOMARKERS FOR PREDICTING IMMUNOTHERAPY RESPONSE OF NON-SMALL CELL LUNG CANCER PATIENTS

Checkpoint blockade immunotherapy (IO) provides improved long-term survival in a subset of advanced stage non-small cell lung cancer (NSCLC) patients. However, highly predictive biomarkers of IO response are an unmet clinical need. In this thesis, pre-treatment clinical covariates and quantitative image-based features (i.e., Radiomics) were utilized to identify parsimonious models that predict rapid disease progression (RDP) phenotypes and survival outcomes among NSCLC patients treated with IO. As part of the thesis, four studies were conducted. First, novel prognostic and predictive computed tomography (CT) radiomic features utilizing radial gradient and radial deviation maps were created. One feature, RD outside-border SD, was found to be associated with overall survival in two independent NSCLC cohorts. Second, clinical-radiomic models that predicted RDP phenotypes, including hyperprogressive disease (HPD), were created in the setting of NSCLC IO. Among 228 NSCLC patients, parsimonious clinical-radiomic models with modest to high ability (area under the curves: 0.812 and 0.843) to predict RDP were identified. In the third study, stable and reproducible peritumoral and intratumoral CT radiomic features of lung lesions were identified to reduce the chance of spurious findings. In the fourth and final study, pre-treatment clinical covariates and radiomics were utilized to identify a parsimonious risk-model based on survival outcomes among 332 NSCLC patients treated with IO. The most predictive radiomic feature (GLCM inverse difference) was found to be positively associated with CAIX expression, using a gene-expression and an immunohistochemistry dataset.

Keywords: Radiomics, Lung cancer, Immunotherapy.

ÖZET

KÜÇÜK HÜCRELİ DIŐI AKCİĐER KANSERİ HASTALARININ İMMÜNÖTERAPİ CEVAPLARINI ÖNGÖREN RADYOMİK BT GÖRÜNTÜ BİYOİŐARETLERİNİN TESPİTİ

Kontrol noktası blokajı immünoterapisi (İmT) ileri evre küçük hücreli dışı akciğeri kanseri (KHDAK) hastalarının bir kısmında uzun süreli sağkalım gösteriyor. Bununla birlikte, İmT yanıtını yüksek bir başarı ile tahmin edebilen biyobelirteçler halen karşılanamamış klinik bir ihtiyaçtır. Bu tezde, İmT ile tedavi edilmiş KHDAK hastaları arasında ani hastalık progresyonu (AHP) fenotipleri ve sağkalım sonuçlarını öngören yalın modelleri tanımlamak için tedavi öncesi klinik değışkenler ve nicel görüntütemelli özellikler (yani, Radyomikler) kullanılmıştır. Bu tezin içeriğı bağlamında dört çalışma yapılmıştır. İlk olarak, radyal gradyan ve radyal sapma haritalarını kullanan yeni prognostik ve prediktif bilgisayarlı tomografi (BT) radyomik özellikleri oluşturuldu. Bir özellik, RD outside-border SD, iki bağımsız KHDAK kohortunda genel sağkalım ile ilişkili olarak bulundu. İkinci olarak, KHDAK İmT'nin hiperani progresyonu (HAD) dahil olmak üzere AHP fenotiplerini öngören klinik-radyomik modeller yaratıldı. Toplam 228 KHDAK hastasında AHP'yi öngörmek için orta ila yüksek kabiliyete sahip (eğri altındaki alanlar: 0.812 ve 0.843) klinik-radyolojik modeller oluşturuldu. Üçüncü çalışmada, akciğeri lezyonları kullanılarak stabil ve tekrarlanabilir periferik-tümör ve tümör-içi BT radyomik özellikleri, yanlış bulgu olasılığını azaltmak için tanımlandı. Dördüncü ve son çalışmada, İmT ile tedavi edilen 332 KHDAK hastasının sağkalım sonuçlarına dayanan yalın bir risk modeli tanımlamak için tedavi öncesi klinik değışkenler ve radyomikler kullanıldı. En öngörücü radyomik özellik (GLCM inverse difference), bir gen ifadesi ve bir immünohistokimya kohortu kullanılarak, CAIX ifadesi ile pozitif olarak ilişkili bulundu.

Anahtar Sözcükler: Radyomikler, Akciğeri kanseri, İmmünoterapi.

TABLE OF CONTENTS

ACKNOWLEDGMENTS	iv
ACADEMIC ETHICS AND INTEGRITY STATEMENT	vi
ABSTRACT	vii
ÖZET	viii
LIST OF FIGURES	xiii
LIST OF TABLES	xx
LIST OF SYMBOLS	xxiv
LIST OF ABBREVIATIONS	xxv
1. MOTIVATION AND OBJECTIVES	1
1.1 Thesis hypotheses, objectives and organization	2
2. LUNG CANCER AND RADIOMICS	5
2.1 Lung Cancer	5
2.2 Cancer Heterogeneity and the Utility of Radiomics	7
2.3 Early Detection Radiomic Studies	9
2.4 Prognostic and Survival Outcome Radiomic Studies	12
2.5 Treatment Response Prediction Radiomic Studies	14
2.6 Radiogenomics Studies	16
2.7 Limitations of Radiomics	18
2.8 Future Guidelines	20
2.9 Conclusions	21
3. RADIOMIC FEATURES EXTRACTED FROM RADIAL GRADIENT AND RADIAL DEVIATION MAPS ASSOCIATED WITH SURVIVAL AMONG LUNG ADENOCARCINOMA PATIENTS	22
3.1 Introduction	22
3.2 Materials and Methods	23
3.2.1 Lung Cancer Patients	23
3.2.2 Patient Data	24
3.2.3 Tumor Segmentation	24
3.2.4 Radial Gradient and Radial Deviation Maps and Features	25

3.2.5	Elimination of Non-Reproducible and Redundant Features . . .	26
3.2.6	Radiological Semantic Features	27
3.2.7	Statistical Analyses	28
3.3	Results	28
3.3.1	Patient Demographics	28
3.3.2	Univariable Analyses	29
3.3.3	Multivariable Analyses	33
3.3.4	Combinatorial Analyses	34
3.3.5	Associations with Semantic Radiological Features	34
3.4	Discussion	34
4.	NOVEL CLINICAL AND RADIOMIC PREDICTORS OF RAPID DISEASE PROGRESSION PHENOTYPES AMONG LUNG CANCER PATIENTS TREATED WITH IMMUNOTHERAPY	42
4.1	Introduction	42
4.2	Materials and Methods	43
4.2.1	Study Population and Patient Data	43
4.2.2	CT Tumor Segmentation and Radiomic Feature Extraction . . .	44
4.2.3	Rapid Disease Progression Phenotypes	45
4.2.4	Statistical Analyses	46
4.3	Results	49
4.4	Discussion	55
5.	STABILITY AND REPRODUCIBILITY OF COMPUTED TOMOGRAPHY RADIOMIC FEATURES EXTRACTED FROM PERITUMORAL REGIONS OF LUNG CANCER LESIONS	60
5.1	Introduction	60
5.2	Materials and Methods	61
5.2.1	Moist-run dataset	61
5.2.2	RIDER test-retest dataset	62
5.2.3	Prognostic lung cancer datasets	62
5.2.4	Segmentation Algorithms	62
5.2.5	Peritumoral Masks	63
5.2.6	Radiomic Features	64

5.2.7	Statistical Analyses	65
5.2.8	Survival Analyses	66
5.3	Results	66
5.3.1	Peritumoral Features	66
5.3.2	Intratumoral Features	69
5.3.3	Survival analysis of peritumoral features	75
5.4	Discussion	75
6.	CLINICAL FACTORS AND QUANTITATIVE IMAGE-BASED FEATURES PREDICT IMMUNOTHERAPY RESPONSE AMONG LUNG CANCER PA- TIENTS	80
6.1	Introduction	80
6.2	Materials and Methods	81
6.2.1	Immunotherapy-treated Lung Cancer Patients	81
6.2.2	Radiogenomics Dataset	82
6.2.3	Immunohistochemistry Dataset	82
6.2.4	Prognostic validation Dataset	83
6.2.5	Tumor segmentation and Radiomics Extraction	84
6.2.6	Statistical Analysis	84
6.3	Results	86
6.3.1	Immunotherapy Treated Patient Demographics	86
6.3.2	Clinical Model	87
6.3.3	Radiomics Model	87
6.3.4	CART Analysis	88
6.3.5	Multivariable Analysis	88
6.3.6	Radiogenomics Analysis	93
6.3.7	Immunohistochemistry Analysis	95
6.3.8	Prognostic Validation Datasets	96
6.4	Discussion	97
7.	OVERALL CONCLUSIONS	104
7.1	Radiomic Features Extracted From Radial Gradient and Radial Devi- ation Maps Associated with Survival among Lung Adenocarcinoma Pa- tients (Oncotarget 2018)	104

7.2	Novel Clinical and Radiomic Predictors Of Rapid Disease Progression Phenotypes among Lung Cancer Patients Treated With Immunotherapy (Lung Cancer, 2019)	106
7.3	Stability and Reproducibility of Computed Tomography Radiomic Features Extracted from Peritumoral Regions of Lung Cancer Lesions (Medical Physics, 2019)	107
7.4	Clinical Factors and Quantitative Image-Based Features Predict Immunotherapy Response among Lung Cancer Patients (Under review) . .	108
	APPENDIX A. SUPPLEMENTARY TABLES AND FIGURES	110
	APPENDIX B. LIST OF PUBLICATIONS PRODUCED FROM THE THESIS	137
	B.1 Scientific Journals	137
	B.2 Conference Abstracts	137
	APPENDIX C. OTHER CONTRIBUTIONS TO LITERATURE	139
	REFERENCES	141

LIST OF FIGURES

- Figure 2.1 **The Radiomics Pipeline.** Using standard-of-care imaging studies, tumor(s) are segmented by an automatic or semi-automatic algorithm and approved by a radiologist. Radiomic features are computationally extracted from region-of-interests (ROIs) within and around tumor. Radiomic image features that are redundant and non-reproducible features are eliminated, and a final set of features are combined with clinical data and conventional biomarkers (e.g., immunohistochemistry, liquid biopsies, and molecular markers). The data are analyzed and modeled to identify the most informative data elements that can be used to improve decision support for diagnosis, risk prediction, prognostication, or treatment response. 8
- Figure 3.1 Cartoon image of the four tumor masks. The region inside the black line is the tumor mask, the orange area is the core mask, the red area is the border mask, and the region outside the black line is the outside mask (combination of the half part of red region and whole blue region). 26

- Figure 3.2 Kaplan-Meier survival curves for the following features **a)** Radial gradient border standard deviation in the training cohort, **b)** Radial gradient outside-tumor separation mean in the training cohort, **c)** Radial deviation outside-border separation standard deviation in the training cohort, **d)** Radial gradient outside-border separation standard deviation (2D) in the training cohort, **e)** Radial deviation tumor standard deviation in the training cohort, **f)** For the combination of radial gradient outside-tumor separation mean (RGOTSM) and radial deviation outside-border separation standard deviation (RDOBSSD) features. Hazard ratio with 95% confidence interval is calculated for the entire group instead of subgroups. HR = 3.65; 95% CI (1.89 – 7.05g) For the combination of radial gradient outside-tumor separation mean (RGOTSM) and radial deviation outside-border separation standard deviation (RDOBSSD) features in the test cohort. 32
- Figure 3.3 Volume of interests (VOI) for two lung cancer patients with extreme differences in clinical outcomes. Radial deviation image features for the corresponding VOIs for these lung cancer patients. **a)** A tumor of a patient (Patient ID [PID]: 33) who deceased after 9 months. **b)** A tumor of a patient (PID: 75) with who had an ongoing survival after 60 months. 40
- Figure 4.1 Overall survival and progression-free survival for the training and test cohorts. 47
- Figure 4.2 **Response groups of immunotherapy patients.** The non-responders were defined as patients who developed PD < 2 months (time-to-progression < 2 months). Patients who discontinued treatment and had an accelerated tumor growth were defined as HPD. The PD without HPD patients had PD on first follow-up and did not meet the HPD criteria. Responder patients were ones who developed had an ongoing response (PR/CR or SD) or developed PD at least 2 months after the initiation of therapy (TTP \geq 2 months). 48

- Figure 4.3 Example of an HPD patient that had SD prior to the initiation of immunotherapy but developed rapid tumor growth on first follow-up and experienced more than 2-fold increase from pre-treatment tumor growth versus treatment. 48
- Figure 4.4 **AUROC**s for clinical only models, radiomic only models, and combined clinical-radiomic models. **a)** TTP < 2 months versus TTP \geq 2 months group. The clinical features in the models were: Previous lines of therapy, presence of hepatic metastasis, presence of bone metastasis and NLR. The radiomics features were: Radial gradient border SD-2D, 3D Laws E5L5E5, border 3D Laws E5E5L5 and border quartile coefficient of dispersion. **b)** HPD vs non-HPD. The clinical feature in the models was: RMH prognostic score and the radiomic feature was: border NGTDM strength. 55
- Figure 4.5 Kaplan-Meier curves for OS from initiation of therapy to date of last follow-up or date of death. The patient subgroups were based on RECIST at first follow-up for PD, SD, PR, and HPD. 58
- Figure 4.6 Kaplan-Meier curve for PFS based on novel cut-points identified by CART analysis points using individual probability of each patient for the TTP < 2 months versus TTP \geq 2 months model. CART analysis did not identify a cut-point for the HPD analysis. 59
- Figure 5.1 **Initial parameters used on segmentation algorithms.** Algorithm 1 uses a bounding circle while algorithms 2 and 3 uses seed points as initial parameters. 63
- Figure 5.2 **Peritumoral masks.** Image on the left shows a CT scan ROI. Image on the middle image shows a peritumoral region that is not bounded by the lung parenchyma mask. Image on the right shows peritumoral region bounded by the lung parenchyma mask. Red region is removed from the peritumoral region as it lies outside the lung parenchyma. 64

- Figure 5.3 Concordance correlation coefficient groups of peritumoral features bounded by lung parenchyma. The green boxes represent higher ($CCC > 0.95$), yellow boxes represent moderate ($CCC \geq 0.75$ and $CCC \leq 0.95$) and red boxes represent lower ($CCC < 0.75$) CCCs. 68
- Figure 5.4 Concordance correlation coefficient groups of intratumoral features. The green boxes represent higher ($CCC > 0.95$), yellow boxes represent moderate ($CCC \geq 0.75$ and $CCC \leq 0.95$) and red boxes represent lower ($CCC < 0.75$) CCCs. 73
- Figure 5.5 **Whisker-box plots of CCCs by feature categories.** a) CCCs of features extracted from test-retest dataset (RIDER) from intratumoral ROI and peritumoral ROIs. b) CCCs of features extracted using Algorithm 1 and Algorithm 2 from intratumoral ROI and peritumoral ROI (3 mm) of Moist-run dataset. *Texture features included GLCM, GLRLM, GLSZM and NGTDM features. 74
- Figure 5.6 Kaplan-Meier graphs of overall survival for Moffitt adenocarcinoma dataset (right), MAASTRO adenocarcinoma dataset (middle), and Radiogenomics dataset (left) utilizing radiomics modelling of peritumoral features (0-3 mm). 78
- Figure 6.1 Overall survival graph for the training, test and validation cohorts and progression-free survival graph for the training and test cohorts. 87
- Figure 6.2 Correlation matrix for the radiomic features that were significantly associated with overall survival in the univariable analysis. The feature in the final parsimonious model was GLCM inverse difference and it is found to be correlated with nine other features shown inside the green box. 89
- Figure 6.3 The CART was used to identify patient risk groups based on a model containing one radiomic feature and two clinical features. Patients were grouped from low risk to very-high risk based on the CART decision nodes and terminal nodes. 89

- Figure 6.4 Kaplan-Meier survival curves estimates for overall survival (top) in the training (left), test (middle) and validation cohorts (right), and progressive-free survival (bottom) in the training (left) and test cohorts (right). 92
- Figure 6.5 Time-dependent AUC curves for Cox regression models based on 6, 12, 24 and 36 months for training (top) and test cohorts (bottom). The AUC values were statistically not different between training and test cohorts. 93
- Figure 6.6 Association between GLCM inverse difference CT radiomic feature and high CAIX expression. **a)** Whisker-box plots representing the association between CAIX expression using merck 2DQ892208 probset and GLCM inverse difference. High and low GLCM inverse difference was found using novel cut-point defined by CART analysis **b)** Scatter plot showing the linear relationship between CAIX expression using merck 2DQ892208 probset and GLCM inverse difference. CART defined cut-off point was used to differentiate high (blue) and low (red) GLCM inverse difference. **c)** Whisker-box plots representing the association between CAIX expression using merck NM001216 probset and GLCM inverse difference. High and low GLCM inverse difference was found using novel cut-point defined by CART analysis. **d)** Scatter plot showing the linear relationship between CAIX expression using merck NM001216 probset and GLCM inverse difference. CART defined cut-off point was used to differentiate high (blue) and low (red) GLCM inverse difference. **e)** Whisker-box plots representing the association between CAIX expression on IHC staining and GLCM inverse difference CT radiomic feature. High and low GLCM inverse difference was found using novel cut-point defined by CART analysis. **f)** Scatter plot showing linear relationship between pathologist H-score for CAIX and computer derived (Aperio positive pixel count algorithm) automated CAIX scoring. 97

- Figure 6.7 Representative cases for testing the agreement between GLCM inverse difference and CAIX IHC expression. Correlation between high CAIX and high CT radiomic feature is seen on left side and correlation between low CAIX and low CT radiomic feature is seen on right side. 98
- Figure 6.8 Kaplan-Meier survival plots of patients dichotomized by radiomics score. Same cut-off point was used for dichotomizing the cohorts. **a)** Training cohort **b)** Test cohort **c)** Validation cohort, **d)** Gene-expression cohort **e)** NLST cohort **f)** Moffitt adenocarcinoma cohort **g)** MAASTRO adenocarcinoma cohort. 100
- Figure A.1 Kaplan-Meier survival curves for the features in the test cohort. **a)** RG outside-tumor separation **b)** RD outside-border separation standard deviation. 111
- Figure A.2 Kaplan-Meier curves based on cut-points using individual probability of each patient from the combined radiomic-clinical models. The first column were based on median cut-point and the second column was based on and tertiles. The top row is for TTP < 2 months versus TTP \geq 2 months analysis and the bottom row is for HPD vs non-HPD analysis. 114
- Figure A.3 **Concordance correlation coefficient groups of peritumoral features not bounded by lung parenchyma.** The green boxes represent higher (CCC > 0.95), yellow boxes represent moderate (CCC \geq 0.75 and CCC \leq 0.95) and red boxes represent lower (CCC < 0.75) CCCs. 135
- Figure A.4 Overall survival and progression-free survival for the six risk groups identified by CART in the training cohort. Groups 2, 3 and groups 4, 5 were combined for the analysis in Figure 6.4. 136

Figure A.5 **Patients on low and very-high risk groups. First column represents the primary target lesion CT scan.** Second column represents the tumor segmentation. Third column represents a gradient image of the segmented area for visualization of the tumor texture. Patient on the top was identified as a low risk patient to immunotherapy and had a less dense tumor phenotype with lower GLCM inverse difference score. Patient on the bottom was identified as a very-high risk patient and had a dense tumor phenotype with higher GLCM inverse score.



LIST OF TABLES

Table 2.1	Description of screening and early detection radiomic studies.	11
Table 2.2	Description of prognostic and survival outcome radiomic studies.	13
Table 2.3	Description of treatment response prediction radiomic studies.	15
Table 2.4	Description of radiogenomic studies.	17
Table 3.1	Semantic features analyzed.	27
Table 3.2	Patient characteristics in the training and test cohorts.	30
Table 3.3	Log-rank tests and Cox proportional hazards models for overall survival in the training cohort.	31
Table 3.4	Demographics and imaging parameters by image dichotomized features in training cohort.	35
Table 3.5	Cox proportional hazards models for overall survival in the test cohort.	36
Table 3.6	Demographics and imaging parameters by image dichotomized features in test cohort.	37
Table 3.7	Association between semantic features and radial gradient and radial deviation features.	38
Table 4.1	Patient demographics by rapid disease progression phenotypes.	50
Table 4.2	Disease burden characteristics of patients by rapid disease progression phenotypes.	51
Table 4.3	Hematology results of patients by rapid disease progression phenotypes.	52
Table 4.4	Mutational status results of patients by rapid disease progression phenotypes.	52
Table 4.5	Multivariable models of clinical covariates associated with TTP < 2 months vs TTP \geq 2 months.	53
Table 4.6	Multivariable models of clinical covariates associated with HPD vs non-HPD.	54
Table 4.7	Performance statistics from the combined multivariable models and SMOTE sub-sampled multivariable models.	57

Table 5.1	Initial parameter and algorithm comparison by Jaccard index scores.	67
Table 5.2	Distribution of stability groups of all peritumoral radiomic features extracted from ROIs that are bounded by a lung mask for different initial parameters.	69
Table 5.3	Distribution of stability groups of all peritumoral radiomic features extracted from ROIs that are bounded by a lung mask for different segmentation algorithms.	70
Table 5.4	Distribution of reproducibility groups of peritumoral radiomic features extracted from ROIs that are bounded by a lung mask.	71
Table 5.5	Distribution of reproducibility groups of peritumoral radiomic features extracted from ROIs that were not bounded by a lung mask.	72
Table 6.1	Patient characteristics by the training and test cohorts.	90
Table 6.2	Overall survival and progression free survival rates by training and test cohorts and patient risk groups.	91
Table 6.3	Univariable and multivariable Cox regression analysis for overall survival for the training and test cohorts.	94
Table 6.4	Univariable and multivariable Cox regression analysis for progression-free survival for the training and test cohorts.	95
Table 6.5	Univariable and multivariable Cox regression analysis for overall survival for the validation cohort.	96
Table 6.6	Patient characteristics by CART risk groups for the training cohort.	99
Table A.1	Radial gradient and radial deviation imaging features.	110
Table A.2	Log-rank p-values for the 17 features that are reproducible and non-redundant.	111
Table A.3	The 5-year survival rates for dichotomized image features.	112
Table A.4	Univariable analysis of the association between radiomic features and rapid disease progression phenotypes.	113
Table A.5	Distribution of stability groups of statistical peritumoral radiomic features extracted from ROIs that are bounded by a lung mask.	115

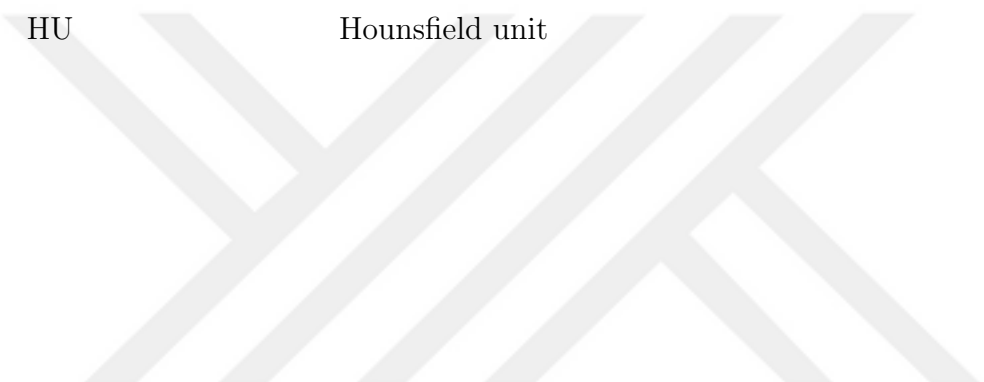
Table A.6	Distribution of stability groups of histogram peritumoral radiomic features extracted from ROIs that are bounded by a lung mask.	116
Table A.7	Distribution of stability groups of GLCM, GLRLM, GLSZM and NGTDM texture features peritumoral radiomic features extracted from ROIs that are bounded by a lung mask.	117
Table A.8	Distribution of stability groups of 3D Laws texture peritumoral radiomic features extracted from ROIs that are bounded by a lung mask.	118
Table A.9	Distribution of stability groups of wavelet texture peritumoral radiomic features extracted from ROIs that are bounded by a lung mask.	119
Table A.10	Distribution of stability groups of all peritumoral radiomic features extracted from ROIs that are not bounded by a lung mask.	120
Table A.11	Distribution of stability groups of statistical peritumoral radiomic features extracted from ROIs that are not bounded by a lung mask.	121
Table A.12	Distribution of stability groups of histogram peritumoral radiomic features extracted from ROIs that are not bounded by a lung mask.	122
Table A.13	Distribution of stability groups of texture peritumoral radiomic features extracted from ROIs that are not bounded by a lung mask.	123
Table A.14	Distribution of stability groups of 3D Laws peritumoral radiomic features extracted from ROIs that are not bounded by a lung mask.	124
Table A.15	Distribution of stability groups of wavelet peritumoral radiomic features extracted from ROIs that are not bounded by a lung mask.	125
Table A.16	Distribution of stability groups of all intratumoral radiomic features.	126
Table A.17	Distribution of stability groups of intratumoral statistical features.	127
Table A.18	Distribution of stability groups of intratumoral histogram features.	128
Table A.19	Distribution of stability groups of features intratumoral GLCM, GLRLM, GLSZM and NGTDM texture features.	129
Table A.20	Distribution of stability groups of intratumoral 3D Laws texture features.	130

Table A.21	Distribution of stability groups of intratumoral wavelet texture features.	131
Table A.22	Distribution of stability groups of intratumoral size features.	132
Table A.23	Distribution of stability groups of intratumoral shape features.	133
Table A.24	Distribution of reproducibility groups of intratumoral radiomic features.	134



LIST OF SYMBOLS

mm	Millimeter
kVp	Peak kilovoltage
U/L	Units per litre
g/dL	Grams per decilitre
mg/dL	Milligrams per decilitre
L	Litre
HU	Hounsfield unit



LIST OF ABBREVIATIONS

PD-1	Programmed death-1
PD-L1	Programmed death ligand-1
NSCLC	Non-small-cell lung cancer
IHC	Immunohistochemistry
RECIST	Response evaluation criteria in solid tumors
CT	Computed tomography
RG	Radial gradient
RD	Radial deviation
RIDER	Reference image database to evaluate therapy response
OS	Overall survival
NLST	National Lung Screening Trial
LDCT	Low-dose helical computed tomography
NELSON	Nederlands-Leuvens Longkanker Screenings Onderzoek
EGFR	Epidermal growth factor receptor
ALK	Anaplastic lymphoma kinase
DICOM	Digital imaging and communications in Medicine
SUV	Standardized uptake value
LIDCIDRI	Lung Image Database Consortium
IDRI	Image Database Resource Initiative
PET	Positron emission tomography
ROI	Region-of-interest
AUC	Area under the curve
TCIA	The Cancer Imaging Archive
LASSO	Least absolute shrinkage and selection operator
PPV	Positive predictive value
FP	False positive
SMOTE	Synthetic minority oversampling technique
CV	Cross validation

GGO	Ground-glass opacity
ANN	Artificial neural network
SVM	Support vector machine
TNM	Tumor, Node, Metastasis
SABR	Stereotactic ablative radiation therapy
KRAS	Kirsten rat sarcoma viral oncogene homolog
DCE	Dynamic contrast-enhanced
MRI	Magnetic resonance imaging
IBSI	Image Biomarker Standardization Initiative
RQS	Radiomics quality score
CPU	Central processing unit
GPU	Graphics processing unit
AI	Artificial intelligence
VOI	Voxel-of-interest
CADe	Computer-aided detection
CADd	Computer-aided diagnosis
MCC	H. Lee Moffitt Cancer Center
MAASTRO	Maastricht Radiation Oncology Clinic
LuTA	Lung Tumor Analysis
2D	2-dimensional
3D	3-dimensional
CCC	Concordance correlation coefficient
NCI	National Cancer Institute
SD	Standard deviation
HR	Hazard ratio
TTP	Time-to-progression
TGR	Tumor growth rate
ECOG	Eastern Cooperative Oncology Group
LDH	lactate dehydrogenase
HPD	Hyperprogressive disease
PD	Progressive disease

PR	Partial response
CR	Complete response
SD	Stable disease
TTF	Time-to-treatment failure
AUROC	Area under the receiver-operator characteristic
CART	Classification and regression tree
NLR	Neutrophils to lymphocytes ratio
PFS	Progression-free survival
TME	Tumor microenvironment
RMH	Royal Marsden Hospital
MDA	MD Anderson
OR	Odds ratio
CI	Confidence interval
NGTDM	Neighborhood gray tone difference matrix
GLCM	Gray level co-occurrence matrix
GLRLM	Gray level run-length matrix
GLSZM	Gray level size zone matrix
TMB	Tumor mutational burden
LFC	Logarithmic fold-change
CAIX	Carbonic anhydrase IX
FAM83F	Family with Sequence Similarity 83 Member F

1. MOTIVATION AND OBJECTIVES

Lung cancer is the leading cause of cancer-related death worldwide and accounts for more deaths than prostate, breast, colon, and pancreatic cancer combined [1]. Immunotherapy targeting programmed death-1 (PD-1) or programmed death ligand-1 (PD-L1) provide durable responses and improved long-term survival in advanced non-small-cell lung cancer (NSCLC) patients [2–7]. However, like many novel therapies, a lack of predictive biomarkers reduces foreknowledge of potential response rates to immunotherapy, which could be used to tailor individual patient’s treatment plans. Currently, PD-L1 expression extracted from immunohistochemistry (IHC) is the only applied biomarker to select patients for immunotherapy yet, PD-L1 expression alone is not adequate to predict response [8, 9]. Recent clinical trials demonstrated that immunotherapy exhibits survival benefit regardless of PD-L1 expression [7, 10] hence, additional biomarkers are significant clinical unmet need.

Medical imaging is one of the key instruments that are intensively used for diagnosis, treatment planning, monitoring and image-guided interventions of cancer patients. However, in current radiology practice, only a few simple quantitative metrics are being used to characterize tumor phenotype. As such, assessment of response to immunotherapy on clinical trials uses consensus guidelines—Response Evaluation Criteria in Solid Tumors (RECIST or iRECIST [11]) that measure only tumor(s) size by medical imaging (e.g., computed tomography [CT]) and how it objectively changes in response to therapy. Although iRECIST is widely used for evaluating immunotherapy response, it is not a baseline (pre-treatment) treatment response predictor. However, there is emerging evidence that standard-of-care medical images contain complementary and interchangeable data orthogonal to other sources such as, demographics, pathology, hematology, genomics, and proteomics. These data can be captured by a high-throughput computing technique called "Radiomics", which involves extracting computational features from a region of interest (i.e., intratumoral or peritumoral regions) that are based on size, shape, intensity and texture that quantify a variety

of phenotypic traits [12]. These quantitative features have been shown to be associated with patient outcomes independent of clinical covariates and patient characteristics [13–18]. Hence, the main objective of this thesis was to develop models utilizing radiomic and clinical covariates that can predict response of NSCLC patients in an immunotherapy setting.

1.1 Thesis hypotheses, objectives and organization

The focus of this study is on the development of radiomic-clinical based models for the prediction of outcomes of NSCLC patients in an immunotherapy setting and the main hypotheses were as follows:

- (i) *Hypothesis 1*: Quantitative image features extracted from radial gradient (RG) and radial deviation (RD) maps of CT scans have the potential to be used as prognostic, diagnostic and predictive markers of lung cancers, including but not limited to pre-surgical CT scans and immunotherapy baseline CT scans.
- (ii) *Hypothesis 2*: Quantitative CT image-based features combined with clinical covariates extracted from pre-treatment NSCLC immunotherapy patients could predict disease progression phenotypes, survival and biological underpinnings of tumors that lead to treatment resistance which can be used as biomarkers for avoiding unnecessary treatments.
- (iii) *Hypothesis 3*: Identifying stable and reproducible quantitative image features is an important precursor prior to conducting analyses of radiomics data as features with low-fidelity will likely lead to spurious findings and unrepeatable models.

Based on these hypotheses, six main objectives were defined as part of this thesis. These objectives overall fall within the aim of assisting physicians as a medical decision support system utilizing quantitative image-based features and clinical covariates that

are easily extractable from standard-of-care scans and patient characteristics. As such the objectives are listed below:

- (i) *Objective 1*: Create a comprehensive review on the developments of the radiomics of lung cancer to guide the scientific community on the current advancements and future directions (Chapter 2).
- (ii) *Objective 2*: Extract features from RG and RD maps from pre-surgical contrast-enhanced thoracic CT scans among patients with lung adenocarcinoma and assess the associations with overall survival (Chapter 3).
- (iii) *Objective 3*: Develop parsimonious models that can identify NSCLC patients treated on clinical trials of immunotherapy that are at risk of rapid disease progression utilizing CT scans and clinical covariates (Chapter 4).
- (iv) *Objective 4*: Assessing the stability and reproducibility of CT peritumoral and intratumoral radiomic features of lung lesions (Chapter 5).
- (v) *Objective 5*: Develop parsimonious models utilizing clinical data and CT scans of NSCLC patients treated on immunotherapy clinical trials that can predict survival outcomes (Chapter 6).
- (vi) *Objective 6*: Finding potential biological and radiological underpinnings of predictive and prognostic radiomic features by comparing with semantic radiology, gene expression profiling and IHC (Chapter 3 and Chapter 6).

This thesis consists a total of seven chapters and three appendixes. Since the thesis was based on multiple-manuscripts, chapters were written in a self-contained manner. However, all analysis on chapters are complementing to reaching the optimal goal of identifying clinical and image-based biomarkers for NSCLC immunotherapy. The chapters of the thesis is as follows:

- (i) *Chapter 2*: Introduction to epidemiology of lung cancer, radiomics and particularly radiomics of lung cancer.

- (ii) *Chapter 3*: Description of the first published manuscript which radiomic features extracted from RG and RD maps were used to assess overall survival of pre-surgical lung adenocarcinoma patients. *Published in Oncotarget 2017 Nov 10; 8(56): 96013–96026.*
- (iii) *Chapter 4*: Description of the second published manuscript which clinical and radiomic predictors of rapid disease progression phenotypes were found in the setting of NSCLC immunotherapy. Features developed on Chapter 3 were also utilized in this study. *Published in Lung Cancer 2019 Mar; 125: 75-79.*
- (iv) *Chapter 5*: Description of the third published manuscript which stability and reproducibility of radiomic features extracted from intratumoral and peritumoral regions were assessed. To measure stability the “Moist run” dataset which consists of multiple segmentations of lesions [19] and to measure reproducibility the Reference Image Database to Evaluate Therapy Response (RIDER) dataset that consists of test-retest data [20] were utilized. *Published in Medical Physics 2019 Sep 08; doi: 10.1002/mp.13808*
- (v) *Chapter 6*: Description of the fourth manuscript (under review) which clinical and radiomic predictors of patient survival in the setting of NSCLC immunotherapy. The models created in this study included only stable and reproducible radiomic features which were identified in Chapter 5. A training, an internal test and an external validation cohort were utilized in this study. Also biological underpinnings of radiomic signature were explored using separate gene-expression and IHC datasets. *Under review at Nature Communications.*
- (vi) *Chapter 7*: Overall conclusions of the thesis were given by summarizing and highlighting the novelty of each study described in previous Chapters. Potential future directions regarding each study were discussed.
- (vii) *Appendix 1*: Supplementary methods, figures and tables are given.
- (viii) *Appendix 2*: List of publications produced from the thesis are given.
- (ix) *Appendix 3*: Other contributions to the literature during the time span of this dissertation are given.

2. LUNG CANCER AND RADIOMICS

2.1 Lung Cancer

Lung cancer is a major public health issue and the leading cause of cancer-related deaths among both men and women in the United States [1]. Approximately 57% of lung cancer patients exhibit metastatic disease stage at the time of diagnosis [21]. Unfortunately, a vast majority of these patients have severe outcomes with 5-year overall survival (OS) rates of only 6% on a distant stage diagnosis [22].

Use of tobacco products is the major cause of lung cancer; however, cessation missions are not enough to win the war against lung cancer as there are other genetic [23] and environmental factors [24], as well as “bad luck” [25] that can lead to a lung cancer development. There is a pressing clinical necessity for accurate and affordable screening methods to catch lung cancers at an earlier stage while it is still localized. One of the major progress that reduce lung cancer mortality rates was based on the study by The National Lung Screening Trial (NLST) which compared standard chest radiography and low-dose helical computed tomography (LDCT) for early detection of lung cancer [26]. A 20% relative reduction in lung cancer mortality was perceived after a median follow-up of approximately 6 years for LDCT compared to chest radiography. Similar results were seen in a European trial, Netherlands-Leuven Longkanker Screenings ONderzoek (NELSON), in 2018 indicating significant reductions in lung cancer mortality and confirming the efficacy of lung cancer screening [27]. At the other end of the detection spectrum, sophisticated therapies have made substantial progress on the fight against advanced lung cancers over the past 20 years. These therapies include targeted therapies such as Erlotinib and Gefitinib (and now Osimertinib) to treat patients with activating epidermal growth factor receptor (EGFR) mutations [28, 29], or Crizotinib for anaplastic lymphoma kinase (ALK) rearranged lung cancers [30]. However, most lung cancers do not have a known driver mutation, and thus many novel therapies fail to be useful because viable, predictive biomarkers cannot be found. An exception

may be immunotherapy that target the checkpoint blockades, where total mutational burden or PD-L1 status may predict response. However, emerging data have shown that durable responses can occur with negative biomarker results and some patients have rapid progression even if these biomarkers are positive [31]. Thus there remains a clinically unmet need to develop more robust and predictive biomarkers for response to checkpoint blockades.

Even though a significant reduction in lung cancer mortality rates has been seen in the last decade [1], pathologic staging remains to be the most important prognostic factor for prognosing lung cancer survival [32]. However, there is marked variability in patient outcomes and survival among patients with the same stage of disease, which suggests that other factors contribute to lung cancer prognosis [33–37].

Emerging targeted treatment options and the rise of precision medicine for cancer has made a remarkable increase in the quantity and types of information that are being extracted from individual patients and tumors. Medical imaging is one of the key instruments that are intensively used for diagnosis, treatment planning, monitoring and image-guided interventions of cancer patients. These data are stored digitally in a standardized Digital Imaging and Communications in Medicine (DICOM) format that readily be used for analysis. However, in current radiology practice, only few simple quantitative metrics are being used to quantify tumor phenotype, such as largest diameters of tumors (i.e., RECIST) by CT or maximum and mean standardized uptake value (SUV) from positron emission tomography (PET), while the remainder of evaluative analytics are handled in a qualitative manner making it user-dependent. However, there is emerging evidence that standard-of-care medical images contain complementary and interchangeable data that are orthogonal to other data types such as, demographics, pathology, hematology, genomics, and proteomics. These imaging data can be captured by a high-throughput computing technique called Radiomics which involves extracting computational image-based features from a region of interest (i.e., intratumoral or peritumoral region) of a medical imaging scan. Features are based on size, shape, intensity and texture that quantify a variety of phenotypic traits [12]. These quantitative radiological imaging features (i.e., Radiomics) are associated with

cancer patients' diagnosis, prognosis, and/or their tumor genotype independent of clinical covariates and other patient characteristics [13–18].

In the remaining of this chapter, the field of Radiomics is described and discussed in the setting of lung cancer. Potential advantages and pitfalls are mentioned thoroughly.

2.2 Cancer Heterogeneity and the Utility of Radiomics

Cancers are heterogeneous across a wide range of spatial and temporal scales which results in regional variations in metabolism, vasculature, oxygenation and gene expression [38,39]. Even with the same histology, tumors may still have unique driver mutations, proteomic profiling, and/or aggressiveness. Technical advances allow extensive molecular characterization of tumor cells in each individual patient which enables precise individualized cancer treatment. However, a single arbitrary sample taken from the tumor using needle biopsy may only represent a small sub-region of the tumor parenchyma potentially generating misleading results via “sampling artifact”. On the other hand, Radiomics can capture tumor phenotype from a 3-dimensional space from multiple sites reducing the sampling-bias and has no time-constraint hence can be captured using longitudinally.

The process of radiomics involves five fundamental steps (a) Image acquisition and digitization, (b) ROI selection and segmentation, (c) Quantitative feature extraction and feature selection, (d) biomarker discovery and modeling and (e) validation (Figure 2.1). The first step involves image-data acquisition and digitization. Radiomics does not require further imaging of patients but rather uses standard-of-care images. Since imaging is a fundamental instrument of cancer treatment from diagnosis to treatment planning and assessing patient response, radiomics have the potential to benefit greatly from this readily available image-data. However, image acquisition protocols vary widely across different medical centers and sometimes even in between the same institutions. Thus, it is crucial to maintain a homogeneous image acquisi-

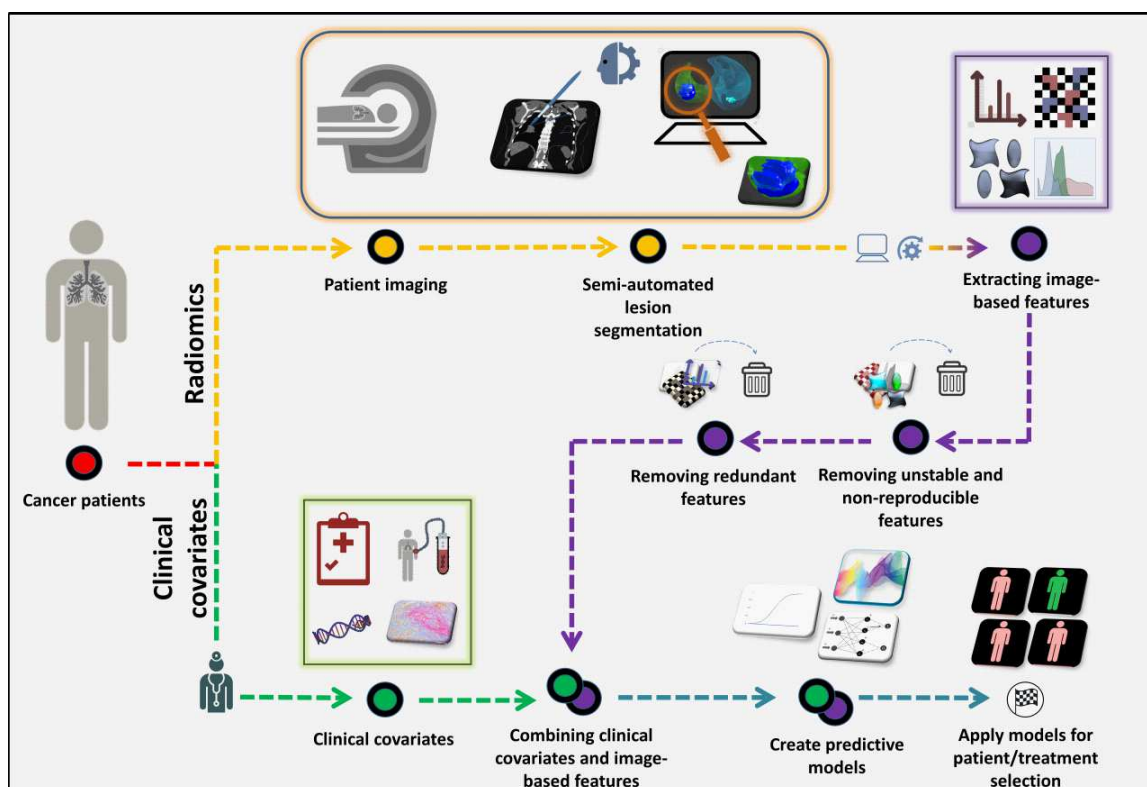


Figure 2.1 The Radiomics Pipeline. Using standard-of-care imaging studies, tumor(s) are segmented by an automatic or semi-automatic algorithm and approved by a radiologist. Radiomic features are computationally extracted from region-of-interests (ROIs) within and around tumor. Radiomic image features that are redundant and non-reproducible features are eliminated, and a final set of features are combined with clinical data and conventional biomarkers (e.g., immunohistochemistry, liquid biopsies, and molecular markers). The data are analyzed and modeled to identify the most informative data elements that can be used to improve decision support for diagnosis, risk prediction, prognostication, or treatment response.

tion parameter space across the cohort being analyzed. The second step involves the ROI selection and segmentation. The choice of ROI is typically the primary tumor or index lesions when the disease is metastatic. Increasingly, however, the peritumoral regions are being used as an ROI. Segmentation of the ROI can directly affect radiomic feature values; in fact some radiomic features are extremely sensitive to segmentation bias [40] rendering them uninformative. Hence, ideally, segmentation methods should produce accurate, reproducible and consistent segmentations in an automated manner and require minimal user input in order to increase stability of features. The third step consists of feature extraction that includes shape, intensity, texture and wavelet as well as location or orientation of the ROI. To reduce overfitting, features that are non-reproducible and unstable could be eliminated and principal component analy-

sis or clustering methods could be used to reduce dimensionality. The fourth step is the biomarker discovery and modelling step which includes creating models utilizing biostatistics or machine learning methods to find the most informative radiomics correlated with the desired end-points of the study (e.g., overall survival, treatment response, genomics data...etc.). To achieve comprehensive models, clinical data such as demographics, genetic information, and/or hematology should be incorporated to the final models. Lastly, the trained models need to be further validated by assessing accuracy on independent external, if not internal datasets. External validation gives highest credibility to the study as it shows that the model can be translated between multiple cohorts and institutions, demonstrating its independence.

Many studies have been conducted on radiomics in the setting of lung cancer. Some of the major studies in the literature based on four different categories: 1- Early detection radiomic studies, 2- Prognostic and survival outcome studies, 3- Treatment response prediction studies and 4- Radiogenomic studies and were discussed below.

2.3 Early Detection Radiomic Studies

Differentiating malignant from benign tumors has been a long ongoing subject well before the introduction of radiomics [41, 42]. With the era of radiomics, a vast number of quantitative feature are able to quantify the lesion phenotype, which has potential to improve diagnostic accuracy. As diagnostic CT and PET/CT scans are routinely obtained during the workup of lung cancer patients, image-based features can provide valuable and readily available complementary decision support information which have translational implications for improved lesion stratification.

After the NLST was made publicly available, many studies were performed on images of this immense cohort. One of the pioneering works was done by Hawkins *et al.* [43] where baseline screening CT scans from the NLST were used to predict indeterminate pulmonary nodules that would subsequently develop into malignant tumors. They used a set of stable and non-redundant radiomic features that were fed into a

group of classifiers where best accuracy was achieved using a random forests classifier generating area under the curve (AUC) of 0.81 for predicting development of cancer in 1 year, with highest and lowest radiomics scores having accuracies of 0.93. Peikert *et al.* [44] created a radiomics model to compare malignant and benign screen-detected indeterminate lung nodules utilizing the NLST dataset. Using least absolute shrinkage and selection operator (LASSO) multivariable analysis, they reported an AUC of 0.939 with 8 features. Huang *et al.* [45] performed a matched case-control study to find the added value of CT image-based features for the early detection of small pulmonary nodules. Utilizing the NLST dataset, they extracted texture features within the tumor ROI as well as the peritumoral ROI. They showed that image-based features increased the positive predictive value (PPV) and reduced the false positive (FP) rates compared to thoracic radiologist evaluations. Cherezov *et al.* [46] improved malignancy prediction accuracy from 74.7% to 81.0% by implementing separate models based on different nodule sizes on NLST dataset. They used Synthetic Minority Oversampling Technique (SMOTE) to overcome the class imbalance. Chae *et al.* [47] utilized texture features to differentiate pre-invasive lesions from invasive pulmonary adenocarcinomas that were marked as part-solid ground-glass opacities (GGOs) on chest CT scans. Their artificial neural network (ANN) model showed an excellent performance using five radiomic features with an AUC of 0.981 on 86 part-solid GGOs. Liu *et al.* [48] identified semantic features (i.e., radiological traits quantified by radiologists) that are predictive for malignancies in lung nodules. Dhara *et al.* [49] utilized 891 nodules from the Lung Image Database Consortium and Image Database Resource Initiative database (LIDC-IDRI) and classified malignant versus benign nodules using support vector machine (SVM). Their models reached an AUC of 0.951 which outperformed methods that required manual segmentation of a trained radiologist. Table 2.1 shows the summary of the screening and early detection radiomic studies.

Table 2.1
Description of screening and early detection radiomic studies.

Author	Imaging modality	Major end-point	N (training + test cohorts)	Study cohort	Number of features in the model	Evaluation metric	Analytical method used	Study description	Validation Type
Hawkins et al. [43]	CT	Diagnosis	176+152	NLST	10	AUC = 0.81	Random forest classifier	Predicting benign vs malignant nodules on a screening cohort	Internal validation with a separate dataset
Peikert et al. [44]	CT	Diagnosis	726	NLST	8	AUC = 0.939	LASSO	Differentiating benign vs malignant from indeterminate nodules	Internal validation utilizing bootstrapping
Huang et al. [45]	CT	Diagnosis	140+46	NLST	5 radiomic + 1 clinical	AUC = 0.9154	Random forest classifier	Differentiating benign vs malignant from small nodules (≤ 20 mm)	Internal validation with a separate dataset
Cherezov et al. [46]	CT	Diagnosis	255+212	NLST	Ranging from 5 to 10	AUC = 0.76 to 0.86	Machine learning	Differentiating benign vs malignant based on nodule size	Internal validation with a separate dataset
Chae et al. [47]	CT	Diagnosis	86	private	2 and 5	AUC = 0.981	Logistic regression + ANN	Differentiating preinvasive lesions from invasive pulmonary adenocarcinomas	No validation
Liu et al. [48]	CT	Diagnosis	102+70	private	4 semantic features	AUC = 0.74 and 0.80	Linear classifier	Differentiating benign vs malignant from incidentally identified nodules	Internal validation with a separate dataset
Dhara et al. [49]	CT	Diagnosis	891	LIDC/IDRI	Unknown	AUC = 0.8488 to 0.9505	SVM	Predicting benign vs malignant nodules	No validation

Abbreviations: CT = computed tomography; AUC = area under curve; SVM = support vector machine; LASSO = least absolute shrinkage and selection operator; ANN = artificial neural network; LIDC/IDRI = Lung Image Database Consortium and Image Database Resource Initiative database.

2.4 Prognostic and Survival Outcome Radiomic Studies

Accurately determining the aggressiveness of tumors is critical as it can help physicians develop a prognosis and provide options to choose between curative and palliative treatments as well as the aggressiveness of the therapy. Pathological staging remains to be the most important prognostic factor for lung cancer survival [32]. However, radiomic studies have shown that image-based biomarkers have the potential to complement other types of biological markers and improve prognosis of lung cancer which will assist physicians on determining the treatment type.

Aerts *et al.* [50] analyzed NSCLC and head and neck cancer patients and validated a CT radiomic signature that had better prognostic performance than covariates such as Tumor, Node, Metastasis (TNM) staging and volume. They found associations between their radiomic signature and gene-expression patterns using gene-set enrichment analysis where the most informative features were found to be correlated with cell cycling pathways. Grove *et al.* [14] developed two CT features; convexity and entropy ratio, which were significantly associated with OS of patients diagnosed with primary lung adenocarcinoma utilizing two independent cohorts. Tunali *et al.* [51] assessed the same cohorts and developed novel radiomic features generated from RG and RD maps that also associated with OS. Coroller *et al.* [52] built a combined model of CT radiomics and clinical predictors that is correlated with distant metastasis while Wu *et al.* [53] utilized fluorine 18 (18F) PET/CT based radiomic features to overcome same problem. Huang *et al.* [54] found radiomic signatures that correlated with disease-free survival. Several studies [55–57] investigated the prognostic performance of CT radiomic features for distant metastasis and loco-regional recurrence after stereotactic body radiation therapy. Win *et al.* [58] showed that heterogeneity on both CT and PET components of PET/CT were significant predictors of survival. Chae *et al.* [47] and She *et al.* [59] found CT radiomic signatures that differentiated indolent versus invasive lung adenocarcinoma. Table 2.2 shows the summary of the screening and early detection radiomic studies.

Table 2.2
Description of prognostic and survival outcome radiomic studies.

Author	Imaging modality	Major end-point	N (training + test cohorts)	Study cohort	Number of features in the model	Evaluation metric	Analytical method used	Study description	Validation Type
Aerts et al. [50]	CT	OS	422+ 225	private + TCIA	4	Concordance index = 0.65	Cox proportional hazards	Prognostic power of radiomic features and the underlying gene-expression patterns.	External validation
Grove et al. [14]	CT	OS	61+47	Will be available on TCIA	1	P value : 0.008	Cox proportional hazards	Prognostic power of newly developed radiomic features	External validation
Tunali et al. [51]	CT	OS	61+47	Will be available on TCIA	1	P value : 0.014	Cox proportional hazards	Prognostic power of newly developed radiomic features	External validation
Coroller et al. [52]	CT	OS & Distant metastasis	98+84	private	3	Concordance index = 0.61	Cox proportional hazards	Predicting distant metastasis	External validation
Wu et al. [53]	PET	Distant metastasis	70+31	private	2	Concordance index = 0.71	Cox proportional hazards	Predicting distant metastasis	Internal validation with a separate dataset
Huang et al. [54]	CT	DFS	141+141	private	5	Concordance index = 0.691	LASSO	Predicting DFS in early stage patients	Internal validation with a separate dataset
Huynh et al.[55]	CT	OS	131	private	13	AUC = 0.667	Correlation analysis (Spearman's)	Early stage disease recurrence prediction of patients treated with SBRT	No validation
Li et al. [56]	CT	OS	92	private	2 radiomic + 1 clinical + 1 semantic	Log-rank p-value = 0.0002	Cox proportional hazards	Early stage disease survival prediction of patients treated with SBRT	No validation
Oikionomou et al. [57]	CT + PET	OS	150	private	7	(logrank chi-square: 8.92, p = 0.002	Cox proportional hazards	Predict clinical outcome in lung cancer patients treated with SBRT	No validation
Win et al. [58]	(FDG) PET/CT	OS	56 + 66	private	2 radiomic + 1 clinical	Cox regression p-value < 0.001	Cox proportional hazards	Predicting OS	External validation
Chae et al. [47]	CT	Prognostic	86	private	2	AUC = 0.981	ANN	Differentiate preinvasive lesions from invasive pulmonary adenocarcinomas	No validation
She et al. [59]	CT	Prognostic	207 + 195	private	5	AUC = 0.95	Logistic regression	Differentiate indolent from invasive pulmonary adenocarcinomas	Internal validation with a separate dataset

Abbreviations: CT = computed tomography; AUC = area under curve; TCIA = The Cancer Imaging Archive; PET = positron emission tomography; OS = overall survival; ANN = artificial neural network; LASSO = least absolute shrinkage and selection operator;

2.5 Treatment Response Prediction Radiomic Studies

Early assessment of a therapeutic efficacy and predicting treatment outcomes would aid clinicians to decide which treatment has the optimal benefit for the patient. Potentially this will eliminate unnecessary treatments reducing toxicities, costs, and increasing patient survival. Thus, accurate and robust early predictive models are clinical needs for all cancer types.

Coroller *et al.* [60] investigated CT radiomic features extracted from primary lung tumors and lymph nodes to predict pathological complete response after neoadjuvant chemoradiation before surgery. They showed that a clinical-radiomics model had the highest median AUC and performed significantly better than the radiomics only or clinical only models. Yu *et al.* [61] retrospectively analyzed training and validation cohorts of NSCLC patients that were treated with surgery or stereotactic ablative radiation therapy (SABR). Their CT radiomic model was able to identify patients by mortality risk and also shown association with distant metastasis. Mattonen *et al.* [62] compared a machine learning radiomics based approach versus a physician to detect local recurrence after SABR. Their radiomic signature consisted of five features which discriminated local recurrence from fibrosis with an AUC of 0.85. Khorrami *et al.* [63] utilized peri- and intratumoral CT radiomic features to predict pemetrexed-based chemotherapy response. They showed that peritumoral features were predictive for time-to-progression. Fave *et al.* [64] utilized delta radiomics (i.e., changes in radiomic features in longitudinal scans) and showed that radiomic feature alterations after radiation therapy may represent tumor response. Tunali *et al.* [65] and Trebeschi *et al.* [66] utilized CT radiomics to address a recent clinical unmet need of identifying immunotherapy response. Both studies had significant performances with AUCs over 0.80 to discriminate responders versus non-responders. Table 2.3 shows the summary of the screening and early detection radiomic studies.

Table 2.3
Description of treatment response prediction radiomic studies.

Author	Imaging modality	Major end-point	N (training + test cohorts)	Study cohort	Number of features in the model	Evaluation metric	Analytical method used	Study description	Validation Type
Coroller et al. [60]	CT	pCR	85	private	2 radiomic + 1 clinical	AUC = 0.67, p < 0.05	Random forest	Predicting pathological response after neoadjuvant chemoradiation	No validation
Yu et al. [61]	CT	OS	147 + 295	private	2	log-rank p-value = 0.017	Random survival forests	Predicting the clinical outcomes for patients with stage I NSCLC treated with SABR	Internal validation with a separate dataset
Mattonen et al. [62]	CT	Local recurrence	45	private	5	FPR = 24.0%, FNR = 23.1%	Machine learning	Assess physician ability to detect timely local recurrence and to compare physician performance with a radiomics tool	No validation
Khorrani et al. [63]	CT	OS and TTP	72 + 53	private	7	AUC = 0.77	Minimum redundancy maximum Cox	Discriminative ability of radiomic features on response to chemotherapy	Internal validation with a separate dataset
Fave et al. [64]	CT	Local recurrence	107	private	1	Log-rank p-value = 0.269	proportional hazards	Assessing radiation therapy response by utilizing delta radiomics	No validation
Tunali et al. [65]	CT	Rapid disease progression	228	private	4 radiomic + 4 clinical	AUC = 0.8040	Logistic regression	Identifying rapid disease progression phenotypes in patients treated with immunotherapy	No validation
Trebeschi et al. [66]	CT	Response to immunotherapy	133+70	private	2	AUC = 0.83, p < 0.001	Machine learning	Identifying radiomic biomarkers for immunotherapy response	Internal validation with a separate dataset

Abbreviations: CT = computed tomography; AUC = area under curve; OS = overall survival; SABR = stereotactic ablative radiation therapy; TTP= time-to-progression; pCR = pathological complete response; FPR = false positive ratio; FNR = false negative ratio.

2.6 Radiogenomics Studies

Cancer is a heterogeneous disease where genomic heterogeneity, molecular and microenvironmental events contribute to the aggressiveness and therapy response. Genomic profiling is being used for appropriate treatment selection for advanced lung cancer [67]. Meanwhile, studies have shown that radiomic features capture the link between the cancer genomics and tumor phenotype which is called radiogenomics. (Nota bene: In some publications, “Radiogenomics” is a term to describe genomic prediction of radiation response patterns, however those studies were not referred in this section).

Wu *et al.* [67] observed that some CT texture features were associated with NSCLC tumor histology. Velazquez *et al.* [68] found clinical-radiomics signatures that differentiated *EGFR* and Kirsten rat sarcoma viral oncogene homolog (*KRAS*) mutations, the most common somatic mutations in lung adenocarcinomas. On the other hand, Gevaert *et al.* [69] utilized semantic features (to predict *EGFR* and *KRAS* mutations; however their models were only able to predict for *EGFR* mutations accurately. Liu *et al.* [70] utilized CT radiomics to predict *EGFR* mutation status in an Asian cohort who had surgically-resected peripheral lung adenocarcinomas. Weiss *et al.* [71] similarly looked at CT texture features that are discriminated *KRAS* mutant tumors from pan-wildtype tumors. Yamamoto *et al.* [72] combined clinical covariates and CT based features to characterize tumors with *ALK*+ rearranged NSCLC. Yoon *et al.* [73] identified clinical and CT and PET radiomic predictors for *ALK/ROS1/RET* fusion-positive lung adenocarcinoma. Zhou *et al.* [74] integrated semantic CT features with next-generation RNA sequencing data to identify radiogenomic biomarkers. They validated 10 metagenes annotated by functional gene enrichment analysis that resulted in significant associations with semantic CT image features. Table 2.4 shows the summary of the radiogenomic studies.

Table 2.4
Description of radiogenomic studies.

Author	Imaging modality	Major end-point	N (training + test cohorts)	Study cohort	Number of features in the model	Evaluation metric	Analytical method used	Study description	Validation Type
Wu et al. [67]	CT	Tumor histology	350	private	5	AUC = 0.72; p = 2.3×10^{-7}	Machine learning	Finding association between radiomic features and the tumor histologic subtypes	No validation
Velazquez et al. [68]	CT	<i>EGFR</i> mutation	353+352	private	21	AUC = 0.70	Random forest	Discriminating between <i>EGFR</i> ⁺ and <i>EGFR</i> ⁻	External validation
Gevaert et al. [69]	CT	<i>EGFR</i> mutation	186	private	5 semantic features	AUC = 0.87	Multivariate decision tree	Discriminating between <i>EGFR</i> ⁺ and <i>EGFR</i> ⁻	No validation
Liu et al. [70]	CT	<i>EGFR</i> mutation	298	private	4 radiomic +2 clinical	AUC = 0.709	Logistic regression	Discriminating between <i>EGFR</i> ⁺ and <i>EGFR</i> ⁻	No validation
Weiss et al. [71]	CT	<i>KRAS</i> mutation	48	private	2	Accuracy = 89.6%	Recursive decision tree	Differentiation between K-ras mutation and pan-wildtype	No validation
Yamamoto et al. [72]	CT	<i>ALK</i> mutation	59 +113	private	3 radiomic +1 clinical	Accuracy = 78.8%	Random forest	Discriminating between <i>ALK</i> ⁺ and <i>ALK</i> ⁻	External validation
Yoon et al. [73]	PET/CT	<i>ALK/ROS1/RET</i> mutations	539	private	4 radiomics + 1 qualitative image features + 2 clinical	Sensitivity = 0.73, Specificity = 0.70	Chi-squared test and Student t test	Predictors of tumors with <i>ALK</i> , <i>ROS1</i> , or <i>RET</i> fusions	Internal validation utilizing CV

Abbreviations: CT = computed tomography; AUC = area under curve; PET = positron emission tomography; CV = cross validation.

2.7 Limitations of Radiomics

Although radiomics have shown promise for potentially improving diagnostic, prognostic, and predictive accuracy of cancer, there are still limitations and hurdles to overcome. The two most important limitations which will be discussed are the lack of reproducible studies and poor study designs that lead to spurious results and create underpowered studies.

There is a broad spectrum of factors contributing to non-reproducible results that range from imaging to segmentation and to statistical flaws. Standard-of-care image acquisition parameters have a wide range. These parameters include, but are not limited to: pixel spacing, slice thickness, reconstruction kernel, kVp, washin and washout periods on PET and dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) scans, administration of contrast agents, as well as gradient strengths, field strengths, echo time, and repetition time on MRI scans. Intra- and inter-scanner variabilities affect these parameters, which can cause radiomic feature distributions to change. To overcome these issues, efforts need to be expended to at least partially standardize acquisition and reconstruction protocols, and to develop post-acquisition corrections for some of these variables, such as linear interpolation to a fixed voxel size. On the other hand, if radiomic features are being extracted from a heterogeneous image acquisition parameter space, features that are less sensitive to these parameters which eventually shown to have higher predictive/prognostic value, should be used eliminating the features that are more sensitive [75]. Meanwhile, computational radiomic feature calculations involve many critical processing steps that include pre-processing, spatial interpolation and intensity discretization. Recently an international group of radiomic researchers formed the Image Biomarker Standardization Initiative (IBSI) to tackle the issue of standardizing radiomic features across multiple institutions [76]. The goal of the initiative was to set consensus and provide benchmarks on the most common radiomic features as well as image processing steps before feature extraction where benchmark calculations were performed on digital synthetic images. All researchers in the field to set their radiomic features are encouraged to use the IBSI standards in order to increase the chance of reproducible research.

Another important factor that affects the reproducibility is the segmentation of the region-of-interest (i.e., tumor parenchyma or peritumoral parenchyma). Manual segmentations are particularly time consuming and often leads to intra- and inter-observer variations. To overcome this, segmentation of the tumors can be performed by fully- or semi-automated algorithms which involve minimal user variations such as simple initializations (e.g., seed point), followed by a computer-based delineation of the ROI. However, to tackle this issue further, segmentation algorithms across institutions have to be standardized to achieve consistent delineations. Nevertheless, many of the features are not reproducible even when acquired within a few minutes using same image acquisition parameters [77] or when the same segmentation algorithms were being used [19]. Hence, researchers working on lung cancer radiomics are encouraged to choose reproducible features by utilizing test re-test datasets, such as RIDER, and stable features by utilizing multiple segmentation datasets such as the Moist run dataset [40].

Another important factor affecting quality of results is incorrect study designs that can increase the rate of false discoveries [39]. With the potential wide range of hyperparameters such as number of filters, feature categories, and other adjustable parameters, there are theoretically an unlimited numbers of radiomic features that can be used. As a matter of fact, many studies include too many features without accounting for multiple testing errors which leads to feature selection bias, false positive results or underpowered studies [78–80]. As a rule of thumb, Chalkidou *et al.* [81] suggested that using at least 10-15 observations (i.e., patients) per predictor variable (i.e., radiomic feature) will realistically reduce false discovery rate. Another potential application is to correct significant p-values for multiple testing using methods such as Bonferroni-Holm or Benjamini-Hochberg methods [82–84]. If estimates of predictive performance are conducted from a cohort of a single institution, multiple-folded repeated cross validation should be performed to minimize the risk of overfitting. However, utilizing an internal and/or ideally an external validation cohort with similar patient demographics and status is the optimal method to validate findings to avoid spurious findings. Researchers could also assess whether the found model can be applied on a distinct patient population (making it predictive) or reflect a *pan-signature* that can be used

on multiple patient subgroups (making it prognostic).

Overall, to assess the quality of a radiomics study, a new metric called radiomics quality score (RQS) has been proposed by Lambin *et al.* [85] which evaluates radiomic studies by a series of questions on internal consistency, reproducibility and clinical applicability. The RQS score does not claim to evaluate the significance of the study yet it quantifies proper study design and scientific usefulness. Researchers are highly suggested to evaluate their studies by RQS and try to get a high score for more repeatable and quality science in the field of radiomics. To create a benchmark for RQS score, a recent meta-analysis [86] analyzed 77 published papers and documented that the mean RQS across all studies was 9.4 ± 5.6 where the potential maximum of RQS is 36.

2.8 Future Guidelines

Radiomics is a tool for clinical decision support which does not have the goal to remove radiologists from clinical practice but rather act as a complementary system that converts radiological examinations into quantitative features that ultimately improve diagnostic accuracy and prediction power. As such, it is intended to extract information utilizing standard-of-care images that routinely available. However, as mentioned before, there are limitations that are still unmet to make the transition from research to clinical practice. The most crucial step to make the leap to clinical translation is testing the clinical utility of such image-based markers by randomized prospective clinical trials. In order to achieve this, proper study designs and high-quality research needs to be put in place to reduce false discovery rates and spurious findings.

With an increasing number of medical images being digitized and being readily available for analysis, applying machine learning methods, specifically deep learning, for predicting patient outcomes attracts many researchers. Deep learning has unprecedented success due to advances in central processing units (CPUs) and graphics processing units (GPUs), the availability of big data due to increased storage and dig-

itization. Deep learning models are most effective when the numbers of samples are sufficiently high, e.g. tens of thousands to millions. However, due to the logistical hurdles associated with sharing medical data between institutions, most studies still fall way too short of reaching these sample sizes to get properly working artificial intelligence (AI) models. To overcome this challenge, researchers use data augmentation methods such as affine transformation [87] or generative adversarial networks [88] to generate artificial samples and/or use transfer learning methods [89,90] where the neural networks are trained using natural images and only the final output layer is trained with the target-task samples (i.e., medical images) and/or use two-dimensional images instead of using three-dimensional images to reduce the input layer size (Note bene: This is contradictory with the premise of radiomics to quantify 3-D tumor phenotype). Another potential way to tackle this is through a centralized databases or using a distributed learning platform where the ‘code’ is shared instead of the data [85]. Although methods are used to overcome the low number of sample sizes, it is believed that more work is necessary to make the transition to clinic for the deep learning methods. Nevertheless, deep learning methods for the segmentation of tumors [91] is a much more applicable process since segmentation outputs are binary and the sample sizes are much bigger due to a pixel wise decision making.

2.9 Conclusions

In this review, an overview of radiomics was given, particularly on the setting of lung cancer. Studies that had impact on the field as well as potential limitations and future directions were summarized.

3. RADIOMIC FEATURES EXTRACTED FROM RADIAL GRADIENT AND RADIAL DEVIATION MAPS ASSOCIATED WITH SURVIVAL AMONG LUNG ADENOCARCINOMA PATIENTS

3.1 Introduction

Lung cancer is the leading cause of cancer-related death in the United States and accounts for more deaths than prostate, breast, colon, and pancreatic cancer combined [1]. Pathological staging is the most important prognostic factor for lung cancer survival [32]. However, there is marked variability in patient outcomes and survival among patients with the same stage of disease, which suggests that other factors contribute to NSCLC prognosis. These prognostic factors include sex, histology, genetic alterations in oncogenes and tumor suppressor genes, co-morbidities, and patient performance status [33–37]. Additionally, there is emerging evidence that radiological and quantitative image features are associated with patient outcomes independent of clinical covariates and patient characteristics [13–18]. As diagnostic CT scans are routinely obtained during the workup of lung cancer patients, image features can provide valuable and readily available complementary decision support information which could have translational implications for improved prediction of patient outcomes and further patient stratification.

With high-throughput computing, it is now possible to rapidly extract a large number of quantitative image features from standard-of-care imaging such as CT. The conversion of digital medical images into mineable high-dimensional data is a process that is known as Radiomics. Radiomics is motivated by the premise that biomedical images contain information that reflects the underlying pathophysiology of the region of interest (i.e., lung tumor) and that these relationships can be revealed via conversion of images to structured data, data-mining, and statistical analysis [12]. As part of this

study, a set of image features extracted from RG and RD maps generated from thoracic CT images were analyzed. For each voxel in the voxel-of-interest (VOI), RD and RG value were calculated which in-turn formed the RG and RD maps. Each voxel in the RD map is defined as the angle between a voxel’s gradient vector and its radial vector (i.e., vector pointing towards the center of mass of the segmented lesion), whereas each voxel in the RG map specified the magnitude of gradient along that voxel’s radial vector. Using these maps, RG and RD features which represent voxel-by-voxel gradient changes from the selected VOIs were generated. As such, these features were expected to be sensitive to changes in tumor shape that occur along radial directions, such as lobulation and border definition, which are important predictive and prognostic features in lung cancer [15, 92, 93].

Features extracted from RG and RD maps were first used in a computer-aided detection (CAD_e) system for eliminating false positive pulmonary nodule candidates on chest X-ray [94]. Messay *et al.* [95] used these features in a computer-aided diagnosis system (CAD_x) to discriminate between benign and malignant nodules. In another study from this group, RG and RD image features were utilized to optimize free parameters of a CT pulmonary nodule segmentation system [96]. As such, the goal of this study was to extract features from RG and RD maps from pre-surgical contrast-enhanced thoracic CT scans among patients with lung adenocarcinomas and assess whether these features were associated with overall survival. Additionally, potential biological underpinnings of these features were explored by analyzing the association between RG and RD image features with semantic radiological features.

3.2 Materials and Methods

3.2.1 Lung Cancer Patients

This retrospective study was approved by the Institutional Review Boards at the University of South Florida and Maastricht University Medical Center. There were two separate cohorts used in this study that have been described elsewhere [14]. Briefly,

the training cohort included 61 patients from the H. Lee Moffitt Cancer Center (MCC) Research Institute, Tampa, Florida and the test cohort included 47 patients from the Maastricht Radiation Oncology Clinic (MAASTRO), Maastricht, Netherlands. All patients were diagnosed with lung adenocarcinoma and underwent surgical resection as first course of therapy. Pre-treatment contrast enhanced CT scans were acquired within two months prior to surgery. Both cohorts included diagnostic pre-treatment contrast-enhanced CT scans acquired between 2006 and 2009 and clinical data including demographics, histology, stage, and vital status information. Follow-up for vital status information occurs annually through passive and active methods.

3.2.2 Patient Data

For the training cohort, clinical data were obtained from Moffitt’s Cancer Registry, which abstracts self-reported patient data and clinical information from patient medical records. Follow-up information for vital status occurs annually through passive and active methods. For this analysis, vital status was updated for the Moffitt patients since the previously published report [14]. Pathological TNM staging was utilized when available and clinical stage was used if pathological staging was unknown. Smoking status was categorized as ever smoker (current or former smoker) or never smoker. Similar data were abstracted and databased from MAASTRO for the test cohort patients.

3.2.3 Tumor Segmentation

All tumors were segmented using an in-house single-click ensemble segmentation algorithm on the Lung Tumor Analysis (LuTA) software program platform (Definiens Developer XD©, Munich, Germany) [97]. After applying the single click approach, the tumor delineations were inspected and edited if needed by a resident expert radiologist. The lung and tumor mask images obtained from LuTA software program were then imported into MATLAB® (Mathworks, Natick, MA) for image feature extraction as

described below.

3.2.4 Radial Gradient and Radial Deviation Maps and Features

Development of the RG and RD image features has been previously described [94,95]. In our study, after the tumors were segmented and center of mass of the tumor was automatically detected, 48 features were extracted from the RG and RD maps (Table A.1) bounded by different masks (described below) which were derived from tumor delineation masks using morphological operations. Since there were variations in image acquisition parameters, tri-linear interpolation was performed by a factor of two for scans acquired with a slice thickness of ≥ 4 mm on the z-axis to create homogeneous spacing between scans. Additionally, pixels were interpolated tri-linearly in x and y directions to 2.50 mm x 2.50 mm.

The masks used were ‘tumor mask’, ‘border mask’, ‘core mask’, and ‘outside mask’. The tumor mask was the region that was delineated semi-automatically using Definiens Developer XD© software (Definiens, Inc., Cambridge, MA). The border mask is a doughnut-shaped region that is created by subtracting the two masks which are formed by a dilation operation followed by an erosion operation on the tumor mask. The region obtained after the erosion operation is the core mask. Structural elements radii used for dilation and erosion morphological operations were 7.5 mm and 12.5 mm for smaller tumors (major axis length [2D] < 100 mm) and 10.0 mm and 15.0 mm for larger sized tumors (major axis length [2D] ≥ 100 mm), respectively. The outside mask is created by implementing dilation to the tumor mask followed by the subtraction of the tumor mask from the dilated region. The structural element used for the dilation morphological operation was 17.5 mm pixels for smaller tumors and 22.5 mm for larger sized tumors, respectively (Figure 3.1). All masks were additionally bounded to the lung parenchyma mask so that the VOIs did not exceed outside of the lung parenchyma. The features were created using the four masks for both 2-dimensional (2D) and 3-dimensional (3D). All 2D features were computed on the slice which included the center of mass of the segmentations. The 3D features features were

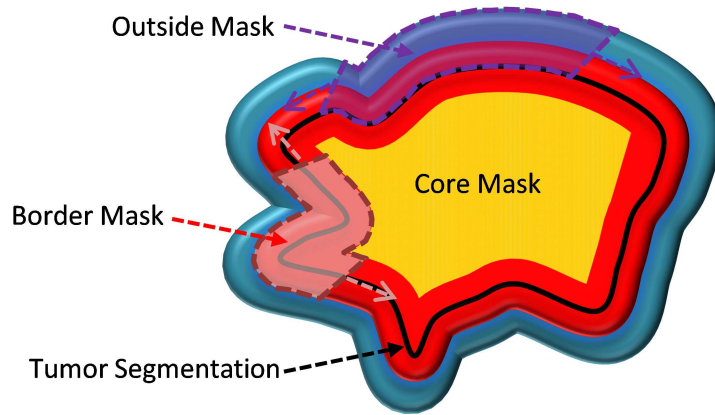


Figure 3.1 Cartoon image of the four tumor masks. The region inside the black line is the tumor mask, the orange area is the core mask, the red area is the border mask, and the region outside the black line is the outside mask (combination of the half part of red region and whole blue region).

a natural extension of the 2D features and were calculated after the first and the last slices of the tumor segmentations were removed in order to reduce the partial volume effects. The separation features were calculated as the difference divided by the sum inside the analyzed masks (e.g., the RD mean outside-border separation feature is the difference of outside RD mean and the border RD mean divided by the sum of the two).

3.2.5 Elimination of Non-Reproducible and Redundant Features

To eliminate the non-reproducible features the RIDER dataset [20] was used and the concordance correlation coefficient (CCC) between test-re-test scans were calculated. The RIDER dataset is a National Cancer Institute (NCI) sponsored project for the guidance of integrating quantitative features across different institutions. The dataset is publicly available in National Biomedical Imaging Archive [20]. A total of 31 patients with unenhanced test-retest chest CT scans were acquired within 15 minutes of each other. The CCCs were calculated to quantify the reproducibility between consecutive scans for patients. The CCC values range from 1 to -1, where 1 indicates a perfect correlation between two variables. Also, whenever two or more features had a Pearson's correlation coefficient > 0.80 , the feature(s) with higher absolute column-wise correlation mean were eliminated.

3.2.6 Radiological Semantic Features

Radiological semantic features for the training cohort were extracted by a clinical radiologist (with more than 7 years of experience) who was blinded to survival status and RD/RG status of the patients. Table 3.1 contains the 13 radiological semantic features that were extracted from the CT scans of the subjects. Briefly, in terms of morphological characteristics, the presence or absence of fissure attachment (defined as a tumor that attaches to the fissure; tumor’s margin is obscured by the margin), pleural attachment (defined as tumor attaches to the pleura other than fissure; tumor’s margin is obscured by the pleura), lobulation, concavity, air bronchogram, calcification, attachment to vessel, and pleural retraction were assessed. The following features were also evaluated, which have been defined elsewhere: dominant attenuation pattern [98], shape, border definition, spiculation [99,100], bubble-like lucency and cavitation [100].

Table 3.1
Semantic features analyzed.

No. Features	Values
1 Fissure attachment	0: absence 1: presence
2 Pleural attachment	0: absence 1: presence
3 Bubble-like lucency / cavitation	0: absence 1: presence
4 Air bronchogram	0: absence 1: presence
5 Calcification	0: absence 1: presence
6 Attachment to vessel	0: absence 1: presence
7 Pleural retraction	0: absence 1: presence
8 Lobulation	0: absence 1: presence
9 Spiculation	0: absence 1: presence
10 Concavity	0: absence 1: presence
11 Attenuation pattern ¹	1: GGO 2: part-solid 3: solid
12 Shape	1: round/oval 2: irregular
13 Border definition	1: well-defined 2: neither 1 or 3 3: poorly defined

Abbreviations: GGO = ground glass opacity.

¹None of the tumors were GGOs.

3.2.7 Statistical Analyses

All statistical analyses were performed using Stata/MP 14.2 (StataCorp LP, College Station, TX). All image features were dichotomized at their median value. Differences in image-based features by demographic features and semantic features were tested using Fisher’s exact test for categorical variables and Student’s t-test for the continuous variables. Survival analyses were performed using Cox proportional hazard regression, Kaplan-Meier curves, and log-rank tests. All survival data were right-censored at 5 years. To reduce the number of features and covariates to a single parsimonious model, a backward elimination approach was applied. The features that were identified as the most informative in the training cohort were then tested in the test cohort.

3.3 Results

3.3.1 Patient Demographics

Among 61 patients in the training cohort, 50.8% were male, 67.2% were aged above 65 years at the date of diagnosis, and 72.1% were either stage I or II. Among the image acquisition parameters, 93.4% of the CT scans were acquired with 120 kVp, 34.4% used B41f as a convolution kernel, 76.9% had an interpolated slice thickness of 2.5 mm, and 34.4% had a pixel resolution ≥ 0.7785 (third quartile). The median time to event (i.e., OS) was 33.5 months for this cohort (Table 3.2).

In test cohort, there were a total of 47 patients of which 53.2% were male, 46.8% were aged above 65 years at the date of diagnosis and 68.1% were either stage I or II. Among the image acquisition parameters, 85.1% of the CT scans were acquired with 120 kVp, 48.9% used A or B as a convolution kernel, 61.7% had an interpolated slice thickness of 2.5 mm, and 78.7% had a pixel resolution ≥ 0.7785 (third quartile). The median time to event (i.e., OS) was 32.0 months for this cohort (Table 3.2).

3.3.2 Univariable Analyses

After eliminating the redundant ($n = 15$) and non-reproducible features ($n = 16$), the log-rank p-values for the remaining 17 features were calculated (Table A.1). Out of these 17 features, two features (RD outside-border separation standard deviation (SD) and RG outside-border separation SD 2D) were significantly associated with OS, (log-rank p-value ≤ 0.05) and three features (RG border SD , RG outside-tumor separation mean, and RD tumor SD) were marginally (log-rank p-value ≤ 0.1) associated with OS (Table 3.3). The log-rank p-values for all 17 features assessed are presented in Table A.2.

The Kaplan-Meier survival curves with a median cutoff for the five features are presented in Figure 3.2a-e and the 5-year survival rates are presented in Table A.3. For the two features that were significantly associated with OS, tumors with high (\geq median) RD outside-border separation SD (Figure 3.2c, Hazard Ratio [HR] = 0.36; 95% CI 0.16-0.81, $p = 0.013$) and radial gradient outside-border separation SD (Figure 3.2d, HR = 0.43; 95% CI 0.20-0.94, $p = 0.035$) were associated with improved OS (Table 3.3). For the three features which were marginally significant to OS, tumors with high RG border SD (Figure 3.2a, HR = 1.92; 95% CI 0.90-4.11, $p = 0.092$) and RD tumor SD (Figure 3.2e, HR = 2.00; 95% CI 0.92-4.34, $p = 0.078$) were associated with poor OS while high RG outside-tumor separation mean was associated with improved OS (Figure 3.2b, HR = 0.48; 95% CI 0.22-1.06, $p = 0.068$).

In an exploratory analysis, low correlation was found between the features that were previously published in these cohorts (entropy ratio and convexity) [14] and RG/RD features used in this analysis (Pearson correlation coefficient < 0.35 for all features). Hence, the RG and RD features provide orthogonal information to other prognostic features.

Table 3.2
Patient characteristics in the training and test cohorts.

Characteristic	Training Cohort	Test Cohort
Number of patients	61	47
Age at diagnosis, N (%)		
< 65	20 (32.8)	25 (53.2)
≥ 65	41 (67.2)	22 (46.8)
Gender, N (%)		
Female	30 (49.2)	22 (46.8)
Male	31 (50.8)	25 (53.2)
Stage, N (%)		
I and II	44 (72.1)	32 (68.1)
III and IV	17 (27.9)	15 (31.9)
Tumor volume, mean cm³ (SD)	19.5 (29.0)	52.4 (130.0)
Tumor max diameter, mean mm (SD)	31.6 (13.8)	38 (21.5)
Overall survival, median months, (SD)	33.5	32
Image acquisition parameters		
Voltage, kVp, N (%)		
120	57 (93.4)	40 (85.1)
130 or 140	4 (6.6)	7 (14.9)
Convolution kernel, N (%)		
A,B	0 (0)	23 (48.9)
B30s,B60f,B70s	2 (3.3)	5 (10.7)
B30f	8 (13.1)	0 (0)
B40f	19 (31.2)	15 (31.9)
B41f	21 (34.4)	0 (0)
Other	11 (18.0)	4 (8.5)
Interpolated slice thickness , N (%)		
1.5 mm	0 (0)	2 (4.3)
2.0 mm	8 (13.1)	13 (27.7)
2.5 mm	40 (65.6)	29 (61.7)
3.0 mm	13 (21.3)	3 (6.3)
Pixel resolution (mm), tertiles, N (%)		
< 0.6926	20 (32.8)	6 (12.8)
≥ 0.6926 to < 0.7785	20 (32.8)	4 (8.5)
≥ 0.7785	21 (34.4)	37 (78.7)

Abbreviations: SD = standard deviation.

Table 3.3

Log-rank tests and Cox proportional hazards models for overall survival in the training cohort.

Covariate	Training cohort (N = 61)							
	Log-Rank P-value ¹	Univariable Model ² OR (95% CI)	P-value	Multivariable Model ³ OR (95% CI)	P-value	Multivariable Model ⁴ OR (95% CI)	P-value	
RG Border SD	0.084	1.92 (0.90 - 4.11)	0.092	
RG Outside-Tumor Mean	0.061	0.48 (0.22 - 1.06)	0.068	0.29 (0.12- 0.66)	0.003	0.31 (0.13 - 0.72)	0.006	
RD Outside-Border SD	0.009	0.36 (0.16 - 0.81)	0.013	0.25 (0.11 - 0.58)	0.001	0.24 (0.10 - 0.58)	0.001	
RG Outside-Border SD (2D)	0.029	0.43 (0.20 - 0.94)	0.035	
RD Tumor SD	0.071	2.00 (0.92 - 4.34)	0.078	
Age	0.439	1.38 (0.60 - 3.16)	0.444	.	.	0.83 (0.34 - 2.05)	0.690	
Gender	0.694	1.16 (0.54 - 2.49)	0.696	.	.	1.05 (0.47 - 2.35)	0.906	
Stage	0.085	1.95 (0.90 - 4.23)	0.093	.	.	2.14 (0.91 - 5.03)	0.082	
Tumor Volume	0.044	2.23 (1.00 - 4.97)	0.051	

Abbreviations: SD = Standard Deviation; OR = odd ratio; CI = confidence interval; RD = Radial deviation; RG = Radial gradient

Bold values are statistically significant.

¹Log-rank p-value for each covariate for overall survival right censored at 5-years. The radiomic features were dichotomized at the median value and the clinical covariates were dichotomized based on Table 1. The univariable analyses were based on 62 patients. But, due to missing patient data (age and gender), the total sample size for the multivariable analyses was 61 patients.²The independent main effect ORs for each covariate³The ORs for the two imaging features in a single model following backward elimination that considered all features and tumor volume.⁴The ORs for both imaging features identified from backward elimination adjusted for clinical covariates.

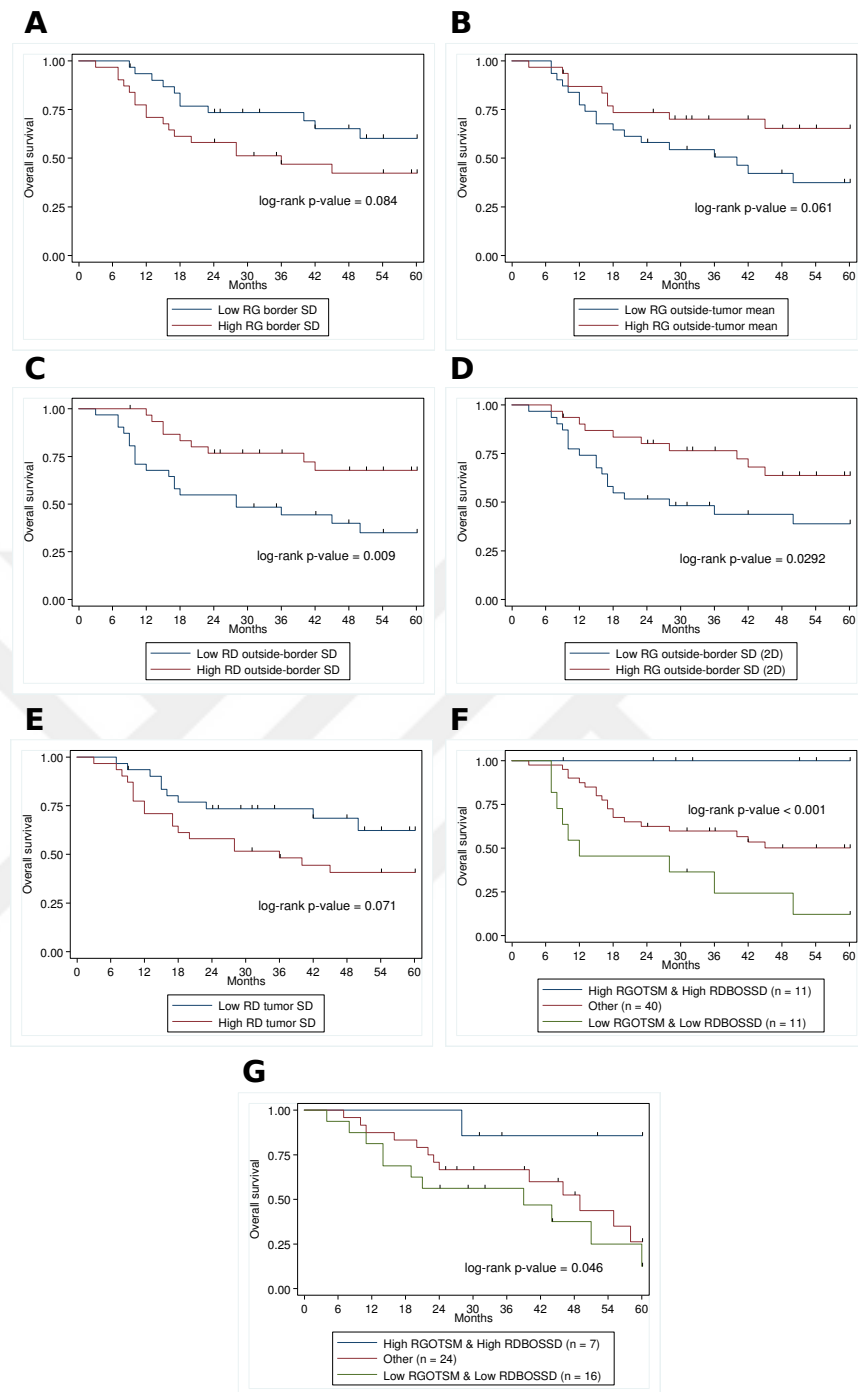


Figure 3.2 Kaplan-Meier survival curves for the following features **a)** Radial gradient border standard deviation in the training cohort, **b)** Radial gradient outside-tumor separation mean in the training cohort, **c)** Radial deviation outside-border separation standard deviation in the training cohort, **d)** Radial gradient outside-border separation standard deviation (2D) in the training cohort, **e)** Radial deviation tumor standard deviation in the training cohort, **f)** For the combination of radial gradient outside-tumor separation mean (RGOTSM) and radial deviation outside-border separation standard deviation (RDOBSSD) features. Hazard ratio with 95% confidence interval is calculated for the entire group instead of subgroups. HR = 3.65; 95% CI (1.89 – 7.05) **g)** For the combination of radial gradient outside-tumor separation mean (RGOTSM) and radial deviation outside-border separation standard deviation (RDOBSSD) features in the test cohort.

3.3.3 Multivariable Analyses

To reduce the number of image features to the most meaningful subset associated with overall survival, a stepwise backward elimination model using a threshold of 0.01 was applied to identify a parsimonious model. All five features and tumor volume were considered for the inclusion in the final model. The two features that remained in a feature-only multivariable model were RD outside-border separation SD (HR = 0.25; 95% CI 0.11 - 0.58, $p = 0.001$) and RG outside-tumor separation mean (HR = 0.29; 95% CI 0.12-0.66, $p = 0.003$). To control for potential confounding, these two features were included in multivariable Cox regression model that included age, gender, and stage; and both features remained significantly associated with OS (Table 3.3). These two features were further analyzed by demographics and imaging parameters and contingency tables were pulled out. In the training cohort, none of the demographics or imaging parameters were significantly associated with RD outside-border separation SD and RG outside-tumor separation mean (Table 3.4).

To determine if the findings could be replicated in an external cohort, these two features were analyzed in a test cohort (Table 3.5) using the median threshold values obtained from the training cohort and found that RD outside-border separation SD was significantly associated with OS (Figure A.1b, HR = 0.36; 95% CI 0.16-0.81, $p = 0.014$) but RG outside-tumor separation mean was not found to be significant (Figure A.1a, HR = 0.75; 95% CI 0.28-2.03 $p = 0.575$). However, for both features, the point estimates were inversely associated with risk of death. When these two features were included in multivariable Cox model that included age, gender and stage; RD outside-border separation SD was significantly associated with OS (HR = 0.40; 95% CI 0.17-0.97, $p = 0.042$) along with age (HR = 2.65; 95% CI 1.07-6.60, $p = 0.035$) and stage (HR = 3.35; 95% CI 1.34-8.36, $p = 0.010$) (Table 3.5). Additionally, among early stage lung cancer patients (stage I and II), RD outside-border separation SD was found to be significantly associated with OS in the training cohort (log-rank p -value = 0.031) and marginally significant in the test cohort (log-rank p -value = 0.097). None of the patient demographics or imaging parameters were significantly associated with RG outside-tumor separation mean and RD outside-border separation SD (except

pixel resolution significantly associated with RG outside-tumor separation mean in test cohort ($p = 0.010$, Table 3.6).

3.3.4 Combinatorial Analyses

In exploratory analyses, we assessed the combinatorial effects of RD outside-border separation SD and RG outside-tumor separation in the training and test cohorts (Figure 3.2f-g and Table A.3). In both cohorts patients who had high values ($>$ median) of both features also had significantly better OS compared to patients who had low values (\leq median) of both features. The subset of early stage patients (stage I and II) were also explored and the combinatorial effect was also found to be significantly associated with OS in the training cohort (log-rank P-value = 0.020). Although the survival pattern was similar in the test cohort, it did not reach statistical significance (log-rank p-value = 0.19).

3.3.5 Associations with Semantic Radiological Features

Three RD/RG radiomic features were significantly associated with three semantic features (Table 3.7): Lobulation, pleural attachment, and border definition. Importantly, the replicated feature was significantly associated with border definition. Specifically, cancers with a well-defined border were significantly more likely to have high ($>$ median) RD outside-border separation. These analyses were restricted to the training cohort only.

3.4 Discussion

In this study extracted RG and RD image features were extracted to determine whether they are associated with lung cancer patient survival. Of the 48 features that were analyzed, 31 features were eliminated because they were not reproducible or

Table 3.4
Demographics and imaging parameters by image dichotomized features in training cohort.

Covariate	Radial gradient outside-tumor separation mean			Radial deviation outside-border separation SD		
	LOW	HIGH	P- Value	LOW	HIGH	P- Value
Sex, N (%)						
Female	14 (45.2)	16 (53.3)		13 (43.3)	17 (54.8)	
Male	17 (54.8)	14 (46.7)	0.612	17 (56.7)	14 (45.2)	0.446
Age, N (%)						
< 65	8 (25.8)	12 (40.0)		8 (26.7)	12 (38.7)	
≥ 65	23 (74.2)	18 (60.0)	0.283	22 (73.3)	19 (61.3)	0.416
Stage, N (%)						
I/II	21 (67.8)	23 (77.4)		23 (77.4)	21 (67.7)	
III/IV	10 (32.2)	7 (22.6)	0.570	7 (22.6)	10 (32.3)	0.570
5- year survival, %	37.40%	65.30%	0.061	34.90%	67.70%	0.009
Voltage, KVp, N (%)						
120	28 (90.3)	29 (96.7)		26 (86.7)	31 (100.0)	
130 or 140	3 (9.7)	1 (3.3)	0.612	4 (13.3)	0 (0)	0.053
Convolution Kernel, N (%)						
A,B	0 (0)	0 (0)		0 (0)	0 (0)	
B30s,B60f,B70s	2 (6.7)	0(0)		1(3.2)	1(3.3)	
B30f	6 (20.0)	2 (6.4)		3 (9.7)	5 (16.7)	
B40f	7 (23.3)	12 (38.7)		12 (38.7)	7 (23.3)	
B41f	10 (33.3)	11 (35.5)		9 (29.0)	12 (40.0)	
Other	5 (16.7)	6 (19.4)	0.270	6 (19.4)	5 (16.7)	0.700
Interpolated Slice Thickness, N (%)						
1.5 mm	0 (0)	0 (0)		0 (0)	0 (0)	
2.0 mm	5 (16.1)	3 (10.0)		3 (10.0)	5 (16.1)	
2.5 mm	17 (54.8)	23 (76.7)		20 (66.7)	20 (64.5)	
3.0 mm	9 (29.1)	4 (13.3)	0.189	7 (23.3)	6 (19.4)	0.861
Pixel Resolution, tertiles N (%)						
< 0.6926 mm	7 (22.6)	13 (43.3)		7 (23.3)	13 (41.9)	
≥ 0.6926 and < 0.7785 mm	10 (32.3)	10 (33.3)		13 (43.3)	7 (22.6)	
> 0.7785 mm	14 (45.1)	7 (23.4)	0.146	10 (33.4)	11 (35.5)	0.172

Bold values are statistically significant.

Table 3.5
Cox proportional hazards models for overall survival in the test cohort.

Covariate	Test cohort (N = 47)			
	Multivariable Model ¹ OR (95% CI)	P-value	Multivariable Model ² OR (95% CI)	P-value
RG Border SD
RG Outside-Tumor Mean	0.75 (0.28 - 2.03)	0.575	0.48 (0.17 - 1.37)	0.172
RD Outside-Border SD	0.36 (0.16 - 0.81)	0.014	0.40 (0.17 - 0.97)	0.042
RG Outside-Border SD (2D)
RD Tumor SD
Age	.	.	2.65 (1.07 – 6.60)	0.035
Gender	.	.	1.43 (0.53 – 3.82)	0.476
Stage	.	.	3.35 (1.34 – 8.36)	0.010
Tumor Volume

Abbreviations: SD = Standard Deviation; OR = odd ratio; CI = confidence interval; RD = Radial deviation; RG = Radial gradient

Bold values are statistically significant.

¹The ORs for from the two imaging features identified in training cohort from backward elimination

²The ORs for both imaging features identified from backward elimination in training cohort adjusted for clinical covariates

they were redundant. The remaining 17 features were subjected to statistical analysis resulting in a parsimonious model containing two highly informative features associated with lung cancer survival. One of the two features (RD outside-border separation SD) was replicated and found to be significantly associated with OS in a separate external cohort (test cohort) of lung cancer patients.

Radiomics is motivated by the premise that quantitative image features reflect the underlying pathophysiology of tumors. In Figure 3.3 VOIs and corresponding RD maps for two patients with substantially different clinical outcomes are presented. The patient (Figure 3.3a) with short survival was deceased after 9 months and had a low (< median) RD outside-border separation SD value while the second patient (Figure 3.3b) was still alive after 60 months had a high (> median) RD outside-border separation SD value.

Table 3.6
Demographics and imaging parameters by image dichotomized features in test cohort.

Covariate	Radial gradient outside-tumor separation mean			Radial deviation outside-border separation SD		
	LOW	HIGH	P- Value	LOW	HIGH	P- Value
Sex, N (%)						
Female	17 (48.6)	5 (41.7)		8 (38.1)	14 (53.9)	
Male	18 (51.4)	7 (58.3)	0.747	13 (61.9)	12 (46.1)	0.381
Age, N (%)						
< 65	21 (60.0)	4 (33.3)		11 (52.4)	14 (53.9)	
≥ 65	14 (40.0)	8 (66.7)	0.180	10 (47.6)	12 (46.1)	1.000
Stage, N (%)						
I/II	25 (71.4)	7 (58.3)		11 (52.4)	21 (80.8)	
III/IV	10 (28.6)	5 (41.7)	0.481	10 (47.6)	5 (19.2)	0.059
5- year survival, %	0.23	0.556	0.347	10.90%	47.40%	0.007
Voltage, KVP, N (%)						
120	29 (82.9)	11 (91.7)		17 (81.0)	23 (88.5)	
130 or 140	6 (17.1)	1 (8.3)	0.659	4 (19.0)	3 (11.5)	0.684
Convolution Kernel, N (%)						
A,B	18 (51.4)	5 (41.7)		9 (42.9)	14 (53.8)	
B30s,B60f,B70s	4 (11.4)	1 (8.3)		3 (14.3)	2 (7.7)	
B30f	0 (0.0)	0(0.0)		0 (0.0)	0 (0.0)	
B40f	10 (28.6)	5 (41.7)		7 (33.3)	8 (30.8)	
B41f	0 (0.0)	0(0.0)		0(0.0)	0(0.0)	
Other	3 (8.6)	1 (8.3)	0.905	2 (9.5)	2 (7.7)	0.877
Interpolated Slice Thickness, N (%)						
1.5 mm	2 (5.7)	0 (0.0)		2 (9.5)	0 (0.0)	
2.0 mm	12 (34.3)	1 (8.3)		7 (33.3)	6 (23.1)	
2.5 mm	19 (54.3)	10 (83.4)		11 (52.4)	18 (69.2)	
3.0 mm	2 (5.7)	1 (8.3)	0.219	1 (4.8)	2 (7.7)	0.381
Pixel Resolution, tertiles N (%)						
< 0.6926 mm	3 (8.6)	3 (25.0)		2 (9.5)	4 (15.4)	
≥ 0.6926 and < 0.7785 mm	1 (2.9)	3 (25.0)		2 (9.5)	2 (7.7)	
> 0.7785 mm	31 (88.5)	6 (50.0)	0.010	17 (81.0)	20 (76.9)	0.877

Bold values are statistically significant.

Table 3.7

Association between semantic features and radial gradient and radial deviation features.

Feature name	Semantic Feature			P- Value
	<i>Lobulation</i>			
	Absent	Present		
Radial gradient outside-border separation SD (2D), N (%)				
LOW	27 (60.0)	4 (23.5)		0.021
HIGH	18 (40.0)	13 (76.5)		
Feature name	<i>Pleural attachment</i>			P- Value
	Absent	Present		
Radial gradient border SD, N (%)				
LOW	28 (62.2)	3 (17.7)		0.004
HIGH	17 (37.8)	14 (82.3)		
Feature name	<i>Border definition</i>			P- Value
	Well Defined	Poorly Defined	Other ¹	
Radial gradient border SD, N (%)				
LOW	13 (81.3)	8 (42.1)	10 (37.0)	0.015
HIGH	3 (18.7)	11 (57.9)	17 (63.0)	
Radial deviation outside-border separation SD², N (%)				
LOW	4 (25.0)	9 (47.4)	18 (66.7)	0.029
HIGH	12 (75.0)	10 (52.6)	9 (33.3)	
Radial gradient outside-border separation SD (2D), N (%)				
LOW	4 (25.0)	14 (73.7)	13 (48.2)	0.018
HIGH	12 (75.0)	5 (26.3)	14 (51.8)	

Bold values are statistically significant.

¹Tumor with neither a well or poorly-defined border.²This feature was replicated and found to be statistically significantly associated with survival in both the training cohort and test cohort

ration SD value. In the original CT-image, both patients have similarly-sized tumors that are speculated; however, the VOI for each RD image have considerably different RD map appearances. By quantifying and analyzing these differences, as performed in this study, RD/RG features were shown to have potential clinical utility by differentiating patients with an aggressive disease and poor patient outcomes versus patients with more indolent disease and improved outcomes. Additionally, by analyzing the correlations of RD/RG features with semantic radiological features, we may have revealed their potential biological underpinnings. Specifically, we found three RG and

RD features that were significantly associated with tumor lobulation, pleural attachment, and border definition (Table 3.7). The replicated feature, which was associated with lung cancer survival in both cohorts, was significantly associated with border definition that has been previously reported to be a prognostic factor in lung cancer [15]. In the current analysis, patients who had well defined border definition were significantly associated with high RD outside-border separation SD. As such, these analyses suggest that RG and RD features may be capturing clinically and biologically relevant radiological information of lung cancer tumors.

Radial image-based features have been previously applied in chest CT CAD systems to; discriminate benign and malignant nodules, and optimize free parameters of tumor segmentation [95,96]. However, in these previous studies means and SDs were calculated from two different masks (region inside tumor and region outside tumor) on the RG and RD maps. By contrast, the means and standard deviations were calculated from four different masks (tumor mask, border mask, core mask, and outside mask) in this study. To the best of our knowledge, the current study is the first to analyze RD/RG features for their association with lung cancer survival and their association with radiological semantic features.

In this study, unique and new features were extracted and analyzed from training and test cohorts originally published by Grove *et al.* [14]. Grove *et al.* reported that convexity and entropy ratio features were significantly associated with OS in the training cohort. Thus, a model that included the convexity and entropy ratio features, RD/RG features, and patient characteristics were explored and was found that RD outside-border separation SD (HR = 0.21), RG outside-tumor separation mean (HR = 0.21), and entropy ratio (HR = 3.28) were significantly associated with OS in the training cohort. However, when the remaining three features were analyzed in the test cohort, only RD outside-border separation SD (HR = 0.34) was found to be significant.

Radiomics have the potential to complement and improve current precision medicine. Limitations of tumor-based biomarkers are: they can be subjective to sampling bias due to the heterogeneous nature of tumors, the requirement of tumor speci-

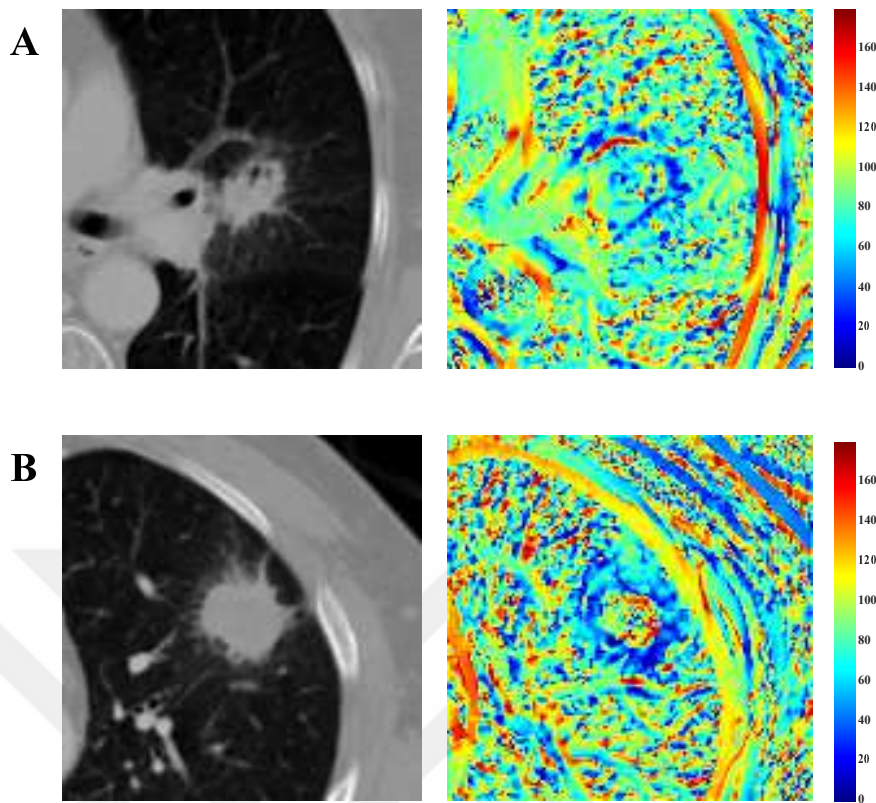


Figure 3.3 Volume of interests (VOI) for two lung cancer patients with extreme differences in clinical outcomes. Radial deviation image features for the corresponding VOIs for these lung cancer patients. **a)** A tumor of a patient (Patient ID [PID]: 33) who deceased after 9 months. **b)** A tumor of a patient (PID: 75) with who had an ongoing survival after 60 months.

mens for biomarker testing, where the assays can be timely and expensive [12, 14, 101]. In contrast, radiomic features can be extracted in real-time from standard-of-care images, do not require timely and often expensive laboratory testing, are not subject to sampling bias and artifact, and are non-invasive. Importantly, radiomic analyses do not subject patients to additional radiation exposure since standard-of-care images are utilized, and radiomic features represent the phenotype of the entire tumor in 3D and not just the portion that was subjected to biomarker testing. Indeed, there is precedence that quantitative image features provide valuable and potentially translational information in lung cancer patient outcomes. Previous studies have shown that tumor shape and density are related to lung cancer survival [12, 14, 50, 101]. Additionally, as tumor shape becomes more eccentric, it has a higher probability of metastatic disease [14], and solid lesions that are differentiable from their outside environment and have high contrast edges tend to be less aggressive [101]. Furthermore, tumors that are connected to lung wall are also associated with poor prognosis [101, 102].

Some limitations and strengths are acknowledged to this study. First, an available dataset [14] of modestly sample sized training and test cohorts were utilized. However, the features that were extracted and analyzed were unique from the prior work [14] and, importantly an image feature was found to be significantly associated with OS in both the training and test cohorts. We applied a rigorous feature reduction approach to eliminate correlated and no-reproducible features, and we utilized a backward reduction approach to identify a single parsimonious model containing the most important features. It is acknowledged that in the combinatorial analyses there were limited numbers of patients in the subgroups and progression-free survival data for these cohorts were not available. A potential limitation of radiomic studies is the range of image acquisition parameters and modalities used which can make it difficult to standardize image features and limit the robustness of computer-extracted features [103]. In future studies, interpolation methods can be investigated to harmonize all data to a smaller range of slice thickness and pitch, which is hypothesized to reduce some acquisition associated variability. However, image acquisition parameters were not found to be associated with RD outside-border separation SD for either cohort (Table 3.4 and Table 3.6). Despite the fact that the study was limited to patients with lung adenocarcinoma, potential histological differences in the analyses were removed and it is believed that this study had numerous strengths that outweigh the potential limitations.

In conclusion, this study identified a RG and RD image feature that was significantly associated with lung cancer survival in both training and test cohorts even after adjusting for clinical covariates. The analyses also revealed a novel combinatorial association of two features which differentiates patients with aggressive disease versus patients with indolent disease, and this was replicated in the test cohort. As such, these findings may have clinical utility to sub-stratify patients based on clinical outcome and identify patients that may need more that aggressive treatment such as neo-adjuvant chemotherapy and aggressive follow-up and management. These features will require confirmation in additional studies and lung cancer patient cohorts.

4. NOVEL CLINICAL AND RADIOMIC PREDICTORS OF RAPID DISEASE PROGRESSION PHENOTYPES AMONG LUNG CANCER PATIENTS TREATED WITH IMMUNOTHERAPY

4.1 Introduction

Immune-checkpoint blockades targeting PD-1 or PD-L1 provide durable responses and improved long-term survival in advanced NSCLC patients [2–7]. However, overall response rates are only about 20-50% and those that do not respond can experience accelerated and lethal progression described as HPD [104]. Though PD-L1 IHC is a widely applied biomarker to select patients for immunotherapy, PD-L1 expression alone is not adequate to predict response [8,9]. Recently, a clinical trial demonstrated that immunotherapy combined with chemotherapy exhibits survival benefit regardless of PD-L1 expression [7]. Hence, additional biomarkers that are highly predictive of positive and negative responses to immune-checkpoint blockades are a significant unmet clinical need.

In this study, clinical data and CT scans of NSCLC patients treated on immunotherapy clinical trials were utilized to develop parsimonious models that identify patients that are at risk of rapid disease progression. Image-based feature (Radiomics) data were extracted to capture peritumoral and intratumoral heterogeneity reflecting the underlying pathophysiology of the ROI [12, 50, 105] that included the lung lesion and surrounding border region of the lung lesion (Figure 2.1). The rapid disease progression phenotypes that were analyzed based on time-to-progression (TTP) and tumor growth rates (TGRs).

4.2 Materials and Methods

4.2.1 Study Population and Patient Data

Based on patient eligibility (Figure 4.1), this study included 228 NSCLC patients that were prospectively enrolled into industry-sponsored clinical trials using PD-1 single agent (Nivolumab, Pembrolizumab), PD-L1 single agent (Durvalumab, Atezolizumab), or the combination of PD-L1 or PD-L1 with cytotoxic T-lymphocyte-associated protein 4 (Ipilimumab, Tremelimumab) as a second agent. All patients were treated between June 2011 and June 2016 at the Moffitt Cancer Center, Tampa, Florida. Other common inclusion criteria included: Patients who were diagnosed with histologically- or cytological-documented NSCLC with advanced/metastatic stage disease with at least one measurable lesion (≥ 10 mm), Eastern Cooperative Oncology Group (ECOG) Performance Status of 0 or 1, and provided written informed consent. Common exclusion criteria included: a concurrent medical condition requiring the use of immunosuppressive medications or immunosuppressive doses of systemic or absorbable topical corticosteroids, and presence of any active autoimmune disease. Patient data were obtained from Moffitt's Cancer Registry, Moffitt's Collaborative Data Services Core, and through manual abstraction from electronic medical records. Access to these data for retrospective analyses was approved by the University of South Florida Institutional Review Board.

Moffitt's Cancer Registry abstracts information from patient electronic medical records on demographics, history of smoking, stage, histology, treatment, and vital status. Follow-up for vital status occurs annually through active (i.e., chart review and directly contacting the patient, relatives, and other medical providers) and passive methods (i.e., mortality records). Hematology data were obtained from the CDCS and included: lactate dehydrogenase (LDH), serum albumin, lymphocytes, white blood cells, neutrophils, fibrinogen, and neutrophils to lymphocytes ratio (NLR). Manually abstracted data included: targeted mutations (*EGFR*, *KRAS*), body mass index, history of systemic treatment(s) for current lung cancer staging, ECOG performance status, RECIST diameters of target and non-target lesions, best overall responses,

number of metastatic sites (number of organs that have metastatic lesions), metastatic sites prior to treatment, Royal Marsden Hospital (RMH) prognostic score, and MD Anderson (MDA) risk score. The RMH prognostic score ranges from 0 to 3 and one point was added when serum albumin is < 3.5 (g/dl), one point was added when LDH $>$ upper limit normal (225 U/L), and one point was added when number of organ metastases > 2 [106]. The MDA risk model score ranges from 0 to 5 and one point was added when serum albumin was < 3.5 (g/dl), one point was added when LDH $>$ upper limit normal (225 U/L), one point was added when number of organ metastases > 2 , one point was added when a presence of a gastrointestinal tumor, and one point was added when ECOG > 0 [107].

4.2.2 CT Tumor Segmentation and Radiomic Feature Extraction

Pre-treatment contrast-enhanced thoracic CT scans were performed ≤ 30 days prior to the initiation of immunotherapy were utilized to extract quantitative image features. The CT imaging studies were acquired using two different manufacturers (General Electric Healthcare, Waukesha, Wisconsin and Siemens, Erlangen, Germany). About 95% of the CT scans were acquired with a slice thickness of 3 mm. The median pixel spacing was 0.80 mm by 0.80 mm and seven different convolution kernels were used to reconstruct the images with B41f as the most common (65.4%).

Patients who had a pre-treatment CT scan and a RECIST target or a non-target lung lesion were considered for radiomics feature extraction and analysis (Figure 4.1). Whenever a patient had a lung target lesion, the lung lesion with largest RECIST diameter based on the radiology report was selected for radiomic analysis. If a patient did not have a lung target lesion, the non-target lung lesion with the largest RECIST diameter was selected for analysis. Patients that did not meet either condition were excluded from all radiomic analysis. The detected lesions were delineated in 3D using an in-house single-click ensemble region growing segmentation algorithm on the Lung Tumor Analysis software program platform (Definiens Developer XD©, Munich, Germany) [97]. Following the single-click approach, tumor delineations were

either confirmed as is or edited by an expert radiologist with more than 10 years of clinical experience. The binary tumor mask images (i.e., tumor delineations) were then imported into in-house radiomic feature extraction toolboxes created in MATLAB® 2015b (The Mathworks Inc., Natick, Massachusetts) and C++ (<https://isocpp.org>).

A total of 600 radiomic features were extracted from the selected ROIs which included both the tumors and tumor border regions (Figure 3.1). Features based on shape, intensity and texture were extracted using algorithms from the Image Biomarker Standardization Initiative (IBSI) v5 [76]. The radial deviation and radial gradient [<https://github.com/TunaliIlke/RDRG-v2.1>] and wavelet features were defined elsewhere [51]. Prior to the extraction of radiomic features, all images were resampled to a single voxel spacing of 1mm x 1mm x 1mm to standardize spacing across all images. Texture features were extracted using a fixed bin width of 25 Hounsfield units (HU). All radiomic features were linearly rescaled between -1 and 1. Non-reproducible features were eliminated by calculating the CCC between test/re-test scans using the RIDER dataset [20]. The CCCs were calculated for each radiomic feature to quantify the reproducibility between consecutive scans and the features with a CCC < 0.80 were removed.

4.2.3 Rapid Disease Progression Phenotypes

Two rapid disease progression phenotypes were generated as dependent variables (Figure 4.2):

- (i) Patients who had a TTP < 2 months versus the patients who had a TTP \geq 2 months;
- (ii) Patients who had a HPD versus patients without a HPD (non-HPD);

where TTP was defined as time from initiation of immunotherapy to disease progression (by RECIST or clinical evaluation) excluding deaths. Patients in the TTP

≥ 2 months group included patients that either had a progressive disease (PD), stable disease (SD) or partial response (PR) on their last follow-up, while all patients in the $TTP < 2$ months group had a PD on their last follow-up. Cancer progression was confirmed by a second follow-up scan (to rule out pseudo-progression) or by clinical evaluation of the treating clinician. The tumor growth rate $TGR_{\text{pre-treatment}}$ was calculated using baseline scan (CT scan prior to initiation of therapy) and pre-baseline scan while the $TGR_{\text{treatment}}$ was calculated using the baseline scan and the first follow-up scan as defined by Champiat *et al.* [104]. The TGR rate was calculated from the difference of tumor volumes from consecutive CT scans: $TGR_{\text{pre-treatment}}$ and $TGR_{\text{treatment}}$ (Figure 4.3).

HPD patients were a subset of patients that exhibited PD and accelerated TGR < 2 months (Figure 4.3) and were defined based on a stringent set of criteria adapted from prior studies [104, 108, 109]:

- (i) Greater than a two-fold increase from $TGR_{\text{pre-treatment}}$ to $TGR_{\text{treatment}}$, and
- (ii) PD on first follow-up scan by RECIST (RECIST 1.1 or iRECIST) criteria, and
- (iii) Time-to-treatment failure (TTF) < 2 months.

4.2.4 Statistical Analyses

All statistical analyses were performed using Stata/MP 14.2 (StataCorp LP, College Station, Texas) and R Project for Statistical Computing version 3.4.3 (<http://www.r-project.org/>). Differences for the clinical covariates were tested using Fisher’s exact test for categorical variables and the Mann-Whitney’s rank sum test and analysis of variance test for continuous variables. Survival analyses were performed using Kaplan-Meier curves, and the log-rank tests. For the radiomics data, non-reproducible features were eliminated using test/re-test scans from the RIDER dataset [20].

Using logistic regression, a covariate reduction model building approach was

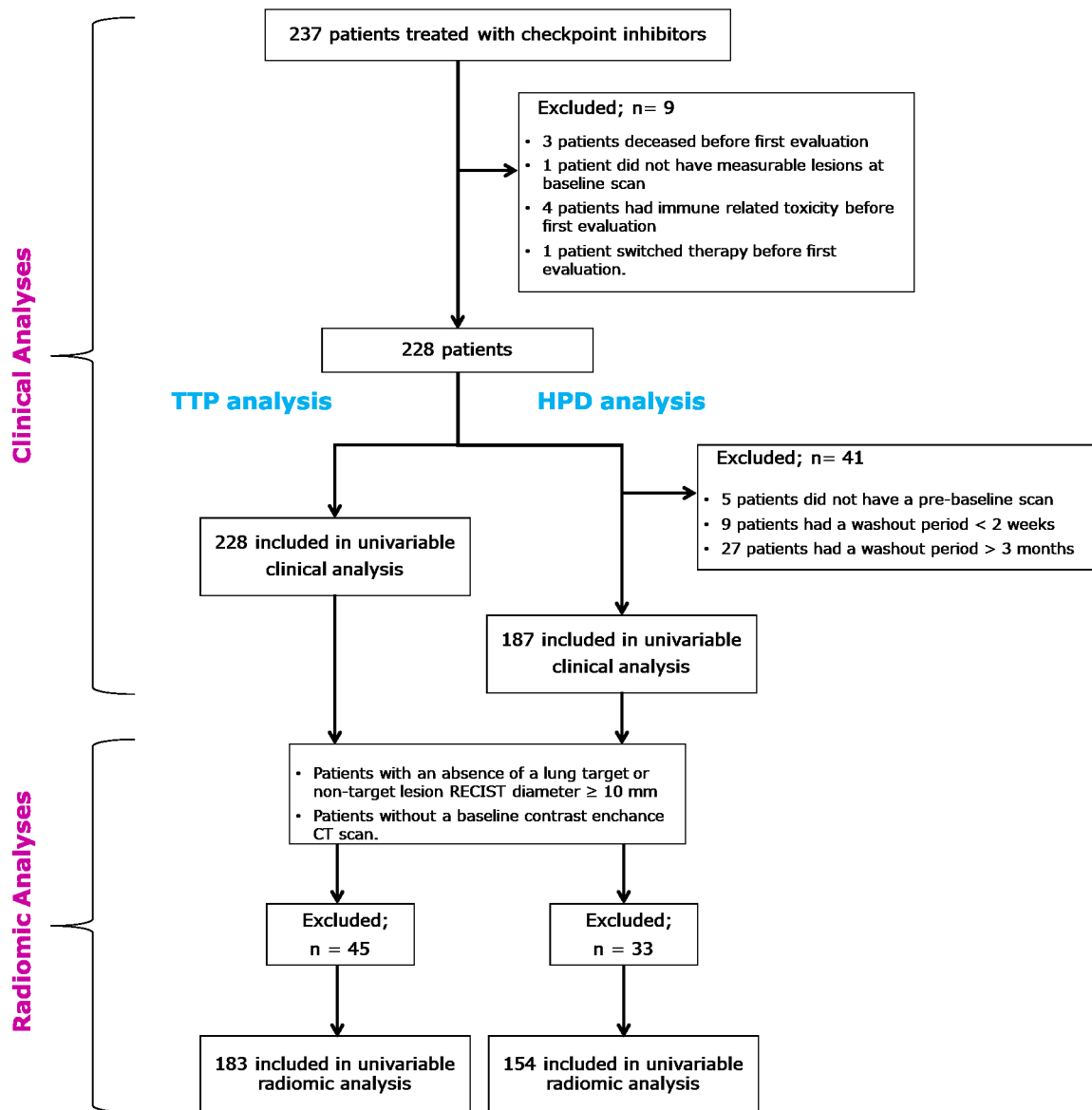


Figure 4.1 Overall survival and progression-free survival for the training and test cohorts.

employed to identify the most informative clinical covariates and radiomic features to predict rapid disease progression phenotypes. The AUROCs were calculated and used to compare the predictive performance of the various models. For the clinical covariates, we only considered covariates that were significantly different ($P < 0.05$) in univariable analysis. Then a stepwise backward elimination approach using a threshold of 0.05 was applied to produce a parsimonious model containing only clinical covariates. For the radiomic features, we only considered features that were significant ($P < 0.05$) or marginally significant in univariable analyses ($P < 0.10$). Any features that

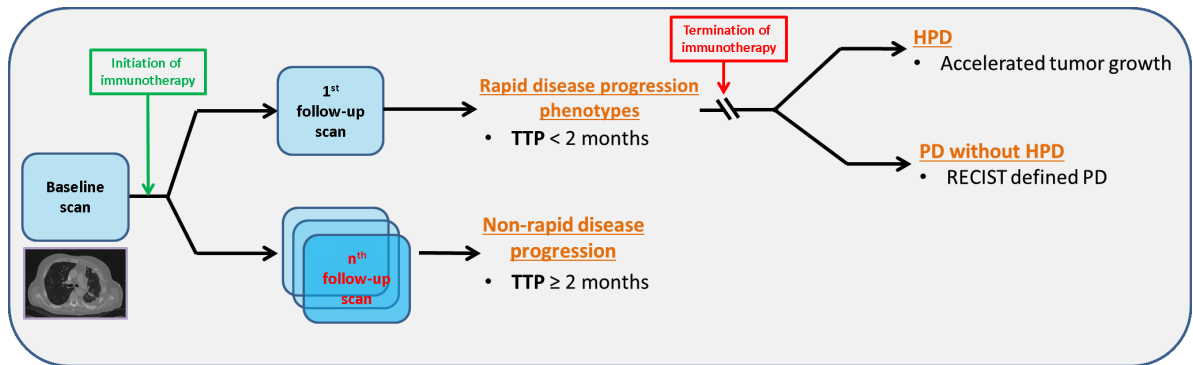


Figure 4.2 Response groups of immunotherapy patients. The non-responders were defined as patients who developed PD < 2 months (time-to-progression < 2 months). Patients who discontinued treatment and had an accelerated tumor growth were defined as HPD. The PD without HPD patients had PD on first follow-up and did not meet the HPD criteria. Responder patients were ones who developed had an ongoing response (PR/CR or SD) or developed PD at least 2 months after the initiation of therapy (TTP ≥ 2 months).

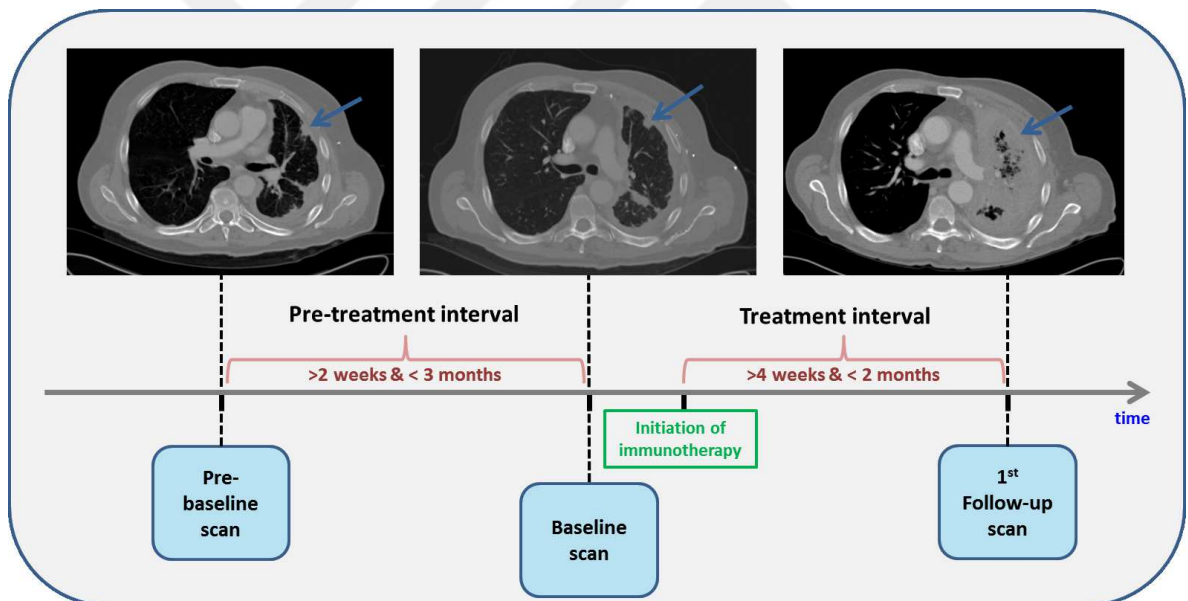


Figure 4.3 Example of an HPD patient that had SD prior to the initiation of immunotherapy but developed rapid tumor growth on first follow-up and experienced more than 2-fold increase from pre-treatment tumor growth versus treatment.

were correlated with tumor volume based on a Pearson's correlation coefficient ≥ 0.80 were removed. If two or more features were correlated based on an absolute Pearson's correlation coefficient ≥ 0.80 , the feature(s) with the higher column-wise absolute correlation mean(s) was/were removed. The remaining radiomic features were included in

a stepwise backward elimination approach using a threshold of 0.05 to produce a parsimonious model of only radiomic features. For rapid disease progression phenotypes, the most informative set of clinical covariates and the most informative set of radiomic features were combined into a single final model.

Individual probability for disease progression of final clinical-radiomic models were calculated and were used to group patients to assess progression-free-survival (PFS). Specifically, the individual probability for each patient was split into percentiles (median and tertiles) and novel cut-points were identified using Classification and Regression Tree (CART). CART is a non-parametric data-mining tool that can fragment data into subgroups by failure time data using [110] and classifies variables through a decision tree composed of binary splits where the split points are optimized based on impurity criterion. For the PFS analysis, an event was defined as death or either clinical or RECIST based progression of cancer and survival data were right-censored at 36 months.

When data are imbalance for case and controls, classification can be biased in favor of the majority class. To overcome this problem, SMOTE technique was applied [111] that increase the number of samples in our minority classes (TTP < 2 months and HPD groups) by creating *synthetic* examples. Five nearest neighbors were used as SMOTE parameter and minority classes were sub-sampled to be equal numbered with the majority samples.

4.3 Results

Patient demographics (Table 4.1), disease burden (Table 4.2), hematology results (Table 4.3) and mutational status results (Table 4.4) were compared for TTP < 2 months vs TTP \geq 2 months and HPD vs non-HPD. The full and parsimonious multivariable models for the clinical covariates are presented for TTP < 2 months vs TTP \geq 2 months (Table 4.5) and HPD vs non-HPD (Table 4.6). For the analysis of TTP < 2 months vs TTP \geq 2 months, the full clinical model included eight clinical

covariates with an AUROC of 0.770, whereas the parsimonious clinical model included four covariates with an AUROC 0.736. In multivariable analysis for HPD versus non-HPD, the full model included two clinical covariates with an AUROC of 0.783 whereas the parsimonious clinical model with one variable feature yielded an AUROC 0.712.

Table 4.1
Patient demographics by rapid disease progression phenotypes.

Characteristic	TTP < 2 months	TTP ≥ 2 months	P-Value	HPD	Non-HPD	P-Value
Total, N (%)	54 (23.7)	174 (76.3)		15 (8.0)	172 (92.0)	
Age at diagnosis, N (%)						
< 65	18 (33.3)	68 (39.1)		5 (33.3)	69 (40.1)	
≥ 65	36 (66.7)	106 (60.9)	0.521	10 (66.7)	103 (59.9)	0.785
Median, (95% CI)	67 (65-68)	67 (65-68)	0.992	67 (57-72)	66 (65-67)	0.834
Sex, N (%)						
Female	24 (44.4)	77 (44.2)		5 (33.3)	77 (44.8)	
Male	30 (55.6)	97 (55.8)	1.000	10 (66.7)	95 (55.2)	0.431
Stage, N (%)						
IIIb	1 (1.9)	7 (4.0)		0 (0)	6 (3.5)	
IV	53 (98.1)	167 (96.0)	0.684	15 (100)	166 (96.5)	1.000
Histology, N (%)						
Adenocarcinoma/Others	36 (66.7)	129 (74.1)		11 (73.3)	120 (69.8)	
Squamous cell carcinoma	18 (33.3)	45 (25.9)	0.299	4 (26.7)	52 (30.2)	1.000
Checkpoint inhibitors, N (%)						
Anti PD-L1	18 (33.3)	43 (24.7)		3 (20.0)	49 (28.5)	
Anti PD-1	18 (33.3)	54 (31.0)		8 (53.3)	50 (29.1)	
Doublet	18 (33.3)	77 (44.3)	0.300	4 (26.7)	73 (42.4)	0.192
ECOG performance status, N (%)						
0	9 (16.7)	43 (24.7)		3 (20.0)	37 (21.5)	
1	45 (83.3)	131 (75.3)	0.267	12 (80.0)	135 (78.5)	1.000
Smoking status						
Never smokers	6 (11.1)	30 (18.0)		1 (6.7)	32 (19.2)	
Ever smokers	48 (88.9)	137 (82.0)	0.292	14 (93.3)	135 (80.8)	0.313
BMI, median, (95% CI)	25 (23.1-26.3)	25.9 (24.9-27.0)	0.079	25.7 (21.2-26.3)	25.7 (24.8-27.0)	0.382

Abbreviations: HPD = hyperprogressive disease; TTP = time-to-progression; BMI = body mass index

Bold P-values are statistically significant

Smoking status was available for 221 patients.

Table 4.2
Disease burden characteristics of patients by rapid disease progression phenotypes.

Characteristic	TTP < 2 months	TTP ≥ 2 months	P-Value	HPD	Non-HPD	P-Value
History of radiation therapy₂						
No	36 (69.2)	135 (79.9)		14 (93.3)	129 (75.9)	
Yes	16 (30.8)	34 (20.1)	0.130	1 (6.7)	41 (24.1)	0.197
Previous lines of systemic treatment						
< 2	27 (50.0)	129 (74.1)		10 (66.7)	120 (69.8)	
≥ 2	27 (50.0)	45 (25.9)	0.001	5 (33.3)	52 (30.2)	0.777
Corticosteroids₃, N (%)						
No	44 (81.5)	149 (85.6)		12 (80.0)	147 (85.5)	
Yes	10 (18.5)	25 (14.4)	0.517	3 (20.0)	28 (14.5)	0.475
Hepatic metastasis, N (%)						
No	35 (64.8)	143 (82.2)		10 (66.7)	138 (80.2)	
Yes	19 (35.2)	31 (17.8)	0.013	5 (33.3)	34 (19.8)	0.316
Bone metastasis, N (%)						
No	31 (57.4)	130 (74.7)		9 (60.0)	126 (73.3)	
Yes	23 (42.6)	44 (25.3)	0.017	6 (40.0)	46 (26.7)	0.366
Brain metastasis, N (%)						
No	47 (87.0)	147 (84.5)		13 (86.7)	148 (86.1)	
Yes	7 (13.0)	27 (15.5)	0.827	2 (13.3)	24 (13.9)	1.000
Number of metastatic sites						
< 3	18 (33.3)	80 (46.0)		5 (33.3)	77 (44.8)	
≥ 3	36 (66.7)	94 (54.0)	0.117	10 (66.7)	95 (55.2)	0.431
RMH prognostic score						
< 2	26 (49.1)	110 (70.5)		4 (26.7)	107 (69.0)	
≥ 2	27 (50.9)	46 (29.5)	0.007	11 (73.3)	48 (31.0)	0.003
TGR_{pre-treatment} percentage, (95% CI)						
	13.4 (5.2-21.0)	14 (11.1-17.6)	0.390	17.1 (1.9-28.4)	13.7 (10.4-17.5)	0.796
RECIST tumor diameter (mm) at baseline, (95% CI)						
	91 (74.4-105.2)	77 (66.6-85.4)	0.175	86 (48.3-105.4)	81 (71.0-89.7)	0.974

Abbreviations: RMH = Royal Marsden Hospital; HPD = hyperprogressive disease; TTP = time-to-progression; TGR = tumor growth rate.

Bold P-values are statistically significant.

Prior systemic treatments were assessed < 6 months prior to the initiation of immunotherapy. Prescribed corticosteroids based on medical record reporting. RMH prognostic score data were available on 209 patients.

Four hundred and nine features were found to be reproducible ($CCC \geq 0.80$). The

Table 4.3
Hematology results of patients by rapid disease progression phenotypes.

Characteristic	TTP < 2 months	TTP ≥ 2 months	P-Value	HPD	Non-HPD	P-Value
Hematology Data, median, (95% CI)₅						
LDH, U/L	324 (225-417)	236 (216-263)	0.046	510 (328-578)	235 (214-280)	0.001
Serum albumin, g/dL	3.8 (3.6-3.9)	4 (3.9-4.1)	<0.001	3.8 (3.4-4.1)	3.9 (3.9-4.0)	0.182
Lymphocytes, 1e+9/L	1.14 (0.98-1.28)	1.31 (1.21-1.40)	0.023	1.16 (0.68-1.37)	1.26 (1.18-1.38)	0.342
WBC, 1e+9/L	7.7 (6.82-8.95)	7.08 (6.79-7.57)	0.180	7.94 (6.24-10.51)	7.39 (7.03-7.97)	0.434
Neutrophils, 1e+9/L	5.65 (4.40-6.73)	4.82 (4.51-5.10)	0.081	5.8 (4.22-8.55)	4.98 (4.68-5.41)	0.284
Fibrinogen, mg/dL	635 (550-676)	491 (467-537)	0.003	552 (440-695)	532 (490-593)	0.441
NLR	4.65 (3.67-5.82)	3.73 (3.35-4.06)	0.007	5.51 (3.41-9.71)	3.82 (3.58-4.29)	0.238

Abbreviations: HPD = hyperprogressive disease; TTP = time-to-progression; TGR = tumor growth rate.

Bold P-values are statistically significant.

Hematology data were available for 208 patients for LDH, 226 patients for serum albumin, WBC and neutrophils, 224 patients for lymphocytes, 102 patients for fibrinogen.

Table 4.4
Mutational status results of patients by rapid disease progression phenotypes.

Characteristic	TTP < 2 months	TTP ≥ 2 months	P-Value	HPD	Non-HPD	P-Value
EGFR mutational status₆						
Not Detected	30 (85.7)	101 (83.5)		9 (100)	105 (84.0)	
Detected	5 (14.3)	20 (16.5)	1.000	0 (0)	20 (16.0)	0.355
KRAS mutational status₆						
Not Detected	17 (70.8)	57 (68.7)		5 (71.4)	58 (69.1)	
Detected	7 (29.2)	26 (31.3)	1.000	2 (28.6)	26 (30.9)	1.000

Abbreviations: HPD = hyperprogressive disease; TTP = time-to-progression; TGR = tumor growth rate.

Bold P-values are statistically significant.

Mutational status data were available for 156 patients for EGFR and 107 patients for KRAS.

reproducible features that were found to be marginally significant ($p < 0.10$) on uni-

variable analysis (Table A.4) were then considered for inclusion in the multivariable analyses. In multivariable analysis the parsimonious radiomics model with only four features yielded an AUROC of 0.717 (Figure 4.4). In multivariable analysis for HPD versus non-HPD, the parsimonious radiomics model with only one feature yielded an AUROC 0.674.

The parsimonious models for the clinical covariates and radiomic features were combined into a final clinical-radiomic model (Figure 4.4). The final clinical-radiomic TTP < 2 months vs TTP \geq 2 months model yielded an AUROC of 0.804 with 83.4% specificity, 63.4% sensitivity and 73.4% accuracy after SMOTE sub-sampling and optimal cut-point approach (Table 4.7). The final clinical-radiomic HPD vs non-HPD model yielded an AUROC score 0.865 (Figure 4.4) with 92.9% specificity 74.0% sensitivity and 82.3% accuracy after SMOTE sub-sampling (Table 4.7). When patients

Table 4.5

Multivariable models of clinical covariates associated with TTP < 2 months vs TTP \geq 2 months.

Covariate	TTP < 2 months vs TTP \geq 2 months			
	Multivariable Model		Parsimonious	
	OR (95% CI)	P-value	Multivariable Model ₁ OR (95% CI)	P-value
Previous lines of systemic treatment	2.13 (0.86-5.29)	0.103	2.62 (1.30-5.30)	0.007
Hepatic metastasis	2.10 (0.91-4.87)	0.084	2.50 (1.14-5.48)	0.022
Bone metastasis	2.48 (1.15-5.33)	0.020	2.54 (1.21-5.33)	0.014
RMH prognostic score	-	-	-	-
LDH₂, U/100 L	1.05 (0.93-1.19)	0.434	-	-
Serum albumin, g/dL	0.33 (0.11-1.02)	0.053	-	-
Lymphocytes, 1e+9/L	0.99 (0.48-2.06)	0.985	-	-
Ratio of: neutrophils/lymphocytes	1.09 (0.99-1.21)	0.073	1.14 (1.06-1.23)	<0.001
AUROC for the model	0.770		0.736	

Abbreviations: TTP, Time-to-progression; OR, odds ratio; CI, confidence interval; AUROC, area under the receiver-operator characteristics.

Bold ORs and P-values are statistically significant.

₁The ORs for image features identified from stepwise backward elimination.

₂LDH count odds ratios were calculated based on U / 100 L.

Table 4.6
Multivariable models of clinical covariates associated with HPD vs non-HPD.

Covariate	HPD vs non-HPD			
	Multivariable Model OR (95% CI)	P-value	Parsimonious Multivariable Model _{1,2} OR (95% CI)	P-value
Previous lines of systemic treatment	-	-	-	-
Hepatic metastasis	-	-	-	-
Bone metastasis	-	-	-	-
RMH prognostic score	4.46 (1.25-15.88)	0.021	6.13 (1.86-20.23)	0.003
LDH ₃ , U/100 L	1.12 (0.94-1.35)	0.213	-	-
Serum albumin, g/dL	-	-	-	-
Lymphocytes, 1e+9/L	-	-	-	-
Ratio of: neutrophils/lymphocytes	-	-	-	-
AUROC for the model	0.783		0.712	

Abbreviations: HPD; Hyperprogressive disease; OR, odds ratio; CI, confidence interval; AUROC, area under the receiver-operator characteristics.

Bold ORs and P-values are statistically significant.

₁The ORs for image features identified from stepwise backward elimination.

₂Because of the low number of HPD patients, a threshold of 0.01 instead of 0.05 was used to produce a more stringent parsimonious model using stepwise backward elimination approach.

₃LDH count odds ratios were calculated based on U / 100 L.

were grouped based on an HPD included RECIST at first follow-up (Figure 4.5), HPD patients had significantly lower OS. Patient-level probabilities from the final TTP < 2 months vs TTP ≥ 2 months model was sub-grouped and analyzed for survival outcomes where patients with lower probability scores had significantly improved PFS. These results were consistent when the probability scores were categorized by CART (Figure 4.6) and percentiles (Figure A.2).

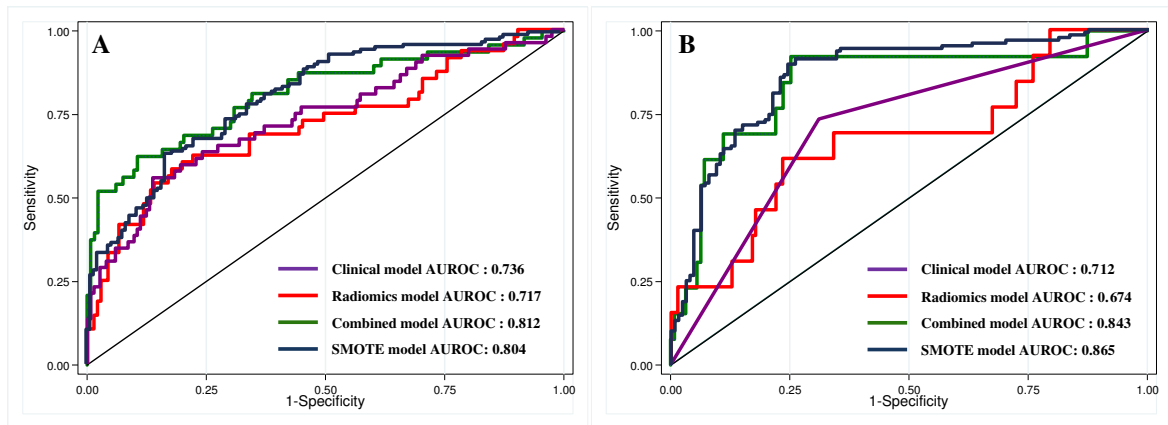


Figure 4.4 AUROCs for clinical only models, radiomic only models, and combined clinical-radiomic models. a) TTP < 2 months versus TTP \geq 2 months group. The clinical features in the models were: Previous lines of therapy, presence of hepatic metastasis, presence of bone metastasis and NLR. The radiomics features were: Radial gradient border SD-2D, 3D Laws E5L5E5, border 3D Laws E5E5L5 and border quartile coefficient of dispersion. b) HPD vs non-HPD. The clinical feature in the models was: RMH prognostic score and the radiomic feature was: border NGTDM strength.

4.4 Discussion

Although immunotherapy can provide clinical benefit among late stage lung cancer patients, published studies have reported that only 20-50% of patients exhibit durable responses [2–7, 112]. Moreover, rapid disease progression is associated with rapid and lethal outcomes (Figure 4.5), particularly HPD [104]. As such, in this study novel parsimonious models were identified containing highly informative clinical data and radiomic features to predict rapid disease progression phenotypes of NSCLC patients with modest to high AUROCs ranging from 0.804 to 0.865 (Figure 4.4) and accuracies ranging from 73.4% to 82.3%.

For the analyses comparing TTP < 2 months vs TTP \geq 2 months, the combined final radiomic-clinical model produced an AUROC of 0.804 (with 83.4% specificity, 63.4% sensitivity, and 73.4% accuracy) based on four clinical covariates and four radiomic features. The clinical covariates in the final model were presence of hepatic and bone metastasis, previous lines of systemic therapies, and NLR. Prior studies have shown that presence of hepatic and/or bone metastases are associated with poor out-

comes on NSCLC patients [113,114]. In the bone marrow, the resident and infiltrated stromal cells and their molecular mediators create a unique bone metastatic microenvironment that leads to persistent immunosuppression [114] and the liver is a known site of immune tolerance which may be the cause of resistant liver metastasis lesions to immunotherapy. Elevated baseline-derived NLR have been associated with reduced OS and progression-free survival (PFS) in NSCLC treated with immunotherapy [113,115]. Increased NLR is a measure of reduced lymphocytes which suppresses the host immune response where elevated NLR in the blood could be an implication of inflammatory cells being present in tumor microenvironment (TME) that leads to immune evasion or escape from the immune intervention [116]. Tumors that were exposed to higher number of systemic treatments could result in a clonal selection of resistant tumor cells that are able to evade immune-surveillance. Another possibility is that these tumors might have an innate resistance to any cancer therapy thereby exhibiting an “immune-desert” phenotype. Three of the four radiomic features (RG border SD-2D, border quartile coefficient of dispersion, border 3D Laws E5E5L5,) were extracted from the border regions of the tumors, which is the immediate outside of the tumors that may be capturing data related to the TME. RG border SD-2D feature has previously shown to be associated with OS and semantic radiology of lung cancer patients [51]. Border quartile coefficient of dispersion, calculates the dispersion of the distribution of border region intensities which quantifies the degree of heterogeneity of the TME. Two 3D Laws features were also identified (3D Laws E5L5E5 and border 3D Laws E5E5L5) that quantify voxel-by-voxel differences and capture spatial and textural heterogeneity across the ROI [117].

Table 4.7

Performance statistics from the combined multivariable models and SMOTE sub-sampled multivariable models.

Performance statistics	TTP < 2 months vs TTP ≥ 2 months₁		HPD vs non-HPD₂	
	Combined model without SMOTE	Combined model with SMOTE	Combined model without SMOTE	Combined model with SMOTE
Area under ROC curve (95% CI)	0.812 (0.731-0.890)	0.804 (0.752- 0.855)	0.843 (0.712-0.973)	0.865 (0.818-0.911)
Accuracy (95% CI)	84.50% (78.42-89.47)	71.2% (65.36-76.53)	90.71% (84.64-94.96)	78.74% (73.19-83.61)
Specificity	97.00%	74.44%	99.21%	66.14%
Sensitivity	50.00%	67.91%	7.69%	91.34%
PPV	85.71%	72.80%	50.00%	72.96%
Optimal cut-point				
Optimal cut-point	0.369	0.563	0.12	0.591
Area under ROC curve (95% CI)	0.812 (0.731-0.890)	0.804 (0.752- 0.855)	0.843 (0.712-0.973)	0.865 (0.818-0.911)
Accuracy (95% CI)	81.81% (75.36-87.11)	73.41% (67.68-78.61)	75.71% (65.75-82.56)	82.28% (77.02-86.77)
Specificity	88.72%	83.46%	74.80%	74.02%
Sensitivity	62.50%	63.43%	92.31%	90.55%
PPV	66.67%	79.44%	26.68%	77.70%

Abbreviations: CI= confidence interval; HPD = hyperprogressive disease; TTP = time-to-progression; PPV = positive predictive value.

₁The clinical features in the model were: Previous lines of therapy, presence of hepatic metastasis, presence of bone metastasis and NLR. The radiomics features were: Radial gradient border SD-2D, 3D Laws E5L5E5, border 3D Laws E5E5L5 and border quartile coefficient of dispersion.

₂The clinical feature in the model was RMH prognostic score and the radiomic feature was border NGTDM strength.

The final clinical-radiomic model for HPD versus non-HPD analysis included RMH prognostic score and one radiomic feature and revealed an AUROC of 0.865 with 92.9% specificity 74.0% sensitivity and 82.3% accuracy. The RMH prognostic score is a clinical prognostic score which has been shown to predict OS of immunotherapy patients [104] and advanced staged lung cancer patients [118]. The radiomic feature, border Neighborhood Grey Tone Difference Matrix (NGTDM) strength, quantifies the similarity of pixel intensities within a neighborhood to derive heterogeneity and texture of the TME [119]. Though the biological underpinnings of HPD are currently unknown, published studies have speculated that checkpoint blockades may be in a proliferative activating oncogenic signaling response [104], aberrations in driver genes may be co-amplifying pathways that initiate HPD [108] and regulatory T-cells that suppress antitumor T-cell responses may promote tumor proliferation [109]. As radiomic features have been shown to capture biological information and genomic pathways [50] and correlate with genomic mutations [120], we speculate that these features associated with rapid disease progression could be capturing one or more of these characteristics. Nonetheless, further research is needed to elucidate the mechanisms of rapid disease progression and HPD and of the biology of the radiomic features that predicts disease outcomes. Both limitations and strengths of this study are acknowledged. This is an

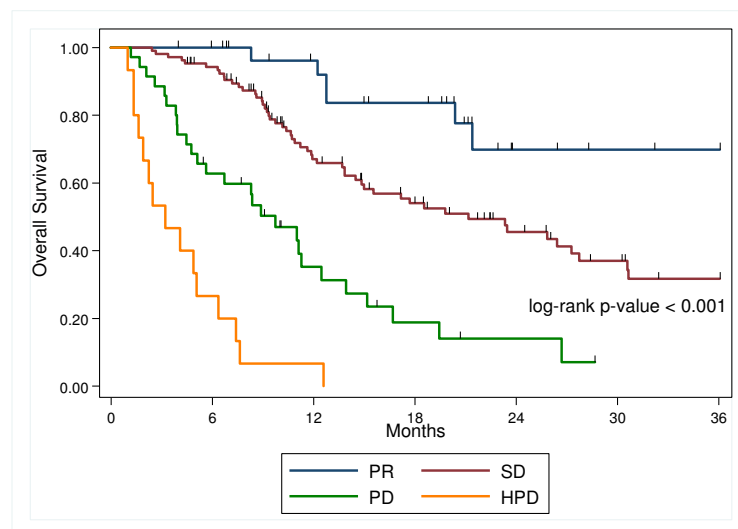


Figure 4.5 Kaplan-Meier curves for OS from initiation of therapy to date of last follow-up or date of death. The patient subgroups were based on RECIST at first follow-up for PD, SD, PR, and HPD.

early report demonstrating the potential utility of radiomics to predict rapid disease

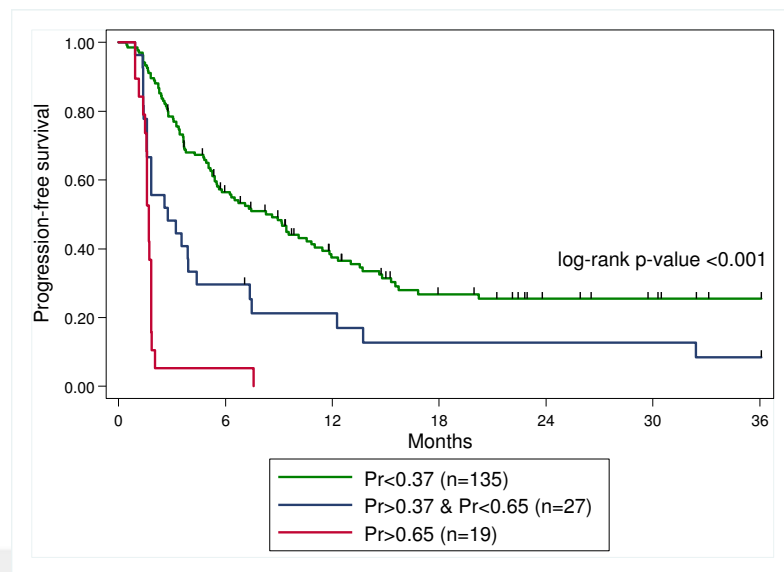


Figure 4.6 Kaplan-Meier curve for PFS based on novel cut-points identified by CART analysis points using individual probability of each patient for the $TTP < 2$ months versus $TTP \geq 2$ months model. CART analysis did not identify a cut-point for the HPD analysis.

progression phenotypes; these results will need to be replicated in independent test and validation cohorts. PD-L1 expression data were not available for these patients since the IHC assays were performed by the industry sponsors. Though, prior studies have revealed that PD-L1 expression has limitations in predicting immunotherapy response [121, 122]. A recent clinical trial study demonstrated that regardless of PD-L1 expression, immunotherapy combined with chemotherapy offers survival benefit [7]. Thus, inclusion of PD-L1 status may add little or no improvement to the performance of our models.

Because of the complexity in objective immunotherapy response, including hyperprogressive disease, pseudo-progression, and acquired resistance, there is a pressing challenge to identify biomarkers to predict patients that are least likely to respond. In this early report, novel models containing informative clinical covariates and radiomic image features were identified to predict rapid disease phenotypes including HPD. Additional research with independent test and validation will be needed to demonstrate the clinical utility of these findings.

5. STABILITY AND REPRODUCIBILITY OF COMPUTED TOMOGRAPHY RADIOMIC FEATURES EXTRACTED FROM PERITUMORAL REGIONS OF LUNG CANCER LESIONS

5.1 Introduction

Radiomics is the process of converting standard-of-care medical images into quantitative image-based data that can subsequently be analyzed using conventional biostatistics, machine learning approaches, and artificial intelligence [12]. Conventional radiomic features based on shape, size, intensity, and texture are typically extracted from the intratumoral ROI to quantify the cancer phenotype [105]. These radiomic features, of which many are beyond visual acuity, have shown to be significantly associated with cancer detection, diagnosis, prognosis, prediction of response to treatment, and monitoring of disease status [17, 43, 50, 64, 123]. However, there has been a renewed interest in quantitative characterization of the peritumoral region, the area immediately surrounding the tumor parenchyma, since this region is involved in immune infiltration, blood and lymphatic vascular networks, and stromal inflammation [124–127]. Early efforts preceding the “modern era of Radiomics” demonstrated that peritumoral image-based features have diagnostic and predictive utility [94, 128–130]. Recent efforts have shown the clinical utility of peritumoral radiomic features in studies of lung, breast, and head and neck cancers [51, 131–135].

Prior studies have established that some radiomic features are sensitive to tumor segmentation and/or image acquisition hence unstable and non-reproducible [40, 136, 137], where stability is defined as the consistency of a feature across different segmentations and reproducibility is defined as the consistency of a feature across image acquisition parameters such as patient position and respiration phase. Identifying stable and reproducible features is an important precursor prior to conducting

analyses of radiomic data since features with low-fidelity will likely lead to spurious findings and unreproducible models. Though aforementioned studies have characterized the stability [40] and reproducibility [136,137] of intratumoral radiomics, no such study to date has been conducted on peritumoral radiomic features.

To address the gap in this domain, this study was conducted to assess the stability and reproducibility of peritumoral radiomic features of lung lesions captured by thoracic CT scans. This study is also different from prior work conducted on intratumoral radiomics in that the majority of the radiomic features that were evaluated in this study were standardized through algorithms defined by IBSI [76]. To measure stability we utilized the “Moist-run” dataset [19] from The Cancer Imaging Archive and to measure reproducibility we utilized the RIDER dataset that consists of test-retest data [20]. Peritumoral ROIs with incremental distances of 3 mm to 12 mm from the tumor boundary were generated by applying morphological image processing operations on tumor segmentation masks. The clinical utility of stable and reproducible peritumoral features was tested on three previously published lung cancer datasets using OS as the endpoint. The stable and reproducible features identified in this study could be applied to a feature selection pipeline for CT radiomic analyses.

5.2 Materials and Methods

5.2.1 Moist-run dataset

The Moist-run dataset was utilized to measure radiomic feature stability. This dataset was constructed by the Quantitative Imaging Network as a lung segmentation challenge [19] and consists of 40 CT images of 40 NSCLC patients from five collections of Digital Imaging and Communications Medicine series and one thoracic phantom. Each patient in the dataset had one lesion of interest and the phantom scan had 12 lesions of interest which totals to 52 lesions of interests. The images on this dataset were previously de-identified.

5.2.2 RIDER test-retest dataset

To measure reproducibility of radiomic features, the RIDER test-retest dataset was utilized [20]. This NCI dataset was developed to generate an initial consensus on how to harmonize data collection and analysis for quantitative imaging methods. This dataset consisted of 32 NSCLC patients who had two non-contrast chest CT scans that were acquired within 15 minutes of each other using the same scanner, acquisition and processing parameters. As such, the only variability between the test and re-test scans would be attributed to patient orientation, respiration, and movement. The images on this dataset were previously de-identified.

5.2.3 Prognostic lung cancer datasets

To test the applied utility of stable and reproducible peritumoral features, three previously published datasets were utilized. One dataset was used for training and two datasets for validation. The training dataset included 62 surgically resected lung adenocarcinoma patients from the H. Lee Moffitt Cancer Center & Research Institute who had CTs two months prior to surgery [14, 51]. The first validation cohort included 47 lung adenocarcinoma patients from MAASTRO, Maastricht, Netherlands [14, 51] and the second validation cohort was a radiogenomic dataset [14, 51] that included 103 adenocarcinoma patients who had pre-surgery CTs.

5.2.4 Segmentation Algorithms

The lesions on the Moist run dataset were previously segmented using three different segmentation algorithms. Each segmentation algorithm was implemented using three different initial parameters (i.e., seed point or bounding circle, Figure 5.1) hence; nine segmentations per lesion were obtained. Algorithm 1 uses marker-controlled watersheds, geometric active contours and Markov random fields inside a user drawn bounding circle ROI surrounding the lesion. Algorithm 2 requires a single-click inside

the lesion as an initial parameter which then automatically generates multiple seed points inside the tumor. Subsequently, a click and grow algorithm was used to generate multiple segmentations that are combined to generate a consensus segmentation. Algorithm 3 uses a “seed circle” as an initial parameter and applies a two-dimensional region growing technique followed by automatic removal of blood vessels and lung parenchyma. Further details of the segmentation algorithms were previously described elsewhere [40]. The lesions on the RIDER dataset were previously segmented using

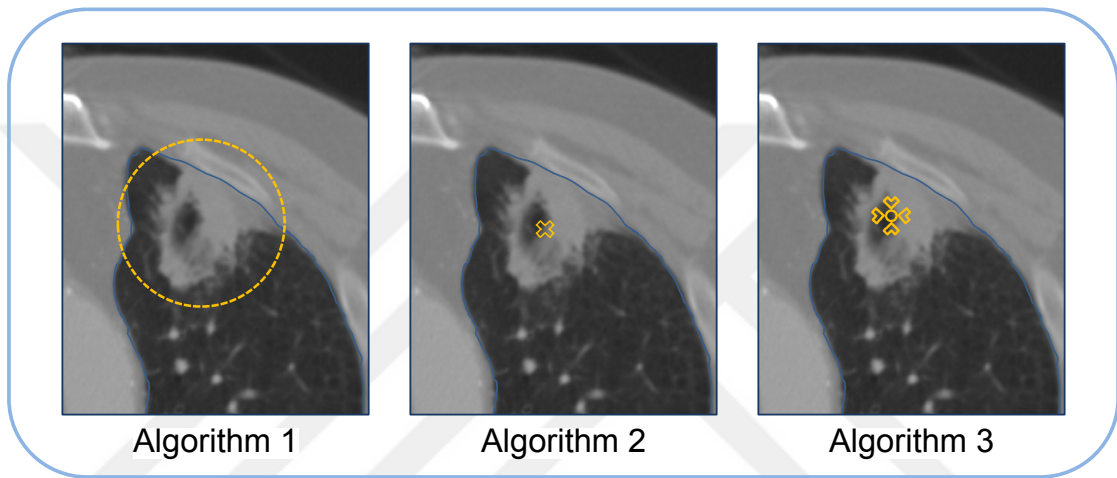


Figure 5.1 Initial parameters used on segmentation algorithms. Algorithm 1 uses a bounding circle while algorithms 2 and 3 uses seed points as initial parameters.

a semi-automatic single-click ensemble region growing segmentation algorithm on the LuTA software program platform (Definiens Developer XD©, Munich, Germany) [97]. The segmentation work flow contained four steps: 1) Pre-processing of automatic organ segmentation; 2) Semi-automated correction of pulmonary boundary; 3) Click and Grow execution; 4) A manual refinement by an expert if needed. Further details of the segmentation algorithms were previously published elsewhere [77].

5.2.5 Peritumoral Masks

Peritumoral masks were generated as a natural extension of the tumor segmentations by using morphological image processing operations. A disk-shaped structural element with a radius of intended peritumoral distance was used for morphological

dilation on tumor segmentations, followed by removal of the tumor region to create “doughnut-shaped” peritumoral masks. Intervals of 3, 6, 9 and 12 mm outside the tumor were used to create the peritumoral masks. For the first analysis, the peritumoral regions were bounded by the lung parenchyma and for the second analysis the peritumoral regions were not bounded by lung parenchyma (i.e., peritumoral regions were allowed to extend outside of lung parenchyma, Figure 5.2). The MATLAB® (version 2018a) scripts to create peritumoral masks from intratumoral masks are available at https://github.com/TunaliIlke/peritumoral_regions/.

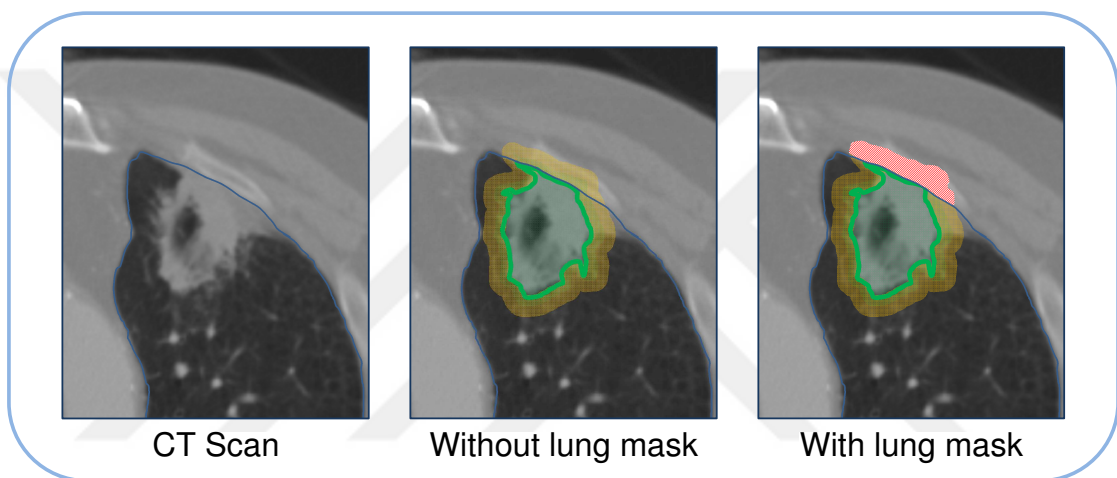


Figure 5.2 Peritumoral masks. Image on the left shows a CT scan ROI. Image on the middle image shows a peritumoral region that is not bounded by the lung parenchyma mask. Image on the right shows peritumoral region bounded by the lung parenchyma mask. Red region is removed from the peritumoral region as it lies outside the lung parenchyma.

5.2.6 Radiomic Features

All images were linearly resampled to a single voxel spacing of 1mm x 1mm x 1mm to standardize spacing across all images. A total of 264 statistical, histogram and texture radiomic features were extracted from the selected peritumoral and intratumoral ROIs using in-house toolboxes created in C++ (<https://isocpp.org>). Texture features included gray level co-occurrence matrix (GLCM), gray level run-length matrix (GLRLM), gray level size zone matrix (GLSZM) and neighboring gray tone difference matrix (NGTDM), 3D Laws and wavelet features. All features (except wavelet texture

features) were extracted using standardized algorithms defined by the IBSI v5 [76]. Wavelet features were extracted from the peritumoral and intratumoral ROIs using algorithms described elsewhere [77]. Histogram, co-occurrence, GLRLM, GLSZM and NGTDM texture features were extracted using a common bin width of 25 HUs. Additionally, 41 IBSI standardized shape and size features were extracted from intratumoral masks.

5.2.7 Statistical Analyses

Statistical analyses were performed using Intercooled Stata/MP 14.2 (Stata-Corp LP, College Station, TX) and R Project for Statistical Computing version 2.13.1 (<http://www.r-project.org>). Stability and reproducibility of features were assessed using CCC. For each feature, CCCs were calculated between different segmentation algorithms, initial parameters and test-retest scans. The CCC values range from 1 to -1, where 1 indicates a perfect correlation between two variables. Similarity between different segmentation approaches were computed using the Jaccard index:

$$J(Seg_1, Seg_2) = \frac{|Seg_1 \cap Seg_2|}{|Seg_1 \cup Seg_2|} \quad (5.1)$$

where Seg_1 and Seg_2 are the two segmentation masks being compared. Differences between initial parameters and algorithms by varying distances were tested using Fisher's exact tests.

Briefly, survival analyses were performed using Kaplan-Meier survival estimates and the log-rank test. The OS was the main endpoint for these analyses and an event was defined as date of death. The OS was assessed from date of first treatment (e.g., surgery) to the date of death or date of last follow-up. The survival data were right censored at 60-months. All P-values were 2-sided and a P-value less than or equally to 0.05 was deemed statistically significant.

5.2.8 Survival Analyses

Utilizing the training cohort (Moffitt adenocarcinoma dataset) a univariable Cox regression model was created and stable and reproducible features (0-3 mm, not bounded by a lung mask) significantly associated with OS were retained. To produce a parsimonious model, the remaining radiomic features were included in a stepwise backward elimination Cox regression model using a threshold of 0.01 for inclusion. CART models were created identify patient risk groups utilizing the final radiomic features. Afterward, using the features found in the training cohort model, CART models were re-created utilizing the two validation cohorts.

5.3 Results

Table 5.1 presents the similarities between segmentations using Jaccard indices between different initial parameters and algorithms being used. The results demonstrate high similarities (Jaccard index > 0.90) between segmentations that were computed using different initial parameters. On the other hand, moderate similarities (Jaccard index > 0.80) were observed between segmentations that were computed using different segmentation algorithms.

5.3.1 Peritumoral Features

Figure 5.3 presents CCC groups (high, moderate, low) of peritumoral radiomic features with respect to different algorithms and different initial parameters. The green boxes represent high ($CCC > 0.95$), yellow boxes represent moderate ($CCC \geq 0.75$ & $CCC \leq 0.95$) and red boxes represent low ($CCC < 0.75$) CCCs. A high CCC indicates that the radiomic feature is not sensitive to variation in segmentations, whereas a low CCC indicates that radiomic feature is sensitive to the difference in segmentations. As peritumoral distance increased, there were significantly higher numbers of moderate or highly stable features (Table 5.2 and Table 5.3). The statistical (Table A.5), histogram

Table 5.1

Initial parameter and algorithm comparison by Jaccard index scores.

Initial parameter comparisons	Jaccard index
Algorithm 1-initial parameter 1 vs initial parameter 2	0.973
Algorithm 1-initial parameter 1 vs initial parameter 3	0.969
Algorithm 1-initial parameter 2 vs initial parameter 3	0.979
Algorithm 2-initial parameter 1 vs initial parameter 2	0.948
Algorithm 2-initial parameter 1 vs initial parameter 3	0.955
Algorithm 2-initial parameter 2 vs initial parameter 3	0.962
Algorithm 3-initial parameter 1 vs initial parameter 2	0.943
Algorithm 3-initial parameter 1 vs initial parameter 3	0.955
Algorithm 3-initial parameter 2 vs initial parameter 3	0.942
Algorithm comparisons¹	
Algorithm 1 vs Algorithm 2	0.81
Algorithm 1 vs Algorithm 3	0.827
Algorithm 2 vs Algorithm 3	0.805

¹Algorithms were compared using segmentations created by random selections of initial parameters (1-2-3) for each lesion.

(Table A.6), and a subset of texture features (Table A.7) were found stable and reproducible (Table 5.4) for different initial parameters however, 3D Laws (Table A.8) and wavelet texture (Table A.9) features were found to be significantly less stable and reproducible. Overall, the inter-stability (i.e. stability across different segmentation algorithms) was observed to be significantly lower than the intra-stability (i.e. stability across different initial parameters of same algorithms). The overall reproducibility of features were not significantly different as peritumoral distances changed; although a subset of texture features (GLCM, GLRLM, GLSZM and NGTDM) were slightly more reproducible for peritumoral distances above 3 mm (Table 5.4). All aforementioned analyses were performed for features extracted from peritumoral regions that were bounded by a lung parenchyma mask (Figure A.3). The stability of features where



Figure 5.3 Concordance correlation coefficient groups of peritumoral features bounded by lung parenchyma. The green boxes represent higher ($CCC > 0.95$), yellow boxes represent moderate ($CCC \geq 0.75$ and $CCC \leq 0.95$) and red boxes represent lower ($CCC < 0.75$) CCCs.

a lung parenchyma mask was not used to bound the peritumoral region was consistent with the analysis where a lung parenchyma mask was used (Tables A.10-A.15). However, peritumoral features were significantly more reproducible with the increasing peritumoral distances (Table 5.5).

Table 5.2

Distribution of stability groups of all peritumoral radiomic features extracted from ROIs that are bounded by a lung mask for different initial parameters.

	Distance ¹				P- Value
	0 – 3 mm	0 – 6 mm	0 – 9 mm	0 – 12 mm	
Algorithm 1-Initial parameter 1 vs Initial parameter 2					
CCC < 0.75	30 (11.4)	0 (0)	10 (3.8)	21 (8.0)	
CCC ≥ 0.75 & CCC ≤ 0.95	63 (23.9)	9 (3.4)	26 (9.9)	8 (3.0)	
CCC > 0.95	171 (64.7)	255 (96.6)	228 (86.3)	235 (89.0)	<0.001
P-value		<0.001	<0.001	0.001	
Algorithm 1-Initial parameter 1 vs Initial parameter 3					
CCC < 0.75	14 (5.3)	1 (0.4)	9 (3.4)	29 (11.0)	
CCC ≥ 0.75 & CCC ≤ 0.95	55 (20.8)	32 (12.1)	63 (23.9)	8 (3.0)	
CCC > 0.95	195 (73.9)	231 (87.5)	192 (72.7)	227 (86.0)	<0.001
P-value		<0.001	<0.001	<0.001	
Algorithm 1-Initial parameter 2 vs Initial parameter 3					
CCC < 0.75	33 (12.5)	2 (0.8)	8 (3.0)	10 (7.6)	
CCC ≥ 0.75 & CCC ≤ 0.95	34 (12.9)	24 (9.1)	38 (14.4)	0 (0)	
CCC > 0.95	197 (74.6)	238 (90.1)	218 (85.6)	244 (92.4)	<0.001
P-value		<0.001	0.022	<0.001	
Algorithm 2-Initial parameter 1 vs Initial parameter 2					
CCC < 0.75	57 (21.6)	36 (13.6)	44 (16.7)	44 (16.7)	
CCC ≥ 0.75 & CCC ≤ 0.95	122 (46.2)	86 (32.6)	52 (19.7)	11 (4.2)	
CCC > 0.95	85 (32.2)	142 (53.8)	168 (63.6)	209 (79.1)	<0.001
P-value		<0.001	0.004	<0.001	
Algorithm 2-Initial parameter 1 vs Initial parameter 3					
CCC < 0.75	40 (15.2)	22 (8.3)	37 (14.0)	45 (17.1)	
CCC ≥ 0.75 & CCC ≤ 0.95	124 (47.0)	76 (28.8)	39 (14.8)	10 (3.8)	
CCC > 0.95	100 (37.8)	166 (62.9)	188 (71.2)	209 (79.1)	<0.001
P-value		<0.001	<0.001	<0.001	
Algorithm 2-Initial parameter 2 vs Initial parameter 3					
CCC < 0.75	55 (20.8)	6 (2.3)	8 (3.0)	45 (17.1)	
CCC ≥ 0.75 & CCC ≤ 0.95	95 (36.0)	74 (28.0)	23 (8.7)	10 (3.8)	
CCC > 0.95	114 (43.2)	184 (69.7)	233 (88.3)	209 (79.1)	<0.001
P-value		<0.001	<0.001	<0.001	
Algorithm 3-Initial parameter 1 vs Initial parameter 2					
CCC < 0.75	182 (68.9)	146 (55.3)	121 (45.8)	97 (36.7)	
CCC ≥ 0.75 & CCC ≤ 0.95	78 (29.6)	99 (37.5)	108 (40.9)	122 (46.2)	
CCC > 0.95	4 (1.5)	19 (7.2)	35 (13.3)	45 (17.1)	<0.001
P-value		<0.001	0.023	0.096	
Algorithm 3-Initial parameter 1 vs Initial parameter 3					
CCC < 0.75	41 (15.5)	33 (12.5)	19 (7.2)	11 (4.2)	
CCC ≥ 0.75 & CCC ≤ 0.95	133 (50.4)	103 (39.0)	42 (15.9)	45 (17.1)	
CCC > 0.95	90 (34.1)	128 (48.5)	203 (76.9)	208 (78.7)	<0.001
P-value		0.004	<0.001	0.335	
Algorithm 3-Initial parameter 2 vs Initial parameter 3					
CCC < 0.75	49 (18.6)	48 (18.2)	30 (11.4)	12 (4.5)	
CCC ≥ 0.75 & CCC ≤ 0.95	156 (59.1)	112 (42.4)	66 (25.0)	49 (18.6)	
CCC > 0.95	59 (22.3)	104 (39.4)	168 (63.6)	203 (76.9)	<0.001
P-value		<0.001	<0.001	0.001	

¹P-values were generated using Fisher's Exact test comparing 0-6 mm vs. 0-3 mm, 0-9mm vs. 0-6 mm, and 0-12 mm vs. 0-9 mm, respectively.

5.3.2 Intratumoral Features

Figure 5.4 presents CCC groups of intratumoral radiomic features with respect to different algorithms and different initial parameters. The majority of the inter-

Table 5.3

Distribution of stability groups of all peritumoral radiomic features extracted from ROIs that are bounded by a lung mask for different segmentation algorithms.

	Distance ¹				P- Value
	0 – 3 mm	0 – 6 mm	0 – 9 mm	0 – 12 mm	
Algorithm 1 vs Algorithm 2					
CCC < 0.75	148 (56.1)	102 (38.6)	69 (26.1)	73 (27.6)	
CCC ≥ 0.75 & CCC ≤ 0.95	108 (40.9)	140 (53.0)	151 (57.2)	142 (53.8)	
CCC > 0.95	8 (3.0)	22 (8.4)	44 (16.7)	49 (18.6)	<0.001
	P-value	0.001	<0.001	0.722	
Algorithm 1 vs Algorithm 3					
CCC < 0.75	165 (62.5)	139 (52.6)	118 (44.7)	81 (30.7)	
CCC ≥ 0.75 & CCC ≤ 0.95	84 (31.8)	90 (34.1)	76 (28.8)	121 (45.8)	
CCC > 0.95	15 (5.7)	35 (13.3)	70 (26.5)	62 (23.5)	<0.001
	P-value	0.005	0.001	<0.001	
Algorithm 2 vs Algorithm 3					
CCC < 0.75	180 (68.2)	147 (55.7)	128 (48.5)	84 (31.8)	
CCC ≥ 0.75 & CCC ≤ 0.95	77 (29.2)	96 (36.4)	105 (39.8)	139 (52.7)	
CCC > 0.95	7 (2.6)	21 (7.9)	31 (11.7)	41 (15.5)	<0.001
	P-value	0.002	0.158	<0.001	

¹P-values were generated using Fisher's Exact test comparing 0-6 mm vs. 0-3 mm, 0-9mm vs. 0-6 mm, and 0-12 mm vs. 0-9 mm, respectively.

stabilities were low ($CCC < 0.75$) while intra-stabilities were more frequently moderate or high (Table A.16). Most size and shape features were found to be highly stable for different initial parameters (Figure 5.4). Intensity, a subset of a texture features (GLCM, GLRLM, GLSZM and NGTDM), size and shape features were at moderately or highly reproducible, while 3D Laws and wavelet features were less reproducible (Table A.24). For all feature categories, intratumoral features had lower median CCC values than their corresponding peritumoral features for both reproducibility and stability assessments (Figure 5.5).

Table 5.4

Distribution of reproducibility groups of peritumoral radiomic features extracted from ROIs that are bounded by a lung mask.

	Distance ¹				P- Value ²
	0 – 3 mm	0 – 6 mm	0 – 9 mm	0 – 12 mm	
All features					
CCC < 0.75	68 (25.8)	55 (20.8)	71 (26.9)	80 (30.3)	
CCC ≥ 0.75 & CCC ≤ 0.95	121 (45.8)	148 (56.1)	138 (52.3)	113 (42.8)	
CCC > 0.95	75 (28.4)	61 (23.1)	55 (20.8)	71 (26.9)	<0.001
P-Value		0.068	0.263	0.081	
Statistical features					
CCC < 0.75	3 (15.8)	2 (10.5)	3 (15.8)	3 (15.8)	
CCC ≥ 0.75 & CCC ≤ 0.95	1 (5.3)	5 (26.3)	4 (21.0)	4 (21.0)	
CCC > 0.95	15 (78.9)	12 (63.2)	12 (63.2)	12 (63.2)	0.731
P-Value		0.272	1.000	1.000	
Histogram features					
CCC < 0.75	5 (17.9)	2 (7.1)	4 (14.3)	4 (14.3)	
CCC ≥ 0.75 & CCC ≤ 0.95	7 (25.0)	15 (53.6)	10 (35.7)	10 (35.7)	
CCC > 0.95	16 (57.1)	11 (39.3)	14 (50.0)	14 (50.0)	0.512
P-Value		0.086	0.467	1.000	
Texture^c features					
CCC < 0.75	3 (4.8)	0 (0)	0 (0)	1 (1.6)	
CCC ≥ 0.75 & CCC ≤ 0.95	28 (45.2)	39 (62.9)	33 (53.2)	31 (50.0)	
CCC > 0.95	31 (50.0)	23 (37.1)	29 (46.8)	30 (48.4)	0.198
P-Value		0.038	0.363	0.857	
3D Laws texture features					
CCC < 0.75	44 (35.2)	44 (35.2)	44 (35.2)	68 (54.4)	
CCC ≥ 0.75 & CCC ≤ 0.95	79 (63.2)	81 (64.8)	81 (64.8)	57 (45.6)	
CCC > 0.95	2 (1.6)	0 (0)	0 (0)	0 (0)	0.002
P-Value		0.615	1.000	0.003	
Wavelet texture features					
CCC < 0.75	13 (43.3)	7 (23.3)	20 (66.7)	4 (13.3)	
CCC ≥ 0.75 & CCC ≤ 0.95	6 (20.0)	8 (26.7)	10 (33.3)	11 (36.7)	
CCC > 0.95	11 (36.7)	15 (50.0)	0 (0)	15 (50.0)	<0.001
P-Value		0.305	<0.001	<0.001	

¹P-values were generated using Fisher's Exact test comparing 0-6 mm vs. 0-3 mm, 0-9mm vs. 0-6 mm, and 0-12 mm vs. 0-9 mm, respectively.

²P-value was generated using Fisher's Exact test for the overall distributions of the four peritumoral distance (3 x 4 contingency table)

³Features consist GLCM, GLRLM, GLSZM and NGTDM texture features.

Table 5.5

Distribution of reproducibility groups of peritumoral radiomic features extracted from ROIs that were not bounded by a lung mask.

	Distance ¹				P- Value ²
	0 – 3 mm	0 – 6 mm	0 – 9 mm	0 – 12 mm	
All features					
CCC < 0.75	66 (25.0)	49 (18.6)	89 (33.7)	73 (27.7)	
CCC ≥ 0.75 & CCC ≤ 0.95	158 (59.9)	118 (44.7)	92 (34.9)	58 (22.0)	
CCC > 0.95	40 (15.1)	97 (36.7)	83 (31.4)	133 (50.3)	<0.001
	P- Value	<0.001	<0.001	<0.001	
Statistical features					
CCC < 0.75	3 (15.8)	2 (10.5)	3 (15.8)	3 (15.8)	
CCC ≥ 0.75 & CCC ≤ 0.95	10 (52.6)	1 (5.3)	0 (0)	0 (0)	
CCC > 0.95	6 (31.6)	16 (84.2)	16 (84.2)	16 (84.2)	<0.001
	P- Value	0.001	1.000	1.000	
Histogram features					
CCC < 0.75	2 (7.1)	2 (7.1)	3 (10.7)	3 (10.7)	
CCC ≥ 0.75 & CCC ≤ 0.95	21 (75.0)	5 (17.9)	5 (17.9)	2 (7.1)	
CCC > 0.95	5 (17.9)	21 (75.0)	20 (71.4)	23 (82.1)	<0.001
	P- Value	<0.001	1.000	0.531	
Texture³ features					
CCC < 0.75	3 (4.8)	0 (0)	0 (0)	0 (0)	
CCC ≥ 0.75 & CCC ≤ 0.95	43 (69.4)	25 (40.3)	22 (35.5)	16 (25.8)	
CCC > 0.95	16 (25.8)	37 (59.7)	40 (64.5)	46 (74.2)	<0.001
	P- Value	<0.001	0.711	0.330	
3D Laws texture features					
CCC < 0.75	44 (35.2)	44 (35.2)	68 (54.4)	67 (53.6)	
CCC ≥ 0.75 & CCC ≤ 0.95	79 (63.2)	74 (59.2)	50 (40.0)	27 (21.6)	
CCC > 0.95	2 (1.6)	7 (5.6)	7 (5.6)	21 (24.8)	<0.001
	P- Value	0.271	0.007	<0.001	
Wavelet texture features					
CCC < 0.75	14 (46.7)	1 (3.3)	15 (50.0)	0 (0)	
CCC ≥ 0.75 & CCC ≤ 0.95	5 (16.6)	13 (43.3)	15 (50.0)	13 (43.3)	
CCC > 0.95	11 (36.7)	15 (50.0)	0 (0)	17 (56.7)	<0.001
	P- Value	<0.001	<0.001	<0.001	

Numbers inside parenthesis are the percentage values.

¹P-values were generated using Fisher's Exact test comparing 0-6 mm vs. 0-3 mm, 0-9mm vs. 0-6 mm, and 0-12 mm vs. 0-9 mm, respectively.

²P-value was generated using Fisher's Exact test for the overall distributions of the four peritumoral distance (3 x 4 contingency table)

³Features consist GLCM, GLRLM, GLSZM and NGTDM texture features.

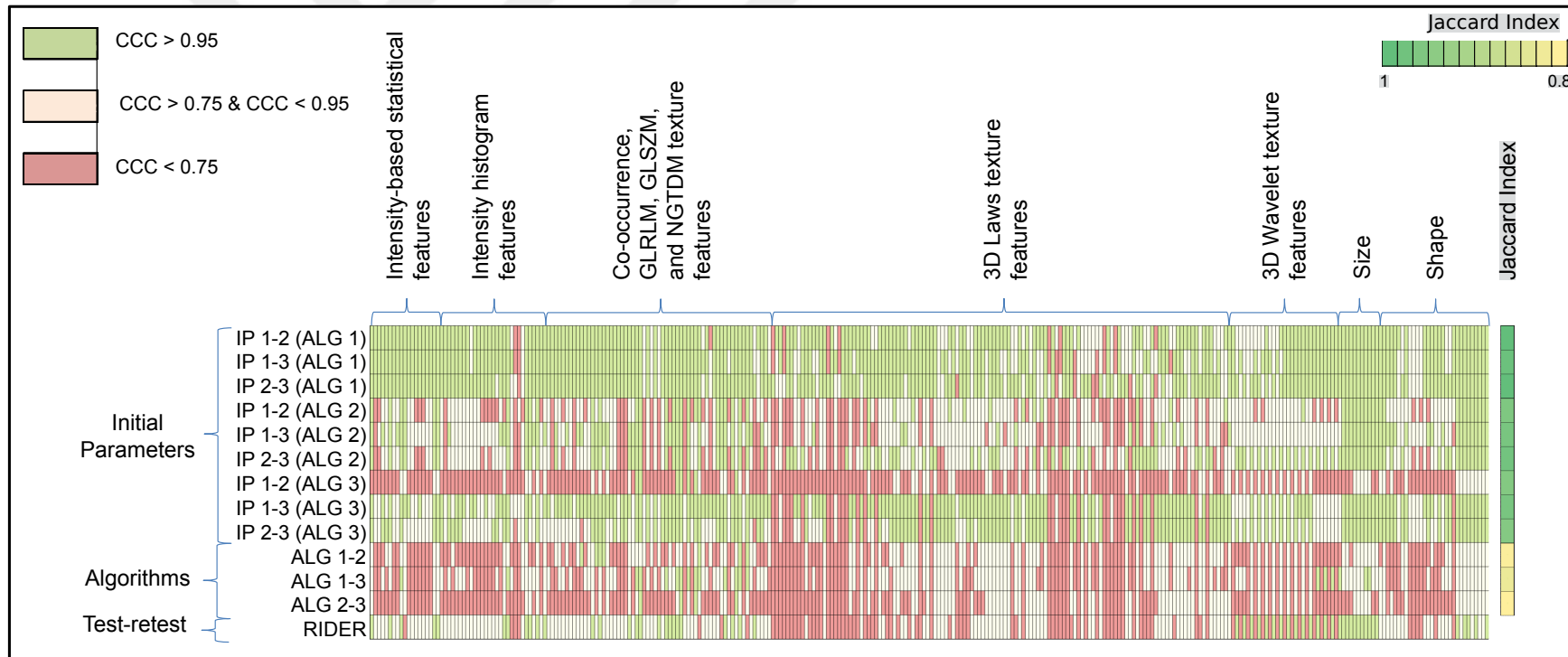


Figure 5.4 Concordance correlation coefficient groups of intratumoral features. The green boxes represent higher ($CCC > 0.95$), yellow boxes represent moderate ($CCC \geq 0.75$ and $CCC \leq 0.95$) and red boxes represent lower ($CCC < 0.75$) CCCs.

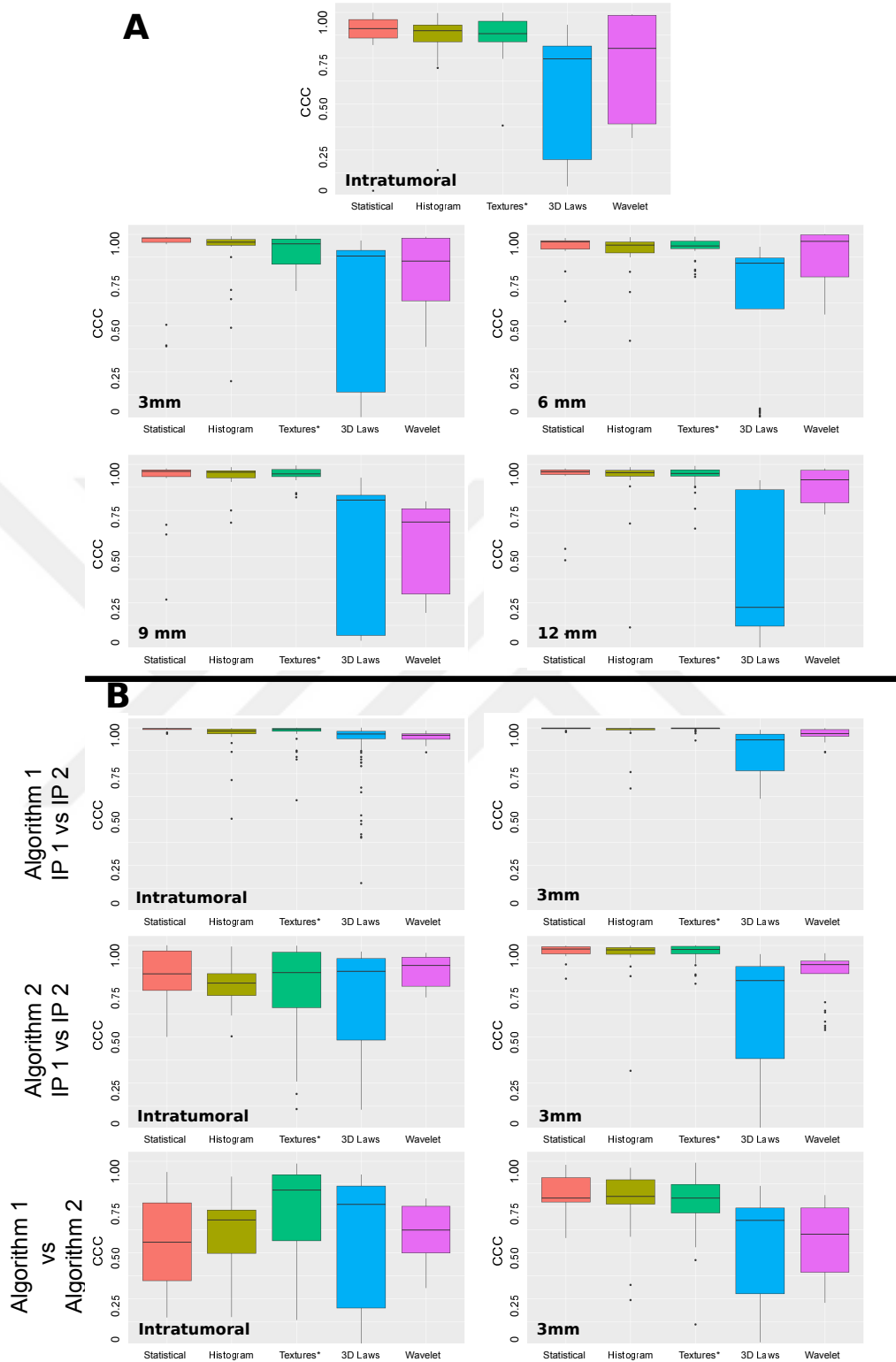


Figure 5.5 Whisker-box plots of CCCs by feature categories. a) CCCs of features extracted from test-retest dataset (RIDER) from intratumoral ROI and peritumoral ROIs. b) CCCs of features extracted using Algorithm 1 and Algorithm 2 from intratumoral ROI and peritumoral ROI (3 mm) of Moist-run dataset. *Texture features included GLCM, GLRLM, GLSZM and NGTDM features.

5.3.3 Survival analysis of peritumoral features

Utilizing the training cohort, univariable Cox regression analyses were conducted using the stable and reproducible peritumoral features (0-3 mm, not bounded by a lung mask, $n = 63$) and the top performing features ($n = 5$, $p < 0.05$) were selected for multivariable analysis. These remaining features were included in a stepwise backward elimination Cox regression model and one feature (F300:3D_Wavelet_P2.L2.C11) remained in the final model. The classification and regression tree analysis identified the optimal cut-point ($\geq 1.18 \times 10^{-4}$) that discriminated by OS in the training dataset and found that patients categorized in high cut-point had significantly worse survival. Applying the novel cut-off point to two independent cohorts showed that F300:3D_Wavelet_P2.L2.C11 was prognostic in all three cohorts (Figure 5.6).

5.4 Discussion

Radiomics are powerful image-based biomarkers that have been successfully applied for cancer detection, diagnosis, prognosis, prediction of response to treatment, and monitoring of disease status by converting standard-of-care medical images into quantitative data [17, 43, 50, 64, 123]. Because the surrounding peripheral areas of the tumors represent the tumor microenvironment, emerging studies have considered the clinical utility of peritumoral radiomic features [51, 131–135]. Overall, a subset of peritumoral features were found to be stable and reproducible. Features found to be stable regardless of the peritumoral distances included statistical and histogram and a subset of texture features (GLCM, GLRLM, GLSZM and NGTDM). This suggests these features are less affected by changes in the ROIs. Also the stability and reproducibility of most 3D Laws and wavelet texture features were found to be inconsistent across the peritumoral regions which have been shown in other studies of intratumoral radiomics [77, 136]. As such, the inclusion of a subset of 3D Laws and wavelet texture features may result in spurious and non-repeatable findings. Also, when the clinical utility of stable and reproducible peritumoral radiomic features were assessed in relation to lung cancer survival, stable and reproducible features were found to be more

likely to get validated than unstable and non-reproducible features. Specifically, the top performing peritumoral feature was a stable and reproducible peritumoral feature utilizing three previously published datasets [14, 51, 138] (Figure 5.6).

Although prior studies have been conducted to assess for stability [40] and reproducibility [136, 137] of intratumoral radiomic features, this is the first study conducted on peritumoral radiomic features. Kalpathy-Cramer *et al.* [40] found that intratumoral size-based CT features were highly stable and shape-based features were less stable. However, their study found that texture-based features were less stable which is consistent with our findings for peritumoral texture-based features. On the other hand, size and shape-based features were not extracted for the peritumoral region in our study because these feature classes explicitly describe the intratumoral ROI. Balagurunathan *et al.* [136] found that most intratumoral features were reproducible utilizing a semi-automatic segmentation method on test-retest CT imaging which was also consistent with our findings however, we also observed that 3D Laws texture features were less reproducible than the rest of the feature groups. A separate study from Balagurunathan *et al.* [137] assessed lung tumor volumes across different segmentation algorithms and found that larger nodules (≥ 8 mm) were more reproducible. However, volumetric analyses of the peritumoral regions were not conducted in this study.

Because peritumoral masks are natural extensions of the intratumoral masks, the radiomic features extracted from the peritumoral and intratumoral regions could yield similar stability and reproducibility. Interestingly, the majority of the intratumoral features were unstable, especially when extracted using different segmentation algorithms (Figure 5.4). However, peritumoral features were found to be more stable and reproducible than their corresponding intratumoral features. This study also showed that peritumoral features further away from the intratumoral region were increasingly more stable. This finding might be related to the existence of homogenous lung parenchyma in distal peritumoral regions compared to intratumoral regions or peritumoral regions proximal to the tumor.

This analyses also revealed that subsets of features were consistently stable irre-

spective of the initial parameter (e.g., seed point) for a given segmentation algorithm. These findings are important since there is no ground truth for initial parameters for any segmentation algorithm and it is essential that features are consistent across different users. Whereas, some of these features were not stable when they were extracted using different segmentation algorithms. These results demonstrate the importance of using the same segmentation algorithm when conducting radiomics research especially when attempting to train, test, and validate findings.

In this study, no significant differences were found on the stability of features that were extracted from ROIs bounded by a lung parenchyma mask versus ROIs that were not bounded by a lung parenchyma mask. Although peritumoral features of lung tumors near the mediastinum or chest wall may be attenuated, our data suggests that these features are still stable. The clinical utility of including outside of the lung parenchyma to the ROI is currently unknown. Notably, pleural invasion by lung tumors is associated with a poor prognosis [139] and peritumoral features extracted from ROIs bounded by lung parenchyma may not accurately capture such a trait. Additionally, the lung parenchyma masks are not always available or are not included in software algorithms.

In 2017 a comprehensive review on the process and developments in radiomics by Lambin *et al.* [85] stated, "...optimal reproducibility and stability enable multicenter studies to maximize the likelihood of a validated radiomic signature being fit-for-purpose in routine clinical use." To meet this goal, assessing the reproducibility and stability using the framework presented here and by others [40, 136] provide groundwork to ensure generalizable studies across datasets and institutions. Because the peritumoral region has unique clinical and biological significance, capturing this information using radiomic analyses has tremendous translational utility as demonstrated from previous studies and this study [51, 131–135].

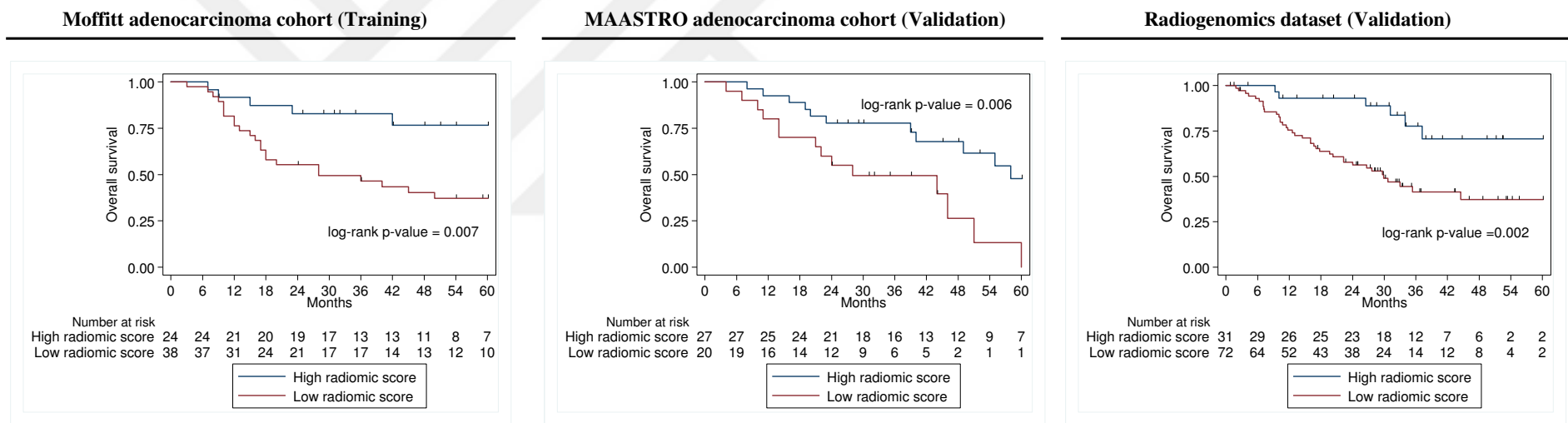


Figure 5.6 Kaplan-Meier graphs of overall survival for Moffitt adenocarcinoma dataset (right), MAASTRO adenocarcinoma dataset (middle), and Radiogenomics dataset (left) utilizing radiomics modelling of peritumoral features (0-3 mm).

In summary, this study identified a subset of stable and reproducible CT radiomic features from the peritumoral region of lung lesions. Because recent studies have shown evidence that peritumoral features have clinical significance [51, 131–135], identifying stable and reproducible features is crucial to minimize spurious and non-repeatable results. The stable and reproducible features identified in this study can be used to guide a feature selection pipeline for assessing the clinical utility of peritumoral CT radiomic features.



6. CLINICAL FACTORS AND QUANTITATIVE IMAGE-BASED FEATURES PREDICT IMMUNOTHERAPY RESPONSE AMONG LUNG CANCER PATIENTS

6.1 Introduction

Immunotherapy, which targets the PD-1 or PD-L1, has demonstrated durable clinical benefit in 20-50% patients with advanced stage NSCLC [2–7]. The patterns of immunotherapy response and progression are complex [140], including, e.g. rapid disease progression [65], hyperprogression [104], and acquired resistance [141]. Because of this complexity, there is a pressing challenge to identify robust predictive biomarkers that can identify patients that are least likely to respond. Though tumor PD-L1 expression by IHC is the only clinically approved biomarker to predict immunotherapy response, recent clinical trials demonstrated significant improvements in clinical outcomes irrespective of PD-L1 expression level [7, 10]. Furthermore, TMB, defined as the total number of mutations per coding area of a tumor genome [142], has been shown to be a superior predictor of immunotherapy response compared to PD-L1 status [143–145]. Despite the potential clinical utility of TMB, there are limitations with its use as tumor specimens have to be sufficient in both quantity and quality [145]. Further, tumors are evolutionarily dynamic and accumulate mutations rapidly [146], and laboratory methods to calculate TMB can be timely and expensive. Moreover, tumor-based biomarkers, including PD-L1 expression, are often subject to sampling bias due to the molecular and cellular heterogeneity of the biopsied tumors [147]. As such, complimentary biomarkers that are predictive, non-invasive, and measured in a timely fashion would have direct translational implications. Quantitative image-based features, or radiomics [12], reflects the underlying pathophysiology and tumor heterogeneity (Figure 2.1) and have many advantages over tissue-based biomarkers as they can be rapidly calculated from standard-of-care medical imaging and they reflect the

entire region-of-interest (e.g., tumor) and not just the portion of the tumor that is assayed.

In Chapter 3, the utility of radiomics to predict rapid disease progression and hyperprogression phenotypes in immunotherapy treated NSCLC patients was demonstrated. Building upon this work, in the present chapter pre-treatment clinical data and radiomic features extracted from CT scans were utilized to develop a parsimonious model to predict survival outcomes among NSCLC patients treated with immunotherapy. The biological underpinnings of the radiomics features were assessed utilizing gene-expression information from a well-annotated radiogenomics NSCLC dataset and IHC data from a separate NSCLC dataset. Furthermore, the radiomic features were assessed for survival in four independent NSCLC cohorts.

6.2 Materials and Methods

6.2.1 Immunotherapy-treated Lung Cancer Patients

This analysis included 332 stage IIIB or IV NSCLC patients that were treated with immunotherapy using PD-1 single agent (Nivolumab, Pembrolizumab), PD-L1 single agent (Durvalumab, Atezolizumab), or combination of PD-L1 or PD-L1 with cytotoxic T-lymphocyte-associated protein 4 (Ipilimumab, Tremelimumab) as second agent. Inclusion criteria included patients having a baseline CT or PET/CT scan less than 90 days prior to the initiation of immunotherapy and at least one RECIST target or non-target lung lesion. The patients were divided into discovery (MCC 1, N = 180), test cohort (MCC 2, N = 90) and validation cohorts (VA, N = 62). Patients in the training cohort were enrolled in clinical trials treated between June 2011 and January 2016 at Moffitt Cancer Center. Patients in the test cohort were treated with immunotherapy between May 2015 and October 2017 where 94.6% were treated as standard-of-care and 5.4% were enrolled in industry-sponsored clinical trials at Moffitt Cancer Center. Patients in the validation cohort were treated with standard-of-care immunotherapy between July 2015 and February 2019 at the James A. Haley Veterans'

Hospital.

Patient data were obtained from electronic medical records and institutional databases including demographics, stage of disease, histology, treatment, vital status, targeted mutations, ECOG performance, RECIST, hematology data, vital status (deceased or alive), and date of death or last follow-up. Date of progression was abstracted and defined as progressive disease from RECIST definition or clinical progression evaluated by the treating clinicians whenever RECIST was not available. This study was approved by the University of South Florida Institutional Review Board.

6.2.2 Radiogenomics Dataset

A previously described dataset [138] of 103 surgically resected adenocarcinoma patients who had pre-surgery CTs and gene expression data was used to identify potential biological underpinnings of the most informative radiomic feature. Briefly, gene expression was IRON-normalized and batch-corrected for RNA quality Pathway and Gene Ontology Enrichment was performed using Clarivate Analytics MetaCore [138].

6.2.3 Immunohistochemistry Dataset

To further investigate the connections between tumor biology and the radiomic features, a dataset of 26 surgically resected NSCLC patients who had pre-surgery CTs and treated at MCC were identified for immunohistochemical staining of carbonic anhydrase (CAIX). However, 16 cases were available with enough tumor tissue and sufficient staining quality.

Slides were stained using a Ventana Discovery XT automated system (Ventana Medical Systems, Tucson, AZ) as per manufacturer's protocol with proprietary reagents. Briefly, slides were deparaffinized on the automated system with EZ Prep solution (Ventana). Heat-induced antigen retrieval method was used in RiboCC (Ven-

tana). The rabbit primary antibody that reacts to CAIX, (ab15086, Abcam, Cambridge, MA) was used at a 1:250 concentration in Dako antibody diluent (Carpenteria, CA) and incubated for 32 min. The Ventana OmniMap Anti-Rabbit Secondary Antibody was used for 20 min. The detection system used was the Ventana ChromoMap kit and slides were then counterstained with Hematoxylin. Slides were then dehydrated and cover slipped as per normal laboratory protocol.

The CAIX proteins were highlighted by brown nuclear staining of CAIX. A computational system was implemented for automated evaluation of positive staining percentage defined as the total number of positive stained pixels by total number of tumor and its immediate microenvironment pixels. Additionally, the automated algorithm calculated scores were compared with board-certified pathologist's H-scoring metric which is defined elsewhere [148].

6.2.4 Prognostic validation Dataset

The radiomics data were further validated for prognosis in four published datasets. Only OS was available for these datasets. The first dataset [14, 51] comprised of 62 adenocarcinoma patients who underwent surgical resection as first course therapy at the Moffitt Cancer Center and had pre-surgery CTs within 2 months prior to surgery. The second dataset [14, 51] comprised of 47 adenocarcinoma patients who underwent surgical resection as first course therapy at the Maastricht Radiation Oncology Clinic and had pre-surgery CTs within 2 months prior to surgery. The third dataset included 234 patients [43, 149] diagnosed with screen-detected incident lung cancers in the National Lung Screening Trial. The fourth dataset was a radiogenomics dataset [138] of 103 adenocarcinoma patients as described above.

6.2.5 Tumor segmentation and Radiomics Extraction

Figure 2.1 presents an overview of the radiomics pipeline used in this study. Pre-treatment contrast-enhanced thoracic CT scans performed ≤ 90 days (median: 10 days) prior to the initiation of immunotherapy (baseline) were retrieved from the picture archiving and communication system and loaded into HealthMyne Quantitative Imaging Decision Support (QIDS) software (<https://www.healthmyne.com>). A radiologist with more than 10 years of clinical experience selected the largest lung tumor of the patients and initialized an automated 3D segmentation algorithm using the HealthMyne® QIDS Rapid Precise Metrics software. The tumor delineation outputs of the 3D segmentation algorithm were either confirmed or edited whenever necessary by the radiologist.

The tumor mask images (i.e., tumor delineations) were imported into an in-house radiomic feature extraction toolboxes created in MATLAB® 2015b (The Mathworks Inc., Natick, Massachusetts) and C++ (<https://isocpp.org>). The CT images were resampled to a single voxel spacing of 1mm x 1mm x 1mm using cubic interpolation to standardize spacing across all images. The HUs in all CT images were then resampled into fixed bin sizes of 25 HUs discretized from -1000 to 1000 HU.

A total of 213 radiomic features were extracted utilizing the training cohort from the intratumoral region ($N = 122$ features) and the peritumoral region 3 mm outside of tumor boundary ($N = 91$ features) using standardized algorithms from the IBSI v5 [76]. Peritumoral regions were bounded by the lung parenchyma mask to exclude any tissue that exceed outside of the lung parenchyma. Unstable and non-reproducible radiomic features were eliminated utilizing methods mentioned on Chapter 5.

6.2.6 Statistical Analysis

All statistical analyses were performed using Stata/MP 14.2 (StataCorp LP, College Station, Texas) and R Project for Statistical Computing version 3.4.3 (<http://www.r->

project.org/). Differences for the clinical covariates were tested using Fisher's exact test for categorical variables and the Mann-Whitney U test for continuous variables. Survival analyses were performed using Kaplan-Meier survival estimates and the log-rank test. The OS and PFS were the two dependent variables. For OS, an event was defined as death and the data were right censored at 36-months. For PFS, an event was defined as death or either clinical or RECIST based progression of cancer and the data were right-censored at 36 months. The index date for both OS and PFS was the date of initiation of immunotherapy.

A rigorous model building approach was employed to reduce the number of covariates and identify the most informative clinical covariates and radiomic features associated with patient survival. For the clinical covariates, univariable Cox regression was performed and covariates significantly ($P < 0.05$) associated with OS were retained. To produce a parsimonious clinical model, the remaining clinical covariates were included in a stepwise backward elimination Cox regression model using a threshold of 0.01 for inclusion. For the radiomic features, univariable Cox regression was performed and radiomic features were retained that were significantly associated with OS after Bonferroni-Holm correction ($P < 0.05$). Radiomic features correlated with tumor volume (Pearson's correlation coefficient ≥ 0.80) were removed. Among the remaining radiomic features, correlated features were identified using an absolute Pearson's correlation coefficient ≥ 0.80 and the feature with the smallest p-value from the univariable analysis was retained. The remaining radiomic features, were utilized to identify a parsimonious radiomics model using a stepwise backward elimination approach applying a threshold of 0.01 for inclusion. The final covariates from the clinical model and the final features from the radiomics model were combined and CART was used to find patient risk groups. CART is a non-parametric approach modified for failure time data [110] that classifies variables through a decision tree composed of splits, or nodes, where the split points are optimized based on impurity criterion. The clinical-radiomics CART model from the training cohort was validated utilizing the test and the validation cohorts. Time-dependent AUCs and CIs were calculated for 6, 12, 24 and 36 months for training and test cohorts. The most predictive Radiomics feature was also validated in four independent cohorts.

For the radiogenomics analysis, the highest prognostic radiomic feature was compared to every gene probesets using two different approaches: correlation and two-group analysis. For the correlation analysis, gene probesets were filtered and determined as statistically significant using the following criteria: Pearson’s correlation with a threshold $|R| > 0.4$, an expression filter with max expression of gene > 5 , and an inter-quartile filter (interquartile range $> \log_2$ [1.2 fold-change]). Gene probesets were filtered and determined as significant using the following criteria based on a Student’s t test $p < 0.001$ and mean logarithmic fold-change (LFC) between high and low prognostic radiomic feature of $LFC > \log_2$ (1.4 fold-change). The significant probesets from the two analyses were intersected yielding a final list of probesets significantly associated with the prognostic radiomic feature.

6.3 Results

6.3.1 Immunotherapy Treated Patient Demographics

Type of checkpoint inhibitor, ECOG performance status, number of previous lines of therapy, serum albumin, lymphocyte counts, and NLR were significantly different between the training and test cohorts (Table 6.1). Also, significant differences were found for OS and PFS between training and test cohorts (36-month OS 32.6% vs. 19.2%, respectively; 36-month PFS 20.8% vs. 9.5%, respectively; Table 6.2) where log-rank P-value was < 0.05 (Figure 6.1).

Median age, sex, smoking status, stage, type of checkpoint inhibitor, ECOG performance status, lymphocyte counts, and NLR were significantly different between the training (MCC cohort 1) and validation cohort (VA cohort, Table 6.1). However, OS were not significantly different between the two cohorts (Figure 6.1). PFS data was not available for validation cohort.

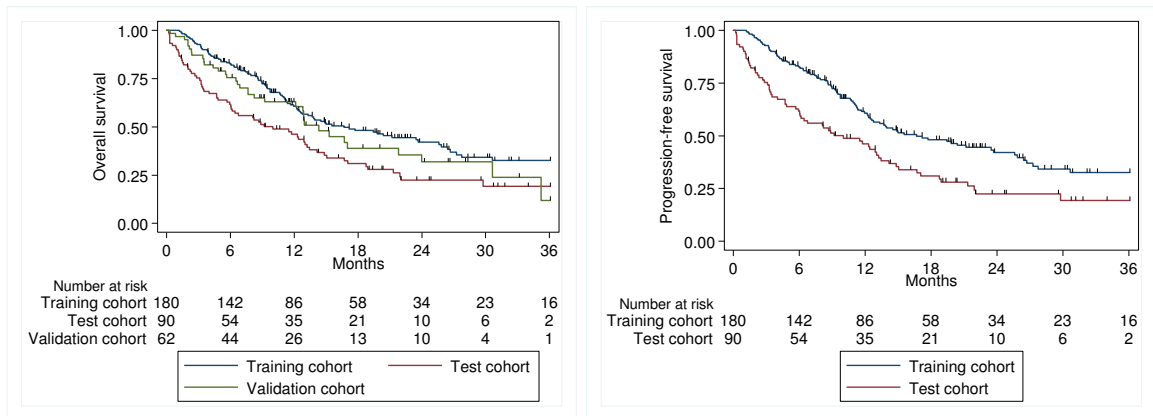


Figure 6.1 Overall survival graph for the training, test and validation cohorts and progression-free survival graph for the training and test cohorts.

6.3.2 Clinical Model

Among the 16 clinical covariates from Table 6.1 that were considered for the clinical model, four clinical features (serum albumin, number of metastatic sites, previous lines of therapy and neutrophils counts) were significantly associated with OS in univariable analysis utilizing the training cohort. The final parsimonious clinical model included two clinical features: serum albumin (HR = 0.33; 95% CI: 0.20-0.52) and number of metastatic sites (HR = 2.14; 95% CI: 1.48-3.11).

6.3.3 Radiomics Model

Among the original 213 intratumoral and peritumoral radiomic features, 67 features were found to be stable and reproducible. Eight of the 67 features were removed because they were correlated with tumor volume. Univariable analysis identified eleven features significantly associated with OS and eight of the nine features were dropped because they were correlated within each other (Figure 6.2). Among the two remaining features (GLCM inverse difference and peritumoral quartile coefficient), stepwise backward elimination approach identified GLCM inverse difference as the most informative radiomic feature (HR = 1.41; 95% CI: 1.19-1.67, $p < 0.001$).

6.3.4 CART Analysis

Based on the two most informative clinical covariates and most informative radiomic feature, CART analysis have found novel cut-off points (Figure 6.3) and classified patients in the training cohort into six risk groups (Figure A.4) which were further collapsed into four risk groups based on OS (Figure 6.4): low-risk (blue), moderate-risk (red), high-risk (green), and very high-risk (yellow). Similar findings were observed for PFS. The risk groups identified in the training cohort were also extracted in the test cohort (Table 6.2 and Figure 6.4) where the time-dependent AUCs were found to be similar for both cohorts for OS (Figure 6.5). Specifically, for 6 months OS, our model achieved an AUC of 0.784 (95% CI: 0.693 – 0.876) and for 24 months the AUC was 0.716 (95% CI: 0.558 – 0.843) for the test cohort.

6.3.5 Multivariable Analysis

A multivariable Cox regression analysis was conducted adjusting for clinical covariates that were significantly different between the training cohort and test cohort (Table 6.1). The HRs were adjusted for ECOG, lymphocyte counts and neutrophils to lymphocytes ratio (Table 6.3) and the high-risk (test cohort HR = 3.33; 95% CI 1.57 – 7.05) and very high-risk (test cohort HR = 5.35; 95% CI 2.14 – 13.36) groups were still found to be associated with significantly worse outcomes compared to the low-risk group (HR = 1.00). The results were consistent when the data were analyzed for PFS (Table 6.4). Utilizing the validation cohort and adjusting for stage, ECOG, lymphocyte count and neutrophils to lymphocytes ratio, the very-high risk group had significantly worse outcomes (HR = 13.81; 95% CI 2.58 – 73.93) compared to the low-risk group (Table 6.5).

Clinical covariates were compared across the four CART risk groups (Table 6.6) and previous lines of therapy, ECOG, white blood cell counts, neutrophils and NLR were found to be significantly different. Multivariable Cox regression was performed adjusting for these potential confounders but did not appreciably alter the HRs for risk

groups (Tables 6.3 6.4).

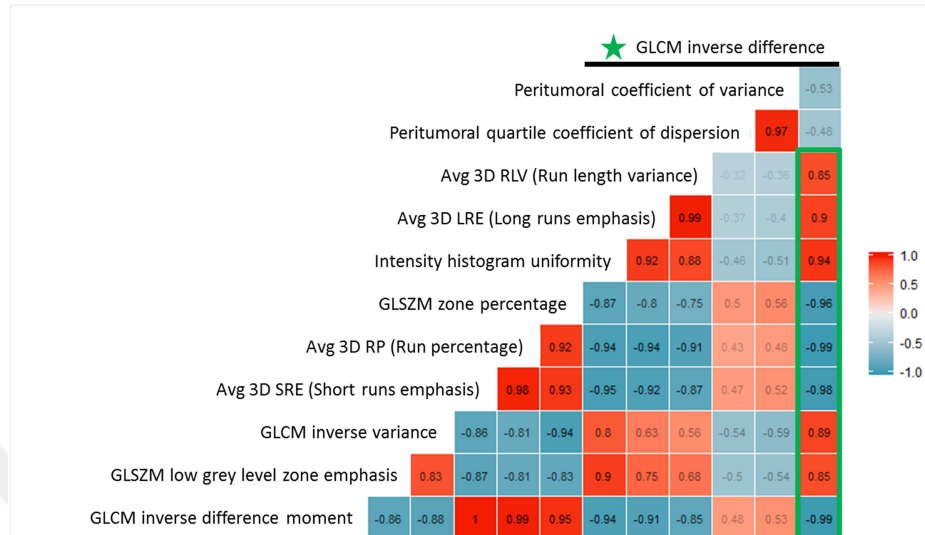


Figure 6.2 Correlation matrix for the radiomic features that were significantly associated with overall survival in the univariable analysis. The feature in the final parsimonious model was GLCM inverse difference and it is found to be correlated with nine other features shown inside the green box.

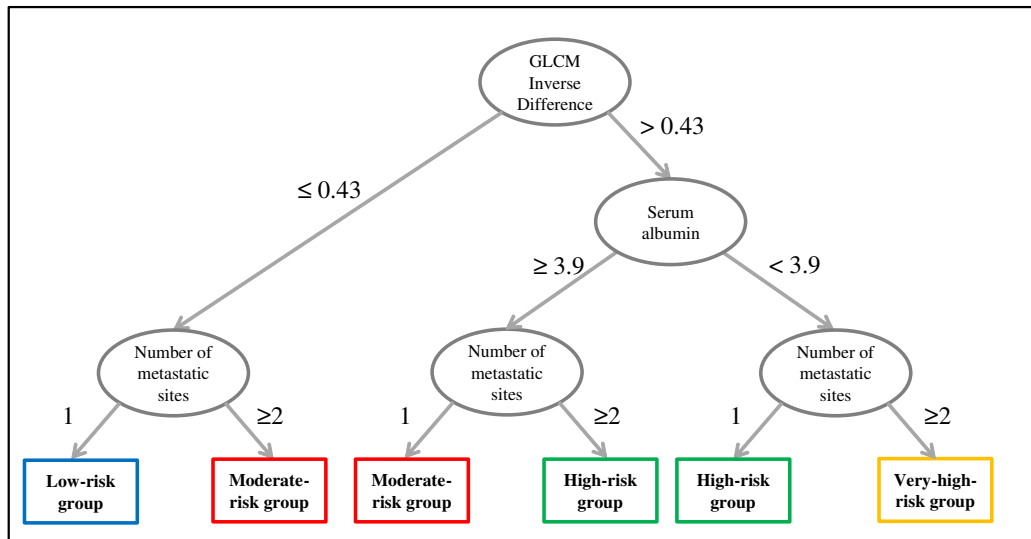


Figure 6.3 The CART was used to identify patient risk groups based on a model containing one radiomic feature and two clinical features. Patients were grouped from low risk to very-high risk based on the CART decision nodes and terminal nodes.

Table 6.1
Patient characteristics by the training and test cohorts.

Characteristic	Training Cohort (N = 180)	Test Cohort (N = 90)	P-Value	Validation Cohort (N = 62)	P-Value
Age at initiation of treatment, N (%)					
<i>Dichotomized</i>					
< 65	68 (37.8)	37 (41.1)		15 (24.2)	
≥ 65	112 (62.2)	53 (58.9)	0.599	47 (75.8)	0.063
Median, (95% CI)	67 (65-68)	67 (64-69)	0.783	68 (67-71)	0.026
Sex, N (%)					
Female	95 (52.8)	43 (47.8)		3 (4.8)	
Male	85 (47.2)	47 (52.2)	0.442	59 (95.2)	<0.001
Smoking status¹					
Never smoker	30 (16.7)	16 (17.8)		2 (3.2)	
Ever smoker	146 (81.1)	74 (82.2)	0.866	60 (96.8)	0.004
Unknown/Missing	4 (2.2)	0 (0)		0 (0)	
Stage, N (%)					
IIIb	6 (3.3)	4 (4.4)		13 (21.0)	
IV	174 (96.7)	86 (95.6)	0.735	49 (79.0)	<0.001
Histology, N (%)					
Adenocarcinoma/others	137 (76.1)	71 (78.9)		43 (69.3)	
Squamous cell carcinoma	43 (23.9)	19 (21.1)	0.648	19 (30.7)	0.135
Checkpoint inhibitors, N (%)					
Anti PD-L1	48 (26.6)	18 (20.0)		8 (12.9)	
Anti PD-1	57 (31.7)	69 (76.7)		54 (87.1)	
Doublet	75 (41.7)	3 (3.3)	<0.001	0 (0)	<0.001
ECOG performance status, N (%)					
0	39 (21.7)	10 (11.1)		12 (19.4)	
1	141 (78.3)	67 (74.4)		39 (62.9)	
2	0 (0)	13 (14.4)	<0.001	11 (17.7)	<0.001
Previous lines of therapy on current diagnosis					
None	70 (43.9)	21 (23.3)		n/a	
1	48 (26.7)	47 (52.2)		n/a	
≥ 2	62 (34.4)	22 (24.4)	<0.001	n/a	-
Number of metastatic sites					
1	82 (46.6)	51 (56.7)		25 (40.3)	
≥ 2	98 (54.4)	39 (43.3)	0.094	37 (59.7)	0.554
EGFR mutational status¹					
Not Detected	107 (59.4)	37 (41.1)		n/a	
Detected	25 (13.9)	5 (5.6)		n/a	
Missing/Inconclusive	48 (26.7)	48 (53.3)	0.355	n/a	-
KRAS mutational status¹					
Not Detected	61 (33.9)	20 (22.2)		n/a	
Detected	29 (16.1)	12 (13.3)	0.664	n/a	
Missing/Inconclusive	90 (50.0)	58 (64.4)		n/a	-
Hematology, median, (95% CI)					
Serum albumin, (g/dL)	4.0 (3.9-4.0)	3.8 (3.6-3.9)	<0.001	3.9 (3.7-4.0)	0.087
Lymphocytes, (1e+9/L)	1.3 (1.2-1.4)	1.0 (0.9-1.2)	<0.001	1.0 (0.9-1.2)	0.014
WBC, (1e+9/L)	7.1 (6.7-7.6)	7.7 (6.8-8.8)	0.246	7.5 (6.7-8.7)	0.383
Neutrophils, (1e+9/L)	4.8 (4.4-5.1)	5.3 (4.6-6.5)	0.131	5.6 (4.8-6.1)	0.329
NLR	3.7 (3.2-4.1)	5.2 (4.0-7.5)	0.002	5.3 (4.1-6.8)	0.004

Abbreviations: CI = confidence interval; NLR = neutrophils to lymphocytes ratio;

Bold P-values are statistically significant.

¹P-values for smoking status, EGFR mutational status and KRAS mutational status were calculated for patients without missing/inconclusive data.

Table 6.2
Overall survival and progression free survival rates by training and test cohorts and patient risk groups.

	Percent survival at:			
	6	12	24	36
	months	months	months	months
<u>Overall survival</u>				
<i>Overall by cohort</i>				
Training Cohort	82.70%	60.80%	42.10%	32.60%
Test Cohort	61.70%	46.20%	22.40%	19.20%
<i>By risk group</i>				
Low-risk				
Training Cohort	100%	95.20%	84.70%	84.70%
Test Cohort	95.00%	85.00%	38.90%	38.90%
Moderate-risk				
Training Cohort	92.60%	76.40%	59.70%	47.90%
Test Cohort	67.80%	56.70%	33.10%	n/a
High-risk				
Training Cohort	81.10%	54.40%	24.90%	15.60%
Test Cohort	62.10%	34.10%	17.10%	8.50%
Very-high-risk				
Training Cohort	59.90%	24.30%	12.20%	0%
Test Cohort	16.70%	11.10%	0%	0%
<u>Progression-free survival</u>				
<i>Overall by cohort</i>				
Training Cohort	47.90%	32.80%	22.80%	20.80%
Test Cohort	37.90%	19.60%	9.50%	9.50%
<i>By risk group</i>				
Low-risk				
Training Cohort	71.40%	71.40%	65.50%	65.50%
Test Cohort	73.70%	46.30%	29.80%	29.80%
Moderate-risk				
Training Cohort	65.90%	43.00%	31.20%	25.00%
Test Cohort	23.90%	19.10%	9.60%	n/a
High-risk				
Training Cohort	43.00%	27.00%	9.80%	9.80%
Test Cohort	41.30%	15.00%	3.80%	n/a
Very-high-risk				
Training Cohort	15.70%	0%	0%	0%
Test Cohort	11.10%	0%	0%	0%

Cells were marked as n/a whenever all of the patients were censored for the given interval.

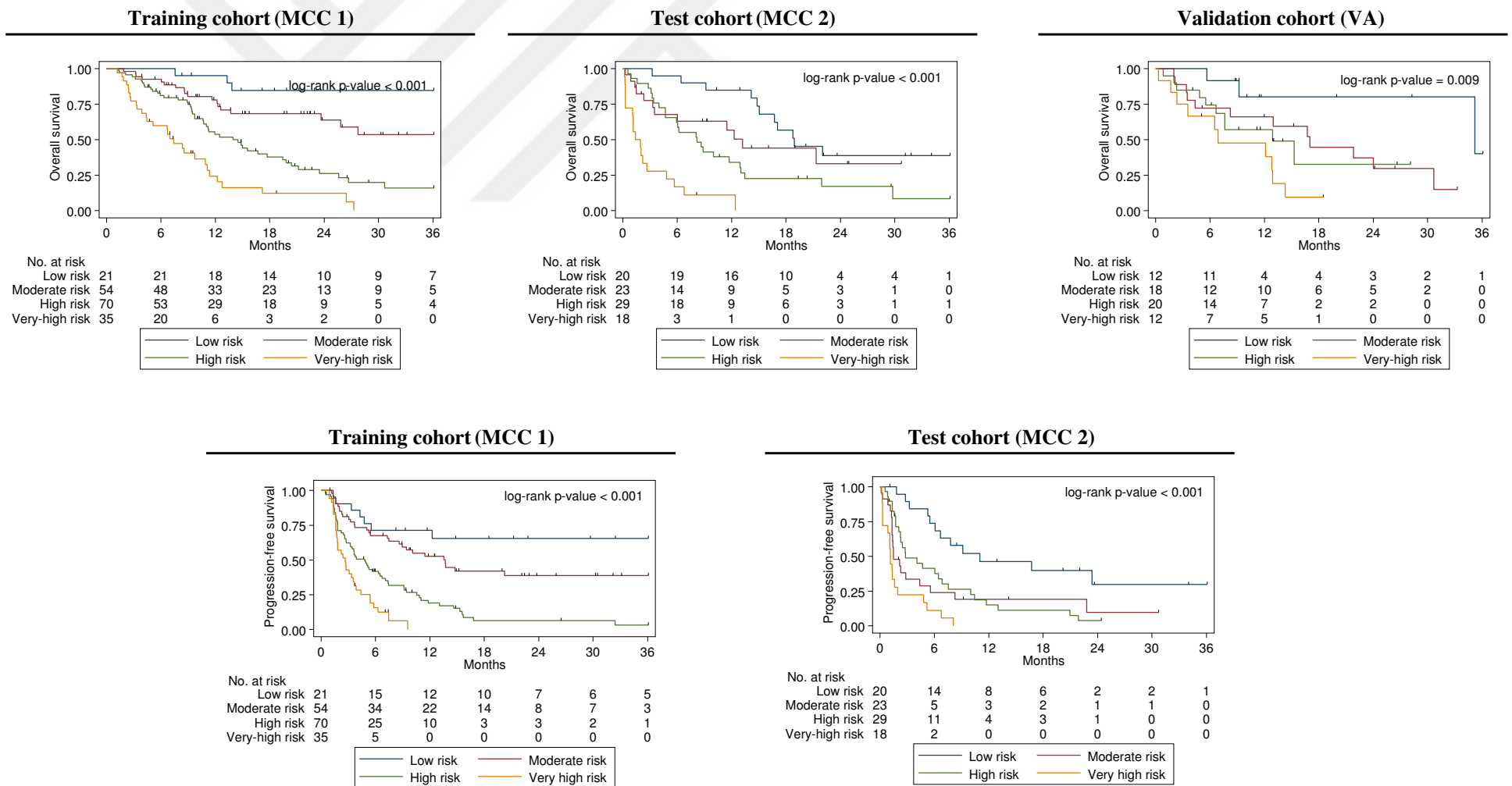


Figure 6.4 Kaplan-Meier survival curves estimates for overall survival (top) in the training (left), test (middle) and validation cohorts (right), and progressive-free survival (bottom) in the training (left) and test cohorts (right).

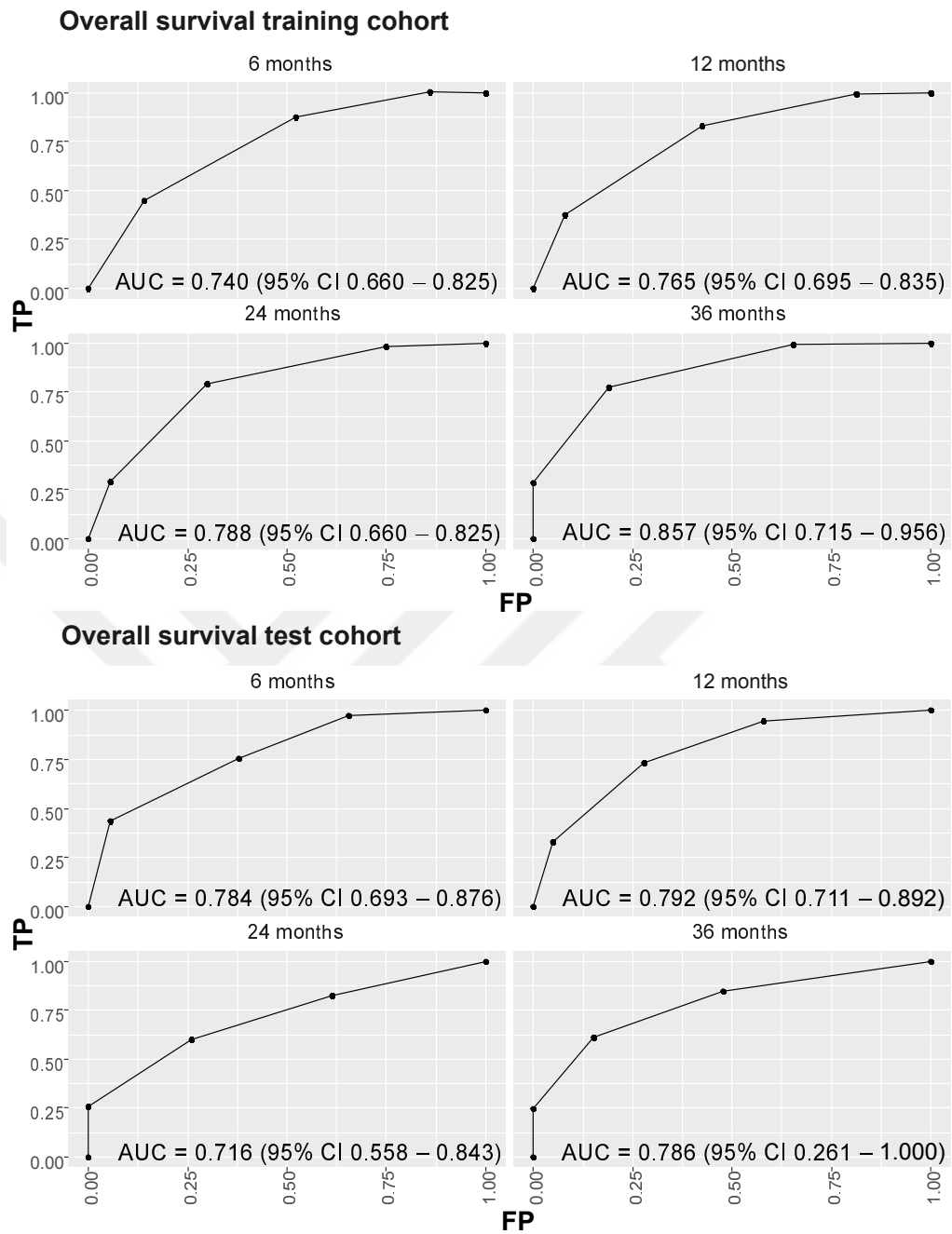


Figure 6.5 Time-dependent AUC curves for Cox regression models based on 6, 12, 24 and 36 months for training (top) and test cohorts (bottom). The AUC values were statistically not different between training and test cohorts.

6.3.6 Radiogenomics Analysis

For two-group analysis, GLCM inverse difference was dichotomized at the previously determined CART threshold (0.43), which was similar to the mean (0.47) and

Table 6.3

Univariable and multivariable Cox regression analysis for overall survival for the training and test cohorts.

	Training cohort (MCC 1)			Test cohort (MCC 2)	
	Univariable Model ¹	Multivariable Model ²	Multivariable Model ³	Univariable Model ¹	Multivariable Model ²
	HR (95% CI)	HR (95% CI)	HR (95% CI)	HR (95% CI)	HR (95% CI)
Overall survival					
Risk group					
Low-risk	1.00 (Reference)	1.00 (Reference)	1.00 (Reference)	1.00 (Reference)	1.00 (Reference)
Moderate-risk	3.79 (1.13 – 12.68)	3.08 (0.89 – 10.66)	3.56 (1.02 – 12.48)	1.70 (0.75 – 3.87)	1.51 (0.66 – 3.51)
High-risk	8.02 (2.47 – 26.09)	7.87 (2.38 – 25.97)	6.98 (2.10 – 23.18)	2.73 (1.33 – 5.63)	3.33 (1.57 – 7.05)
Very-high-risk	19.32 (5.80 – 64.32)	17.33 (5.11 – 58.72)	17.24 (5.09 – 58.36)	10.52 (4.58 – 24.17)	5.35 (2.14 – 13.36)
ECOG	.	1.22 (0.70 – 2.11)	1.20 (0.69 – 2.07)	.	2.63 (1.47 – 4.68)
Pr. treatment	.	.	1.36 (1.01 – 1.81)	.	.
Lymphocytes	.	1.04 (0.74 – 1.46)	.	.	0.73 (0.45 – 1.17)
WBC	.	.	0.98 (0.88 – 1.09)	.	.
Neutrophils	.	.	1.10 (0.89 – 1.34)	.	.
NLR	.	1.01 (0.97 – 1.06)	0.98 (0.92 – 1.05)	.	1.05 (1.02 – 1.08)

Abbreviations: SD = standard deviation; HR = hazard ratio; CI = confidence interval; PFS = progression-free survival; NLR = neutrophils to lymphocytes ratio; WBC = white blood cell; Pr. treatment = previous lines of treatments at current diagnosis

Bold values are statistically significant.

¹The main effects for each risk group with the low risk group as the referent category.

²These models included the clinical covariates that were found to be significant different between the training and test cohorts (Table 6.1) and the risk groups using the low risk group as the referent category.

³These models included the clinical covariates that were found to be significant different between the CART risk groups (Table 6.5).

median (0.45) values in the radiogenomics dataset. Correlation and two-group analyses identified 123 significant probesets representing 91 unique genes that were associated with the GLCM inverse difference radiomic feature. Interestingly, only three probesets (representing two genes) were positively associated with GLCM inverse difference: CAIX and Family With Sequence Similarity 83 Member F (FAM83F). GLCM inverse difference was positively associated with CAIX expression based on two different probesets (Figure 6.6A-D). Median CAIX expression was lower for patients with low GLCM inverse difference (< 0.43) (merck2 DQ892208: 4.61 [95% CI: 4.38 – 5.00]; merck NM001216: 4.48 [95% CI: 4.24 – 4.62]) vs. high GLCM inverse difference (≥ 0.43) (merck2 DQ892208: 6.32 [95% CI: 5.50 – 6.86]; merck NM 001216: 5.66 [95% CI: 5.11 – 6.39]).

Table 6.4

Univariable and multivariable Cox regression analysis for progression-free survival for the training and test cohorts.

	Training cohort (N = 180)			Test cohort (N = 90)	
	Univariable Model ¹	Multivariable Model ²	Multivariable Model ³	Univariable Model ¹	Multivariable Model ²
	HR (95% CI)	HR (95% CI)	HR (95% CI)	HR (95% CI)	HR (95% CI)
Progression-free survival					
Risk group					
Low-risk	1.00 (Reference)	1.00 (Reference)	1.00 (Reference)	1.00 (Reference)	1.00 (Reference)
Moderate-risk	2.02 (0.89 – 4.64)	2.05 (0.88 – 4.76)	2.36 (1.00 – 5.58)	2.96 (1.43 – 6.14)	2.80 (1.34 – 5.85)
High-risk	5.15 (2.33 – 11.36)	5.55 (2.46 – 12.49)	4.89 (2.15 – 11.14)	2.58 (1.29 – 5.14)	3.05 (1.50 – 6.18)
Very-high-risk	9.62 (4.12 – 22.44)	9.03 (3.77 – 21.63)	8.79 (3.66 – 21.11)	7.13 (3.31 – 15.35)	3.95 (1.56 – 8.54)
ECOG	.	1.09 (0.68 – 1.74)	1.05 (0.66 – 1.68)	.	2.33 (1.35 – 4.03)
Prv treatment	.	.	1.32 (1.04 – 1.67)	.	.
Lymphocytes	.	0.83 (0.63 – 1.09)	.	.	0.88 (0.59 – 1.33)
WBC	.	.	1.00 (0.90 – 1.11)	.	.
Neutrophils	.	.	1.04 (0.86 – 1.26)	.	.
NLR	.	1.04 (0.99 – 1.09)	1.01 (0.95 – 1.07)	.	1.05 (1.02 – 1.08)

Abbreviations: SD = standard deviation; HR = hazard ratio; CI = confidence interval; PFS = progression-free survival; NLR = neutrophils to lymphocytes ratio; WBC = white blood cell; Pr. treatment = previous lines of treatments at current diagnosis

Bold values are statistically significant.

¹The main effects for each risk group with the low risk group as the referent category.

²These models included the clinical covariates that were found to be significant different between the training and test cohorts (Table 1) and the risk groups using the low risk group as the referent category.

³These models included the clinical covariates that were found to be significant different between the CART risk groups (Sup Table 2).

6.3.7 Immunohistochemistry Analysis

Further investigating the relationship between CAIX expression of tumors and tumor GLCM inverse difference radiomics showed that patients with higher CAIX expression had a trend towards higher GLCM inverse difference (Figure 6.6E). Automated pathology scoring algorithm was compared with the pathologist scored H-score and shown to have high correlation (Figure 6.6F). Representative cases of patients with high and low GLCM inverse difference and pathologically scored high and low CAIX expressions are shown in (Figure 6.7).

Table 6.5

Univariable and multivariable Cox regression analysis for overall survival for the validation cohort.

	Validation cohort (VA)	
	Univariable Model ¹	Multivariable Model ²
	HR (95% CI)	HR (95% CI)
Overall survival		
Risk group		
Low-risk	1.00 (Reference)	1.00 (Reference)
Moderate-risk	4.07 (0.90 – 18.26)	4.00 (0.83 – 19.17)
High-risk	4.72 (1.02 – 21.94)	4.54 (0.90 – 23.11)
Very-high-risk	9.72 (2.08 – 45.49)	13.81 (2.58 – 73.93)
Age	.	1.00 (0.96 – 1.05)
Stage	.	0.66 (0.23 – 1.93)
ECOG	.	1.98 (0.99 – 3.94)
Lymphocytes	.	1.13 (0.53 – 2.40)
NLR	.	1.09 (0.99 – 1.18)

Abbreviations: SD = standard deviation; HR = hazard ratio; CI = confidence interval; ECOG = Eastern Cooperative Oncology Group; NLR = neutrophils to lymphocytes ratio;

Bold values are statistically significant.

¹The main effects for each risk group with the low risk group as the referent category.

²These models included the clinical covariates that were found to be significant different between the training and test cohorts (Table 6.1) and the risk groups using the low risk group as the referent category.

6.3.8 Prognostic Validation Datasets

GLCM inverse difference was significantly associated with OS in three out of the four independent NSCLC cohorts (Figure 6.8) using previously found CART cut-point (0.43). Although the a priori cut-point for GLCM inverse difference was not significantly associated with OS in the Maastricht patient cohort, GLCM inverse difference as a continuous covariates was significantly associated with OS in a Cox regression model (HR = 2.74; 95% CI 1.04 – 7.24).

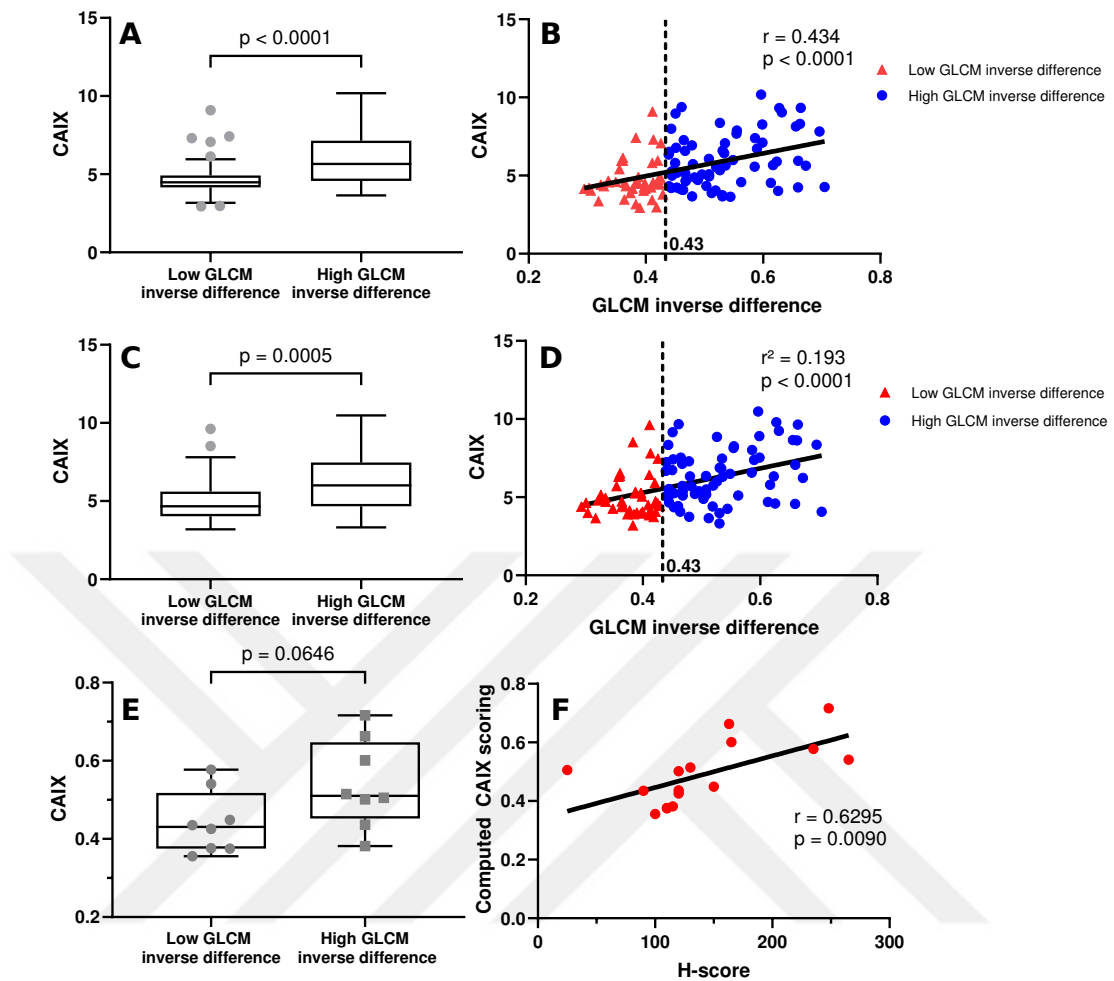


Figure 6.6 Association between GLCM inverse difference CT radiomic feature and high CAIX expression. **a)** Whisker-box plots representing the association between CAIX expression using merck 2DQ892208 probset and GLCM inverse difference. High and low GLCM inverse difference was found using novel cut-point defined by CART analysis **b)** Scatter plot showing the linear relationship between CAIX expression using merck 2DQ892208 probset and GLCM inverse difference. CART defined cut-off point was used to differentiate high (blue) and low (red) GLCM inverse difference. **c)** Whisker-box plots representing the association between CAIX expression using merck NM001216 probset and GLCM inverse difference. High and low GLCM inverse difference was found using novel cut-point defined by CART analysis. **d)** Scatter plot showing the linear relationship between CAIX expression using merck NM001216 probset and GLCM inverse difference. CART defined cut-off point was used to differentiate high (blue) and low (red) GLCM inverse difference. **e)** Whisker-box plots representing the association between CAIX expression on IHC staining and GLCM inverse difference CT radiomic feature. High and low GLCM inverse difference was found using novel cut-point defined by CART analysis. **f)** Scatter plot showing linear relationship between pathologist H-score for CAIX and computer derived (Aperio positive pixel count algorithm) automated CAIX scoring.

6.4 Discussion

Predictive biomarkers that identify lung cancer patients who will experience rapid and lethal outcomes is a critical unmet need as such, patients could avoid in-

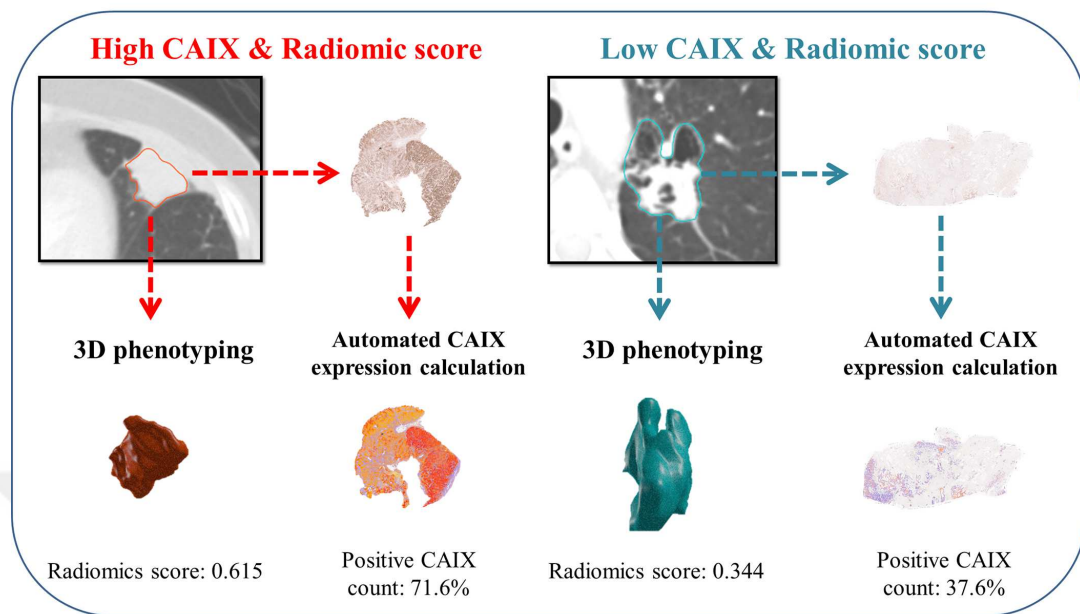


Figure 6.7 Representative cases for testing the agreement between GLCM inverse difference and CAIX IHC expression. Correlation between high CAIX and high CT radiomic feature is seen on left side and correlation between low CAIX and low CT radiomic feature is seen on right side.

effective and expensive treatment. In this study, a rigorous radiomics pipeline and a robust analysis was utilized to identify and successfully test and validate a parsimonious clinical-radiomic model that was significantly associated with survival outcomes and stratified patients into four unique risk groups based on risk of patient death and risk of progression. The very high-risk group was associated with extremely poor OS in all the training, test and independent validation cohorts (Figure 6.4) which may suggest these patients should either avoid immunotherapy altogether or utilize upfront combination treatments that may yield a better response. The most informative radiomic feature, GLCM inverse difference, was positively associated with CAIX expression and further validation demonstrated that GLCM inverse difference was also associated with OS in four independent NSCLC cohorts.

Table 6.6
Patient characteristics by CART risk groups for the training cohort.

Characteristic	Low risk	Moderate risk	High risk	Very-high risk	P-Value
Age at diagnosis, N (%)					
<i>Dichotomized</i>					
< 65	9 (42.9)	22 (40.7)	26 (37.1)	11 (31.4)	
≥ 65	12 (57.1)	32 (59.3)	44 (62.9)	24 (68.6)	0.798
Sex, N (%)					
Female	8 (38.1)	31 (57.4)	40 (57.1)	16 (45.7)	
Male	13 (61.9)	23 (42.6)	30 (42.9)	19 (54.3)	0.323
Smoking status¹					
Never smoker	4 (19.1)	10 (18.9)	12 (17.9)	4 (11.4)	
Ever smoker	17 (80.9)	43 (81.1)	55 (82.1)	31 (88.6)	0.809
Stage, N (%)					
III	2 (9.5)	0 (0)	3 (4.3)	1 (2.9)	
IV	19 (90.5)	54 (100)	67 (95.7)	34 (97.1)	0.138
Histology, N (%)					
Adenocarcinoma/others	17 (81.0)	43 (79.6)	53 (75.7)	24 (68.6)	
Squamous cell carcinoma	4 (19.0)	11 (20.4)	17 (24.3)	11 (31.4)	0.636
Checkpoint inhibitors, N (%)					
Anti PD-L1	4 (19.1)	11 (20.37)	24 (34.3)	9 (25.7)	
Anti PD-1	7 (33.3)	16 (29.6)	25 (35.7)	9 (25.7)	
Doublet	10 (47.6)	27 (50.0)	21 (30.0)	17 (48.6)	0.285
ECOG performance status, N (%)					
0	10 (47.6)	10 (18.5)	15 (21.4)	4 (11.4)	
1	11 (52.4)	44 (81.5)	55 (78.6)	31 (88.6)	0.021
Previous lines of therapy on current diagnosis					
None	10 (47.6)	33 (61.1)	10 (14.3)	17 (48.6)	
1	4 (19.1)	13 (24.1)	24 (34.3)	7 (20.0)	
≥ 2	7 (33.3)	8 (14.8)	36 (51.4)	11 (31.4)	<0.001
Number of metastatic sites					
1	21 (100)	14 (25.9)	47 (67.1)	0 (0)	
≥ 2	0 (0)	40 (74.1)	23 (32.9)	35 (100)	<0.001
EGFR mutational status¹					
Not Detected	14 (77.8)	36 (87.8)	37 (75.5)	20 (83.3)	
Detected	4 (22.2)	5 (12.2)	12 (24.5)	4 (16.7)	0.495
KRAS mutational status¹					
Not Detected	7 (58.3)	17 (60.7)	26 (70.3)	11 (84.6)	
Detected	5 (41.7)	11 (39.3)	11 (29.7)	2 (15.4)	0.401
Hematology, median, (95% CI)					
Serum albumin, (g/dL)	4.0 (3.8-4.2)	4.1 (4.1-4.2)	4.0 (3.9-4.1)	3.6 (3.5-3.7)	<0.001
Lymphocytes, (1e+9/L)	1.0 (0.8-1.4)	1.2 (1.2-1.4)	1.4 (1.3-1.5)	1.2 (0.8-1.6)	0.215
WBC, (1e+9/L)	6.8 (5.1-8.8)	6.9 (6.4-8.0)	6.9 (6.4-7.4)	8.3 (7.4-10.9)	0.023
Neutrophils, (1e+9/L)	4.8 (3.7-6.4)	4.7 (4.1-5.3)	4.4 (3.9-4.9)	6.1 (5.1-7.4)	0.007
NLR	4.1 (2.7-5.7)	3.4 (2.8-4.0)	3.1 (2.8-3.7)	4.6 (3.8-7.0)	0.004

Abbreviations: CI = confidence interval; NLR = neutrophils to lymphocytes ratio;

Bold P-values are statistically significant and P-values for continuous variables were calculated using Kruskal-
¹P-values for Smoking status, EGFR mutational status and KRAS mutational status were calculated for patients without missing/inconclusive data.

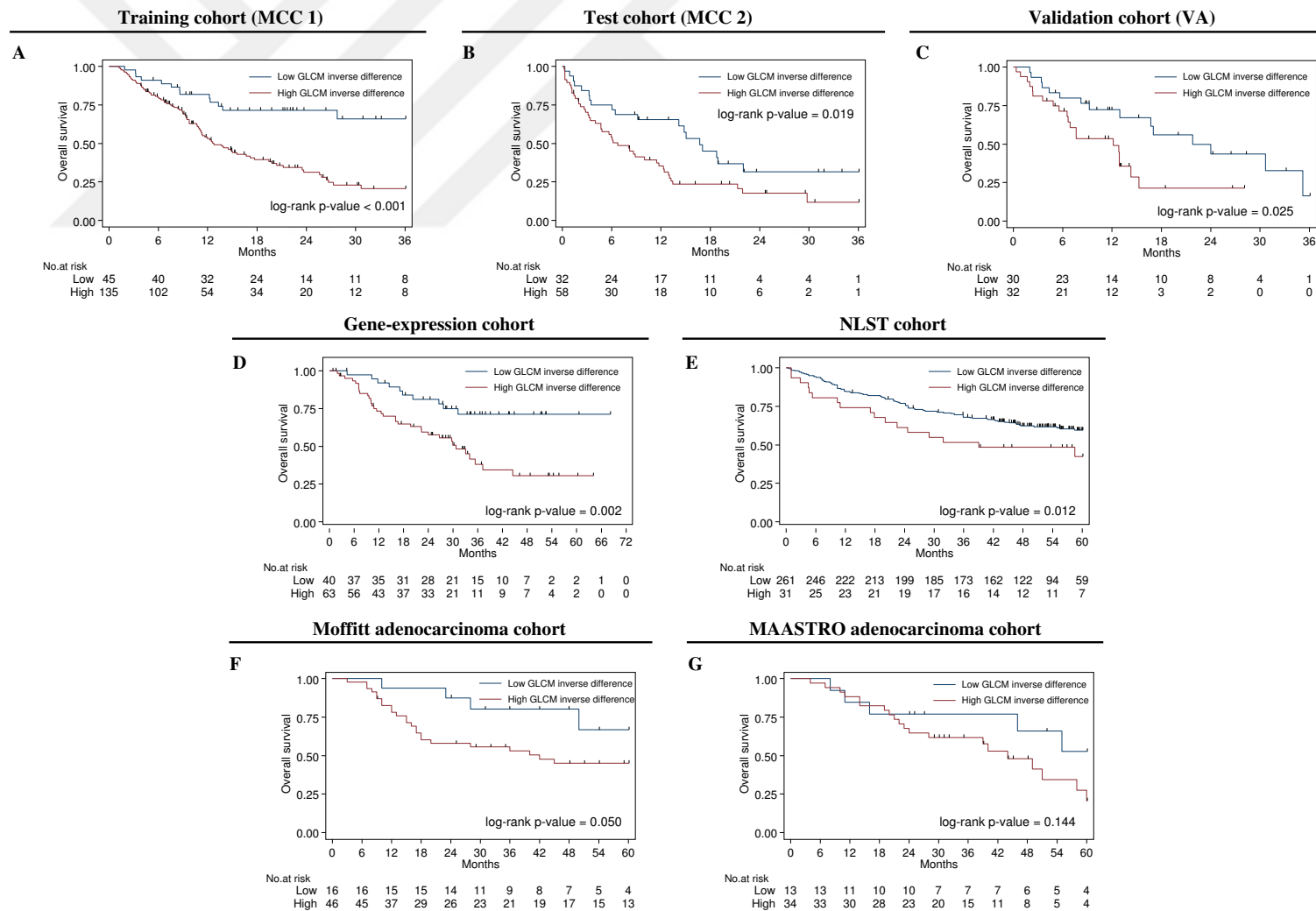


Figure 6.8 Kaplan-Meier survival plots of patients dichotomized by radiomics score. Same cut-off point was used for dichotomizing the cohorts. **a)** Training cohort **b)** Test cohort **c)** Validation cohort, **d)** Gene-expression cohort **e)** NLST cohort **f)** Moffitt adenocarcinoma cohort **g)** MAASTRO adenocarcinoma cohort.

The four final risk groups found in this study were derived from one radiomic feature (GLCM inverse difference) and two clinical covariates (number of metastatic sites and serum albumin). Higher GLCM inverse difference was associated with poor outcomes in four other prognostic validation NSCLC cohorts suggesting a pan-radiomic feature. The GLCM inverse difference is an “avatar feature” that is correlated with nine other radiomic features (Figure 6.2). Dense and uniform lesions were found less likely to respond to treatments as tumors with higher GLCM inverse difference were reflecting this phenotype (Figure A.5). Furthermore, this analyses revealed that the avatar feature is associated to tumor hypoxia since it was positively associated with CAIX expression which is an important pH regulatory enzyme that is upregulated in hypoxic tumors leading to an acidic tumor microenvironment [150] and associated with poor cancer prognosis [151, 152] including NSCLC [153, 154]. Tumor-hypoxia leads to advanced but dysfunctional vascularization and acquisition of epithelial-mesenchymal transition phenotype, resulting in cell mobility and metastasis and alters cancer cell metabolism and contributes to therapy resistance by inducing cell quiescence and immunosuppressive phenotype [155]. The most predictive clinical covariates in this study demonstrate the utility of standard-of-care clinical information to predictive treatment response. Higher number of metastatic sites increases disease burden and can result in mixed responses where one or more lesions may be responding while others are progressing and ultimately resulting in progressive disease. The other clinical covariate, serum albumin, has been shown to be associated with survival in NSCLC patients [156, 157] and is used in cancer prognostic scores including RMH prognostic score [118] and MDA risk score [158]. Lower serum albumin is an indicator of malnutrition, inflammation, and hepatic dysfunction which may lead to worse outcomes. The mechanism of serum albumin in related to immunotherapy response is not yet established yet.

Emerging evidence demonstrates the utility of radiomics as a non-invasive approach to quantify and predict lung cancer treatment response of tyrosine kinase inhibitors [159, 160], platinum-based chemotherapy [63], neo-adjuvant chemo-radiation [60, 161], stereotactic body radiation therapy [55, 62], and immunotherapy [65, 66, 131]. With respect to immunotherapy treatment response, our analysis on Chapter 4 demonstrated that pre-treatment clinical covariates and radiomic features predicted rapid

disease progression phenotypes, including HPD (AUROCs ranging 0.804-0.865) among 228 NSCLC patients treated with single agent or double agent immunotherapy. Sun *et al.* [131] developed and validated a radiomic signature for CD8 cells that predicted clinical outcomes (AUC = 0.67) among 135 patients spanning 15 different cancer types treated with anti-PD-1 or anti-PD-L1 NSCLC patients only represented 22% of their dataset. Trebeschi *et al.* [66] developed a machine learning model that significantly discriminated progressive disease from stable and responsive disease (AUC = 0.83) among 123 NSCLC patients treated with anti-PD1 immunotherapy. The study presented here represents the single largest study population of NSCLC patients treated with immunotherapy.

We acknowledge some limitations of this study. The TMB data is not available for any of the patients and PD-L1 IHC data was only available for 8 patients (4.4%) in the training cohort, 29 patients (32.2%) in the test cohort and for any of the patients in the validation cohort. Thus, the performance of PD-L1 status to predict patient outcomes was not evaluated in this study. However, recent studies have shown that patients respond to immunotherapy regardless of PD-L1 expression [7,10], so inclusion of PD-L1 status may add little or no improvement to predict patient outcomes in our models. Also, there were significant OS differences between the training, the test and the validation cohorts (Figure 6.1) as majority of the patients in the test cohort were treated with standard-of-care immunotherapy while the training cohort comprised of clinical trial patients who usually have better overall performance status. However, the very-high risk groups to immunotherapy found by the developed models were replicated in both test and validation cohorts. The IHC analysis were done for a small group of patients (N = 16) due to limited resources. However, an automated method was used to score IHCs as pathologist H-score metric could lack inter- and intra-variability [148]. Despite these minor weaknesses, this study yields a high radiomic quality score (RQS = 17) which is a stringent metric that quantifies the clinical relevance of a radiomic study [85].

In conclusion, using standard-of-care imaging and clinical covariates a novel parsimonious model was identified and validated that predicts OS and PFS among NSCLC

patients treated with immunotherapy. The prognostic image-based (i.e., GLCM inverse difference) feature was found to be associated with CAIX, an important enzyme up-regulated in hypoxic and acidotic tumors which is related to treatment resistance. The potential clinical application of this work is that baseline radiomics and clinical covariates can identify patients that are unlikely to respond to immunotherapy.



7. OVERALL CONCLUSIONS

Radiomics provides powerful image-based biomarkers that have been successfully applied for cancer detection, diagnosis, prognosis, prediction of response to treatment, and monitoring of disease status by converting standard-of-care medical images into quantitative data [12, 17, 43, 50]. The major aim of this work was to develop radiomic models that are associated with NSCLC patient responses in the setting of checkpoint blockade immunotherapy. This is particularly important as currently, PD-L1 expression extracted from IHC is the only applied biomarker to select patients for immunotherapy; however, PD-L1 expression alone is not adequate to predict response [8, 9]. In fact, recent clinical trials demonstrated that combination immunotherapy (i.e., chemotherapy plus immunotherapy) exhibits survival benefit regardless of PD-L1 expression [7, 10]. The work presented in this thesis provided generalizable radiomic-clinical based multivariable models that were able to predict immunotherapy response. The models created were developed using rigorous radiomic pipelines and biological underpinnings of image-based features that were explored within the boundaries of resources available. Following is a summary of final results obtained in this thesis along with the scientific novelty of the manuscripts that are either published or under review.

7.1 Radiomic Features Extracted From Radial Gradient and Radial Deviation Maps Associated with Survival among Lung Adenocarcinoma Patients (Oncotarget 2018)

In this study, the main hypothesis was that image-based features extracted from RG and RD masks of chest CT exams were predictive of NSCLC patient overall survival. Additionally, the potential biological underpinnings of these features were explored by analyzing the association between RG and RD image features with semantic radiological features.

After eliminating non-reproducible and redundant features, parsimonious models were created where two highly informative RG/RD features were found to be associated with lung cancer survival. One of the two features (RD outside-border separation SD) was replicated and found to be significantly associated with OS in a separate external cohort.

By quantifying and analyzing these differences, as performed in this study, we have shown that RD/RG features may have clinical utility by differentiating patients with an aggressive disease and poor patient outcomes versus patients with more indolent disease and improved outcomes. Semantic radiological analyses revealed that RG/RD features were associated with potential radiologist defined tumor phenotypes. Specifically, three RG/RD features were significantly associated with tumor lobulation, pleural attachment, and border definition (Table 3.7). The replicated feature was significantly associated with border definition semantic feature which has been previously shown to be a prognostic for lung cancer [15].

Overall, this study identified RG and RD image features that were significantly associated with lung cancer survival even after adjusting for clinical covariates. The combinatorial associations of two features were able to differentiate patients with aggressive disease versus patients with indolent disease, and this was replicated in the test cohort. As such, these findings may have clinical utility to sub-stratify patients based on clinical outcome and identify patients that may need more that aggressive treatment such as neo-adjuvant chemotherapy and aggressive follow-up and management. As future work, these features can be used in other scanner settings and cancer histologies. As part of this thesis, these features were used in the work presented in Chapter 4 as rapid disease progression phenotype predictors.

7.2 Novel Clinical and Radiomic Predictors Of Rapid Disease Progression Phenotypes among Lung Cancer Patients Treated With Immunotherapy (Lung Cancer, 2019)

In this study, the main hypothesis was to identify clinical and computational image-based predictors of rapid disease progression phenotypes in NSCLC patients treated with immune-checkpoint blockades. Rapid disease progression phenotypes were based on time-to-progression and/or tumor growth rates. An important recent phenomenon called hyperprogressive disease, an accelerated and lethal progression, among immunotherapy treated patients was also assessed. Rapid disease progression phenotypes are important to predict on baseline (pre-treatment) as currently there are no robust or readily available biomarkers that can predict immunotherapy outcomes. Because of the complexity in objective immunotherapy response, including hyperprogressive disease, pseudo-progression, and acquired resistance, there is a pressing challenge to identify biomarkers to predict patients that are least likely to respond to avoid unnecessary and potentially toxic treatments.

Novel parsimonious models containing highly informative clinical data and radiomic features to predict rapid disease progression phenotypes of NSCLC patients with modest to high AUROCs ranging from 0.804 to 0.865 (Figure 4.4) and accuracies ranging from 73.4% to 82.3% (Table 4.7). Clinical covariates alone have shown to have modest performance in predicting immunotherapy treatment response. However, substantial improvements in the AUROCs were observed when clinical and radiomic features were combined. The clinical covariates in the final model were presence of hepatic and bone metastasis, previous lines of systemic therapies, and NLR. Three of the four radiomic features (RG border SD-2D, border quartile coefficient of dispersion, border 3D Laws E5E5L5) were extracted from the border regions of the tumors, which is the immediate outside of the tumors that may be capturing data related to the TME. Also RG border SD-2D feature has previously shown to be associated with survival on Chapter 3. Although, modest to high AUROCs were found, further research is needed to elucidate the mechanisms of rapid disease progression and HPD and of the biology

of the radiomic features that predicts disease outcomes.

Overall, novel models containing informative clinical covariates and radiomic image features to predict rapid disease phenotypes including HPD were identified in this work. At the time this work was performed, independent test and validation cohorts were not available. Hence, additional research with independent test and validation are needed to demonstrate the clinical utility of these findings and biological underpinnings of the predictive radiomic features found should be investigated. Nevertheless, this proof-of-concept study showed that radiomics have the potential to be utilized as a complementary biomarker to clinical data and provide orthogonal information in the setting of lung cancer immunotherapy.

7.3 Stability and Reproducibility of Computed Tomography Radiomic Features Extracted from Peritumoral Regions of Lung Cancer Lesions (Medical Physics, 2019)

In this study, the main hypothesis was to assess the stability and reproducibility of peritumoral and intratumoral radiomic features of lung lesions captured by thoracic CT scans. Although intratumoral radiomic features were assessed for stability and reproducibility previously, no such study was done utilizing peritumoral radiomic features. This study also differs from prior work conducted on intratumoral radiomics in that the majority of the radiomic features that were evaluated were standardized through algorithms defined by IBSI [76]. Finally, the clinical utility of using such features that are stable and reproducible in relation to lung cancer survival were assessed.

To measure stability the “Moist run” dataset [19] from The Cancer Imaging Archive was used and, to measure reproducibility the RIDER dataset that consists of test-retest data [20] was used. Peritumoral ROIs with incremental distances of 3 mm to 12 mm from the tumor boundary were generated by applying morphological image processing operations on tumor segmentation masks. A subset of peritumoral features

that were stable and reproducible were found. Features found to be stable regardless of the peritumoral distances included statistical and histogram and a subset of texture features (GLCM, GLRLM, GLSZM and NGTDM). This suggests these features are less affected by changes in the ROIs. A majority of the 3D Laws and wavelet texture features were inconsistent across the peritumoral regions, suggesting instability. When the clinical utility of stable and reproducible peritumoral radiomic features were assessed, the most prognostic feature found was also a stable and reproducible feature (Figure 5.6). Interestingly, a majority of the intratumoral features were unstable, especially when extracted using different segmentation algorithms (Figure 5.4). Finally, there were no significant differences in stability for features that were extracted from ROIs bounded by a lung parenchymal mask versus ROIs that were not bounded by a lung parenchymal mask. However, the clinical utility of including extra-pleural tissues into the ROI is currently unknown and future work needs to be done to clarify.

Overall, stable and reproducible IBSI standardized peritumoral and intratumoral features were identified. Because the peritumoral region has unique clinical and biological significance, capturing this information using radiomic analyses has tremendous translational utility as demonstrated from previous studies and this study. The features found in this analysis were further used on Chapter 6 to reduce the chance of spurious findings.

7.4 Clinical Factors and Quantitative Image-Based Features Predict Immunotherapy Response among Lung Cancer Patients (Under review)

In this study, the main hypothesis to be tested that pre-treatment clinical data and radiomic features extracted from CT scans can be used to develop a parsimonious model to predict survival outcomes among NSCLC patients treated with immunotherapy. The biological underpinnings of the radiomics features were assessed utilizing gene-expression information from a well-annotated radiogenomics NSCLC dataset and

were further assessed for survival in four independent internal and external NSCLC cohorts.

Parsimonious models that were associated with survival were further used to stratify patients into unique risk groups based on risk of patient death and progression. The very high-risk group was associated with extremely poor OS and PFS in all the training, test and independent validation cohorts (Figure 6.4). This finding suggests that these patients should either avoid immunotherapy altogether or utilize upfront combination treatments that may yield a better response. Importantly, the most important radiomic feature, GLCM inverse difference, was positively associated with CAIX expression utilizing a radiogenomics NSCLC dataset. The CAIX expression is an important pH regulatory enzyme that is upregulated in hypoxic tumors leading to an acidic tumor microenvironment [162] and associated with poor prognosis [151,152] including NSCLC [153,154]. The two clinical covariates that are associated with survival were number of metastatic sites and serum albumin. Higher number of metastatic sites increases disease burden and can result in mixed responses where one or more lesions may be responding while others are progressing and ultimately resulting in progressive disease. The other clinical covariate, serum albumin, has been shown to be associated with survival in NSCLC patients [156].

Overall, using standard-of-care imaging and clinical covariates, we identified and validated a novel parsimonious model that predicts OS and PFS among NSCLC patients treated with immunotherapy. The prognostic image-based (i.e., GLCM inverse difference) feature was found to be associated with CAIX, an important enzyme unregulated in hypoxic and acidotic tumors which is related to treatment resistance. The potential clinical application of this work is that baseline radiomics and clinical covariates can identify patients that are unlikely to respond to immunotherapy.

APPENDIX A. SUPPLEMENTARY TABLES AND FIGURES

Table A.1
Radial gradient and radial deviation imaging features.

No.	Features	No.	Features
1	C Radial deviation tumor mean	25	C Radial deviation outside mean
2	F Radial deviation tumor SD	26	C Radial deviation outside SD
3	C Radial gradient tumor mean	27	C Radial gradient outside mean
4	C Radial gradient tumor SD	28	C Radial gradient outside SD
5	F Radial deviation tumor mean (2D)	29	F Radial deviation outside mean (2D)
6	R Radial deviation tumor SD (2D)	30	F Radial deviation outside SD (2D)
7	F Radial gradient tumor mean (2D)	31	C Radial gradient outside mean (2D)
8	F Radial gradient tumor SD (2D)	32	F Radial gradient outside SD (2D)
9	R Radial deviation core mean	33	F Radial deviation outside-tumor separation mean
10	R Radial deviation core SD	34	F Radial deviation outside-tumor separation SD
11	R Radial gradient core mean	35	F Radial gradient outside-tumor separation mean
12	F Radial gradient core SD	36	F Radial gradient outside-tumor separation SD
13	R Radial deviation core mean (2D)	37	R Radial deviation outside-tumor separation mean (2D)
14	R Radial deviation core SD (2D)	38	R Radial deviation outside-tumor separation SD (2D)
15	R Radial gradient core mean (2D)	39	R Radial gradient outside-tumor separation mean (2D)
16	R Radial gradient core SD (2D)	40	C Radial gradient outside-tumor separation SD (2D)
17	C Radial deviation border mean	41	F Radial deviation outside-border separation mean
18	C Radial deviation border SD	42	F Radial deviation outside-border separation SD
19	C Radial gradient border mean	43	C Radial gradient outside-border separation mean
20	F Radial gradient border SD	44	C Radial gradient outside-border separation SD
21	F Radial deviation border mean (2D)	45	R Radial deviation outside-border separation mean (2D)
22	R Radial deviation border SD (2D)	46	R Radial deviation outside-border separation SD (2D)
23	C Radial gradient border mean (2D)	47	C Radial gradient outside-border separation mean (2D)
24	R Radial gradient border SD (2D)	48	F Radial gradient outside-border separation SD (2D)

Abbreviations: F = Final set of features that were analyzed after non-reproducible (R) features and correlated features (C) were removed

Table A.2

Log-rank p-values for the 17 features that are reproducible and non-redundant.

No.	Features	Log-rank P-value
2	Radial deviation tumor SD	0.071
5	Radial deviation tumor mean (2D)	0.705
7	Radial gradient tumor mean (2D)	0.697
8	Radial gradient tumor SD (2D)	0.673
12	Radial gradient core SD	0.949
20	Radial gradient border SD	0.083
21	Radial deviation border mean (2D)	0.875
29	Radial deviation outside mean (2D)	0.980
30	Radial deviation outside SD (2D)	0.111
32	Radial gradient outside SD (2D)	0.825
33	Radial deviation outside-tumor separation mean	0.322
34	Radial deviation outside-tumor separation SD	0.646
35	Radial gradient outside-tumor separation mean	0.061
36	Radial gradient outside-tumor separation SD	0.711
41	Radial deviation outside-border separation mean	0.322
42	Radial deviation outside-border separation SD	0.009
48	Radial gradient outside-border separation SD (2D)	0.029

Bold values are statistically significant.

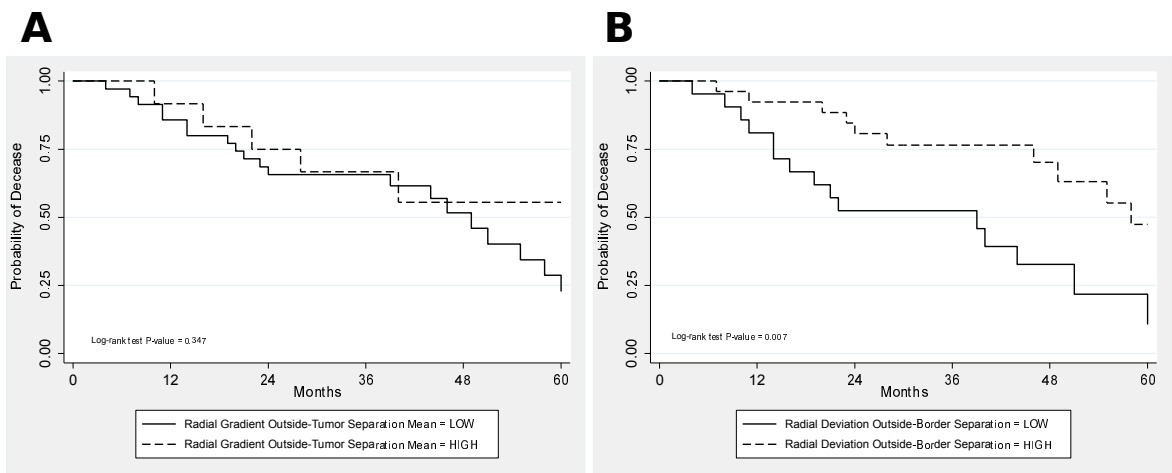


Figure A.1 Kaplan-Meier survival curves for the features in the test cohort. **a)** RG outside-tumor separation **b)** RD outside-border separation standard deviation.

Table A.3

The 5-year survival rates for dichotomized image features.

Features	5-year survival rate	Log-rank P-value
Five significant features from the training cohort		
Fig 3.2A. RG border SD		
High	42.20%	0.084
Low	60.20%	
Fig 3.2B. RG outside-tumor mean		
High	65.30%	0.061
Low	37.40%	
Fig 3.2C. RD outside-border SD		
High	67.70%	0.009
Low	34.90%	
Fig 3.2D. RG outside-border SD (2D)		
High	63.80%	0.029
Low	38.90%	
Fig 3.2E. RD tumor SD		
High	40.80%	0.071
Low	62.40%	
Combinatorial features		
Fig 3.2F. Training cohort		
High/High	100.00%	<0.001
High/Low or Low/High	50.20%	
Low/Low	12.10%	
Fig 3.2G. Test cohort		
High/High	85.70%	0.046
High/Low or Low/High	26.30%	
Low/Low	12.50%	

Abbreviations: SD = standard deviation; RD = radial deviation; RG = radial gradient

Image features were dichotomized at the median value so that “High” \geq median value and “Low” $<$ median value.

Table A.4

Univariable analysis of the association between radiomic features and rapid disease progression phenotypes.

Radiomic Feature	TTP < 2 months vs TTP ≥ 2 months (N = 183)		HPD vs Non-HPD (N = 154)	
	P value	AUROC	P value	AUROC
Avg co-occurrence homogeneity normalized	0.091	0.58	-	-
3D Laws E5L5E5	0.046	0.594	-	-
3D wavelet P1 L2 C1	0.051	0.526	-	-
Border quartile coefficient of dispersion	0.072	0.539	-	-
Border center of mass shift (mm)	0.09	0.536	-	-
B184: 3D Laws E5E5L5	0.078	0.533	-	-
Radial gradient border SD-2D	0.084	0.563	-	-
Border area density minimum volume ellipsoid	-	-	0.093	0.643
Flatness	-	-	0.077	0.663
Border intensity at volume fraction 10	-	-	0.08	0.667
Border NGTDM strength	-	-	0.003	0.674
Border 3D Laws S5R5R5	-	-	0.066	0.771
Border 3D Laws W5W5R5	-	-	0.083	0.649
Radial deviation outside-tumor mean	-	-	0.089	0.683

Abbreviations: TTP = Time-to-progression; HPD = hyperprogressive disease; AUROC = area under the receiver-operator characteristics; SD = standard deviation; NGTDM = neighboring gray tone difference matrix

Bold P-values are statistically significant.

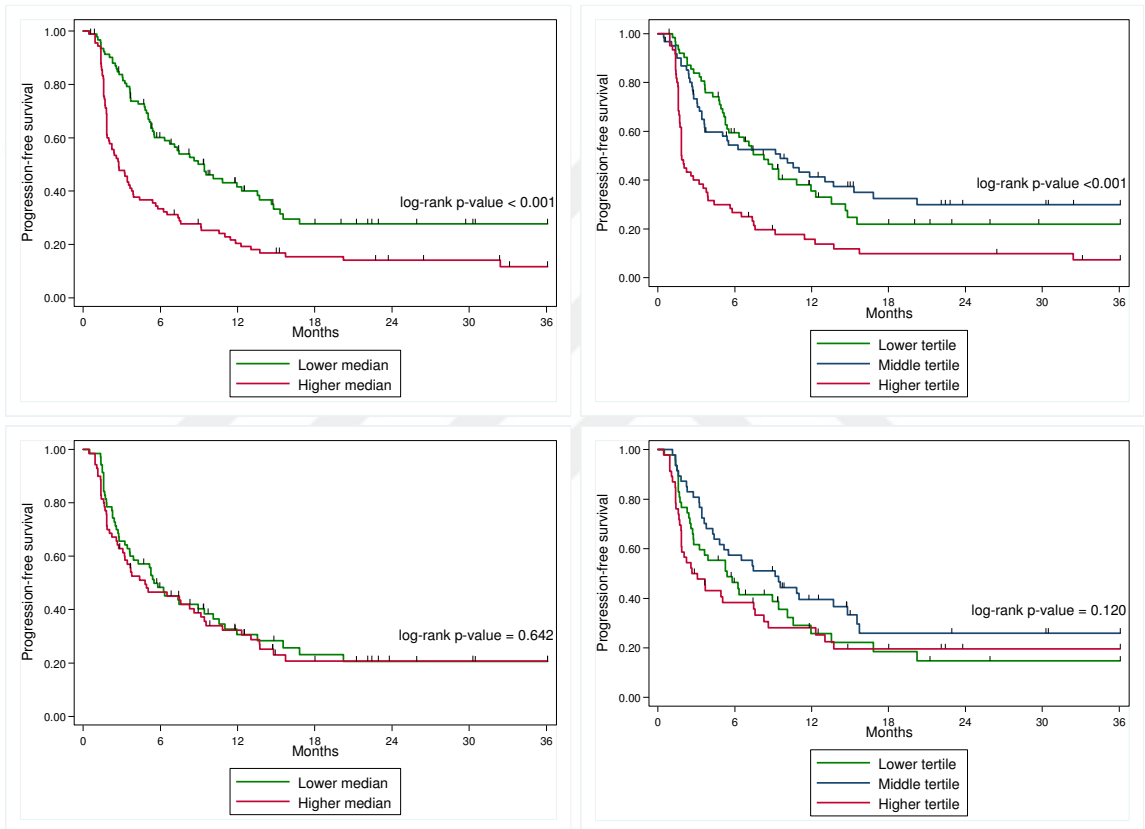


Figure A.2 Kaplan-Meier curves based on cut-points using individual probability of each patient from the combined radiomic-clinical models. The first column were based on median cut-point and the second column was based on and tertiles. The top row is for TTP < 2 months versus TTP ≥ 2 months analysis and the bottom row is for HPD vs non-HPD analysis.

Table A.5

Distribution of stability groups of statistical peritumoral radiomic features extracted from ROIs that are bounded by a lung mask.

	Distance				P- Value
	0 – 3 mm	0 – 6 mm	0 – 9 mm	0 – 12	
Algorithm 1-Initial parameter 1 vs Initial parameter 2					
CCC < 0.75	0 (0)	0 (0)	0 (0)	0 (0)	1.000
CCC ≥ 0.75 & CCC ≤ 0.95	0 (0)	0 (0)	0 (0)	1 (5.3)	
CCC > 0.95	19 (100)	19 (100)	19 (100)	18 (94.7)	
P-value		1.000	1.000	1.000	
Algorithm 1-Initial parameter 1 vs Initial parameter 3					
CCC < 0.75	0 (0)	0 (0)	0 (0)	0 (0)	1.000
CCC ≥ 0.75 & CCC ≤ 0.95	1 (5.3)	0 (0)	0 (0)	1 (5.3)	
CCC > 0.95	18 (94.7)	19 (100)	19 (100)	18 (94.7)	
P-value		1.000	1.000	1.000	
Algorithm 1-Initial parameter 2 vs Initial parameter 3					
CCC < 0.75	0 (0)	0 (0)	0 (0)	0 (0)	1.000
CCC ≥ 0.75 & CCC ≤ 0.95	1 (5.3)	0 (0)	0 (0)	0 (0)	
CCC > 0.95	18 (94.7)	19 (100)	19 (100)	19 (100)	
P-value		1.000	1.000	1.000	
Algorithm 2-Initial parameter 1 vs Initial parameter 2					
CCC < 0.75	0 (0)	0 (0)	0 (0)	0 (0)	0.250
CCC ≥ 0.75 & CCC ≤ 0.95	4 (21.1)	2 (10.5)	2 (10.5)	0 (0)	
CCC > 0.95	15 (78.5)	17 (89.5)	17 (89.5)	19 (100)	
P-value		0.660	1.000	0.486	
Algorithm 2-Initial parameter 1 vs Initial parameter 3					
CCC < 0.75	0 (0)	0 (0)	0 (0)	0 (0)	1.000
CCC ≥ 0.75 & CCC ≤ 0.95	2 (10.5)	0 (0)	0 (0)	0 (0)	
CCC > 0.95	17 (89.5)	19 (100)	19 (100)	19 (100)	
P-value		0.486	1.000	1.000	
Algorithm 2-Initial parameter 2 vs Initial parameter 3					
CCC < 0.75	0 (0)	0 (0)	0 (0)	0 (0)	0.250
CCC ≥ 0.75 & CCC ≤ 0.95	4 (21.1)	2 (10.5)	2 (10.5)	0 (0)	
CCC > 0.95	15 (78.5)	17 (89.5)	17 (89.5)	19 (100)	
P-value		0.660	1.000	0.486	
Algorithm 3-Initial parameter 1 vs Initial parameter 2					
CCC < 0.75	5 (26.3)	0 (0)	0 (0)	0 (0)	0.037
CCC ≥ 0.75 & CCC ≤ 0.95	12 (63.2)	15 (78.5)	15 (78.5)	13 (68.4)	
CCC > 0.95	2 (10.5)	4 (21.1)	4 (21.1)	6 (31.6)	
P-value		0.079	1.000	0.714	
Algorithm 3-Initial parameter 1 vs Initial parameter 3					
CCC < 0.75	0 (0)	0 (0)	0 (0)	0 (0)	<0.001
CCC ≥ 0.75 & CCC ≤ 0.95	7 (36.8)	0 (0)	0 (0)	0 (0)	
CCC > 0.95	12 (63.2)	19 (100)	19 (100)	19 (100)	
P-value		0.008	1.000	1.000	
Algorithm 3-Initial parameter 2 vs Initial parameter 3					
CCC < 0.75	0 (0)	0 (0)	0 (0)	0 (0)	<0.001
CCC ≥ 0.75 & CCC ≤ 0.95	10 (52.6)	3 (15.8)	1 (5.3)	0 (0)	
CCC > 0.95	9 (47.4)	16 (84.2)	18 (94.7)	19 (100)	
P-value		1.000	1.000	1.000	
Algorithm 1 vs Algorithm 2					
CCC < 0.75	3 (15.8)	0 (0)	0 (0)	0 (0)	0.373
CCC ≥ 0.75 & CCC ≤ 0.95	13 (68.4)	14 (73.7)	14 (73.7)	14 (73.7)	
CCC > 0.95	3 (15.8)	5 (26.3)	5 (26.3)	5 (26.3)	
P-value		0.254	1.000	1.000	
Algorithm 1 vs Algorithm 3					
CCC < 0.75	2 (10.5)	0 (0)	0 (0)	0 (0)	0.136
CCC ≥ 0.75 & CCC ≤ 0.95	14 (73.7)	14 (73.7)	11 (57.9)	10 (52.6)	
CCC > 0.95	3 (15.8)	5 (26.3)	8 (42.1)	9 (47.4)	
P-value		0.456	0.495	1.000	
Algorithm 2 vs Algorithm 3					
CCC < 0.75	5 (26.3)	0 (0)	0 (0)	0 (0)	0.049
CCC ≥ 0.75 & CCC ≤ 0.95	12 (63.2)	15 (79.0)	15 (79.0)	14 (73.7)	
CCC > 0.95	2 (10.5)	4 (21.0)	4 (21.0)	5 (26.3)	
P-value		0.079	1.000	<0.001	

Table A.6

Distribution of stability groups of histogram peritumoral radiomic features extracted from ROIs that are bounded by a lung mask.

	Distance				P- Value
	0 – 3 mm	0 – 6 mm	0 – 9 mm	0 – 12	
Algorithm 1-Initial parameter 1 vs Initial parameter 2					
CCC < 0.75	1 (3.6)	0 (0)	1 (3.6)	0 (0)	0.852
CCC ≥ 0.75 & CCC ≤ 0.95	1 (3.6)	1 (3.6)	1 (3.6)	0 (0)	
CCC > 0.95	26 (92.8)	27 (96.4)	26 (92.8)	28 (100)	
P-value		1.000	1.000	0.491	
Algorithm 1-Initial parameter 1 vs Initial parameter 3					
CCC < 0.75	1 (3.6)	1 (3.6)	0 (0)	0 (0)	0.411
CCC ≥ 0.75 & CCC ≤ 0.95	2 (7.2)	0 (0)	2 (7.2)	0 (0)	
CCC > 0.95	25 (89.2)	27 (96.4)	26 (92.8)	28 (100)	
P-value		0.741	0.491	0.491	
Algorithm 1-Initial parameter 2 vs Initial parameter 3					
CCC < 0.75	1 (3.6)	1 (3.6)	1 (3.6)	0 (0)	0.631
CCC ≥ 0.75 & CCC ≤ 0.95	2 (7.2)	0 (0)	1 (3.6)	0 (0)	
CCC > 0.95	25 (89.2)	27 (96.4)	26 (92.8)	28 (100)	
P-value		0.741	1.000	0.491	
Algorithm 2-Initial parameter 1 vs Initial parameter 2					
CCC < 0.75	1 (3.6)	0 (0)	0 (0)	0 (0)	0.281
CCC ≥ 0.75 & CCC ≤ 0.95	6 (21.4)	3 (10.8)	3 (10.8)	1 (3.6)	
CCC > 0.95	21 (75.0)	25 (89.2)	25 (89.2)	27 (96.4)	
P-value		0.295	1.000	0.611	
Algorithm 2-Initial parameter 1 vs Initial parameter 3					
CCC < 0.75	1 (3.6)	0 (0)	0 (0)	0 (0)	0.381
CCC ≥ 0.75 & CCC ≤ 0.95	4 (14.3)	1 (3.6)	1 (3.6)	1 (3.6)	
CCC > 0.95	23 (82.1)	27 (96.4)	27 (96.4)	27 (96.4)	
P-value		0.193	1.000	1.000	
Algorithm 2-Initial parameter 2 vs Initial parameter 3					
CCC < 0.75	1 (3.6)	0 (0)	0 (0)	0 (0)	0.162
CCC ≥ 0.75 & CCC ≤ 0.95	6 (21.4)	2 (7.2)	2 (7.2)	1 (3.6)	
CCC > 0.95	21 (75.0)	26 (92.8)	26 (92.8)	27 (96.4)	
P-value		0.143	1.000	1.000	
Algorithm 3-Initial parameter 1 vs Initial parameter 2					
CCC < 0.75	11 (39.3)	3 (10.8)	2 (7.2)	3 (10.8)	0.006
CCC ≥ 0.75 & CCC ≤ 0.95	17 (60.7)	21 (75.0)	22 (78.5)	18 (64.3)	
CCC > 0.95	0 (0)	4 (14.2)	4 (14.3)	7 (25.0)	
P-value		0.010	1.000	0.562	
Algorithm 3-Initial parameter 1 vs Initial parameter 3					
CCC < 0.75	0 (0)	1 (3.6)	1 (3.6)	0 (0)	<0.001
CCC ≥ 0.75 & CCC ≤ 0.95	11 (39.3)	0 (0)	0 (0)	0 (0)	
CCC > 0.95	17 (60.7)	27 (96.4)	27 (96.4)	28 (100)	
P-value		<0.001	1.000	1.000	
Algorithm 3-Initial parameter 2 vs Initial parameter 3					
CCC < 0.75	1 (3.6)	0 (0)	0 (0)	0 (0)	<0.001
CCC ≥ 0.75 & CCC ≤ 0.95	12 (42.9)	5 (17.9)	1 (3.6)	0 (0)	
CCC > 0.95	15 (53.6)	23 (82.1)	27 (96.4)	28 (100)	
P-value		0.044	0.193	1.000	
Algorithm 1 vs Algorithm 2					
CCC < 0.75	5 (17.9)	3 (10.7)	1 (3.6)	2 (7.1)	0.501
CCC ≥ 0.75 & CCC ≤ 0.95	20 (71.4)	21 (75.0)	20 (71.4)	19 (67.9)	
CCC > 0.95	3 (10.7)	4 (14.3)	7 (25.0)	7 (25.0)	
P-value		1.000	1.000	1.000	
Algorithm 1 vs Algorithm 3					
CCC < 0.75	5 (17.9)	3 (10.7)	0 (0)	1 (3.6)	0.036
CCC ≥ 0.75 & CCC ≤ 0.95	19 (67.9)	17 (60.7)	16 (57.1)	14 (50.0)	
CCC > 0.95	4 (14.3)	8 (28.6)	12 (42.9)	13 (46.4)	
P-value		0.426	0.206	0.789	
Algorithm 2 vs Algorithm 3					
CCC < 0.75	10 (35.7)	2 (7.1)	2 (7.1)	3 (10.7)	0.064
CCC ≥ 0.75 & CCC ≤ 0.95	16 (57.1)	22 (78.6)	22 (78.6)	19 (67.9)	
CCC > 0.95	2 (7.1)	4 (14.3)	4 (14.3)	6 (21.4)	
P-value		0.040	1.000	<0.001	

Table A.7

Distribution of stability groups of GLCM, GLRLM, GLSZM and NGTDM texture features peritumoral radiomic features extracted from ROIs that are bounded by a lung mask.

	Distance				P-Value
	0 – 3 mm	0 – 6 mm	0 – 9 mm	0 – 12	
Algorithm 1-Initial parameter 1 vs Initial parameter 2					
CCC < 0.75	0 (0)	0 (0)	0 (0)	0 (0)	
CCC ≥ 0.75 & CCC ≤ 0.95	1 (1.6)	0 (0)	0 (0)	0 (0)	
CCC > 0.95	61 (98.4)	62 (100)	62 (100)	62 (100)	1.000
P-value		1.000	1.000	1.000	
Algorithm 1-Initial parameter 1 vs Initial parameter 3					
CCC < 0.75	0 (0)	0 (0)	0 (0)	0 (0)	
CCC ≥ 0.75 & CCC ≤ 0.95	1 (1.6)	0 (0)	0 (0)	1 (1.6)	
CCC > 0.95	61 (98.4)	62 (100)	62 (100)	61 (98.4)	1.000
P-value		1.000	1.000	1.000	
Algorithm 1-Initial parameter 2 vs Initial parameter 3					
CCC < 0.75	0 (0)	0 (0)	0 (0)	0 (0)	
CCC ≥ 0.75 & CCC ≤ 0.95	2 (3.2)	0 (0)	0 (0)	0 (0)	
CCC > 0.95	60 (96.8)	62 (100)	62 (100)	62 (100)	0.247
P-value		0.496	1.000	1.000	
Algorithm 2-Initial parameter 1 vs Initial parameter 2					
CCC < 0.75	0 (0)	0 (0)	0 (0)	0 (0)	
CCC ≥ 0.75 & CCC ≤ 0.95	15 (14.2)	1 (1.6)	2 (3.2)	1 (1.6)	
CCC > 0.95	47 (75.8)	61 (98.4)	60 (96.8)	61 (98.4)	<0.001
P-value		<0.001	1.000	1.000	
Algorithm 2-Initial parameter 1 vs Initial parameter 3					
CCC < 0.75	0 (0)	0 (0)	0 (0)	0 (0)	
CCC ≥ 0.75 & CCC ≤ 0.95	6 (9.7)	0 (0)	0 (0)	0 (0)	
CCC > 0.95	56 (90.3)	62 (100)	62 (100)	62 (100)	0.001
P-value		0.028	1.000	1.000	
Algorithm 2-Initial parameter 2 vs Initial parameter 3					
CCC < 0.75	1 (1.6)	0 (0)	0 (0)	0 (0)	
CCC ≥ 0.75 & CCC ≤ 0.95	13 (21.0)	2 (3.2)	2 (3.2)	2 (3.2)	
CCC > 0.95	48 (77.4)	60 (96.8)	60 (96.8)	60 (96.8)	<0.001
P-value		0.002	1.000	1.000	
Algorithm 3-Initial parameter 1 vs Initial parameter 2					
CCC < 0.75	24 (38.7)	4 (6.5)	3 (4.8)	3 (4.8)	
CCC ≥ 0.75 & CCC ≤ 0.95	36 (58.1)	47 (75.8)	32 (51.6)	27 (43.6)	
CCC > 0.95	2 (3.2)	11 (17.7)	27 (43.6)	32 (51.6)	<0.001
P-value		<0.001	0.007	0.630	
Algorithm 3-Initial parameter 1 vs Initial parameter 3					
CCC < 0.75	0 (0)	0 (0)	0 (0)	0 (0)	
CCC ≥ 0.75 & CCC ≤ 0.95	14 (22.6)	3 (4.8)	1 (1.6)	0 (0)	
CCC > 0.95	48 (77.4)	59 (95.2)	61 (98.4)	62 (100)	<0.001
P-value		0.007	0.619	1.000	
Algorithm 3-Initial parameter 2 vs Initial parameter 3					
CCC < 0.75	0 (0)	0 (0)	0 (0)	0 (0)	
CCC ≥ 0.75 & CCC ≤ 0.95	28 (45.2)	4 (6.5)	2 (3.2)	0 (0)	
CCC > 0.95	34 (54.8)	58 (93.5)	60 (96.8)	62 (100)	<0.001
P-value		<0.001	0.680	0.496	
Algorithm 1 vs Algorithm 2					
CCC < 0.75	21 (33.9)	4 (6.4)	2 (3.2)	1 (1.6)	
CCC ≥ 0.75 & CCC ≤ 0.95	39 (62.9)	45 (72.6)	28 (45.2)	24 (38.7)	
CCC > 0.95	2 (3.2)	13 (21.0)	32 (51.6)	37 (59.7)	<0.001
P-value		<0.001	0.001	0.596	
Algorithm 1 vs Algorithm 3					
CCC < 0.75	18 (29.0)	2 (3.2)	1 (1.6)	0 (0)	
CCC ≥ 0.75 & CCC ≤ 0.95	36 (58.1)	38 (61.3)	24 (38.7)	22 (35.5)	
CCC > 0.95	8 (12.9)	22 (35.5)	37 (59.7)	40 (64.5)	<0.001
P-value		<0.001	0.015	0.711	
Algorithm 2 vs Algorithm 3					
CCC < 0.75	23 (37.1)	5 (8.0)	3 (4.8)	2 (3.2)	
CCC ≥ 0.75 & CCC ≤ 0.95	36 (58.1)	44 (71.0)	36 (58.1)	30 (48.4)	
CCC > 0.95	3 (4.8)	13 (21.0)	23 (37.1)	30 (48.4)	<0.001
P-value		<0.001	0.148	0.536	

Table A.8

Distribution of stability groups of 3D Laws texture peritumoral radiomic features extracted from ROIs that are bounded by a lung mask.

	Distance				P-Value
	0 – 3 mm	0 – 6 mm	0 – 9 mm	0 – 12	
Algorithm 1-Initial parameter 1 vs Initial parameter 2					
CCC < 0.75	29 (23.2)	0 (0)	8 (6.4)	12 (9.6)	
CCC ≥ 0.75 & CCC ≤ 0.95	54 (43.2)	7 (5.6)	0 (0)	1 (0.8)	
CCC > 0.95	42 (33.6)	118 (94.4)	117 (93.6)	112 (89.6)	<0.001
P-value		<0.001	<0.001	0.362	
Algorithm 1-Initial parameter 1 vs Initial parameter 3					
CCC < 0.75	13 (10.4)	0 (0)	8 (6.4)	20 (16.0)	
CCC ≥ 0.75 & CCC ≤ 0.95	41 (32.8)	20 (16.0)	36 (28.8)	0 (0)	
CCC > 0.95	71 (56.8)	105 (84.0)	81 (64.8)	105 (84.0)	<0.001
P-value		<0.001	<0.001	<0.001	
Algorithm 1-Initial parameter 2 vs Initial parameter 3					
CCC < 0.75	32 (25.6)	1 (0.8)	7 (5.6)	20 (16.0)	
CCC ≥ 0.75 & CCC ≤ 0.95	29 (23.2)	19 (15.2)	37 (29.6)	0 (0)	
CCC > 0.95	64 (51.2)	105 (84.0)	81 (64.8)	105 (84.0)	<0.001
P-value		<0.001	0.001	<0.001	
Algorithm 2-Initial parameter 1 vs Initial parameter 2					
CCC < 0.75	49 (39.2)	36 (28.8)	44 (35.2)	44 (35.2)	
CCC ≥ 0.75 & CCC ≤ 0.95	75 (60.0)	50 (40.0)	17 (13.6)	7 (5.6)	
CCC > 0.95	1 (0.8)	39 (31.2)	64 (51.2)	74 (59.2)	<0.001
P-value		<0.001	0.001	0.088	
Algorithm 2-Initial parameter 1 vs Initial parameter 3					
CCC < 0.75	32 (25.6)	22 (17.6)	37 (29.6)	44 (35.2)	
CCC ≥ 0.75 & CCC ≤ 0.95	90 (72.0)	46 (36.8)	9 (7.2)	1 (0.8)	
CCC > 0.95	3 (2.4)	57 (45.6)	79 (63.2)	80 (64.0)	<0.001
P-value		<0.001	0.001	0.027	
Algorithm 2-Initial parameter 2 vs Initial parameter 3					
CCC < 0.75	53 (42.4)	6 (4.8)	8 (6.4)	45 (36.0)	
CCC ≥ 0.75 & CCC ≤ 0.95	68 (54.4)	65 (52.0)	17 (13.6)	5 (4.0)	
CCC > 0.95	4 (3.2)	54 (43.2)	100 (80.0)	75 (60.0)	<0.001
P-value		<0.001	<0.001	<0.001	
Algorithm 3-Initial parameter 1 vs Initial parameter 2					
CCC < 0.75	123 (98.4)	124 (99.2)	101 (80.8)	61 (48.8)	
CCC ≥ 0.75 & CCC ≤ 0.95	2 (1.6)	1 (0.8)	24 (19.2)	64 (51.2)	
CCC > 0.95	0 (0)	0 (0)	0 (0)	0 (0)	<0.001
P-value		1.000	<0.001	<0.001	
Algorithm 3-Initial parameter 1 vs Initial parameter 3					
CCC < 0.75	39 (31.2)	29 (23.2)	17 (13.6)	11 (8.8)	
CCC ≥ 0.75 & CCC ≤ 0.95	78 (62.4)	73 (58.4)	31 (24.8)	21 (16.8)	
CCC > 0.95	8 (6.4)	23 (18.4)	77 (61.6)	93 (74.4)	<0.001
P-value		0.011	<0.001	0.102	
Algorithm 3-Initial parameter 2 vs Initial parameter 3					
CCC < 0.75	47 (37.6)	32 (25.6)	30 (24.0)	12 (9.6)	
CCC ≥ 0.75 & CCC ≤ 0.95	78 (62.4)	86 (68.8)	62 (49.6)	31 (24.8)	
CCC > 0.95	0 (0)	7 (5.6)	33 (26.4)	82 (65.6)	<0.001
P-value		0.003	<0.001	<0.001	
Algorithm 1 vs Algorithm 2					
CCC < 0.75	96 (76.8)	69 (55.2)	50 (40.0)	45 (36.0)	
CCC ≥ 0.75 & CCC ≤ 0.95	29 (23.2)	56 (44.8)	75 (60.0)	80 (64.0)	
CCC > 0.95	0 (0)	0 (0)	0 (0)	0 (0)	<0.001
P-value		<0.001	<0.022	0.602	
Algorithm 1 vs Algorithm 3					
CCC < 0.75	125 (100)	119 (95.2)	102 (81.6)	73 (58.4)	
CCC ≥ 0.75 & CCC ≤ 0.95	0 (0)	6 (4.8)	23 (18.4)	53 (41.6)	
CCC > 0.95	0 (0)	0 (0)	0 (0)	0 (0)	<0.001
P-value		0.029	0.001	<0.001	
Algorithm 2 vs Algorithm 3					
CCC < 0.75	125 (100)	125 (100)	106 (84.6)	64 (51.2)	
CCC ≥ 0.75 & CCC ≤ 0.95	0 (0)	0 (0)	19 (15.2)	61 (48.8)	
CCC > 0.95	0 (0)	0 (0)	0 (0)	0 (0)	<0.001
P-value		1.000	<0.001	<0.001	

Table A.9

Distribution of stability groups of wavelet texture peritumoral radiomic features extracted from ROIs that are bounded by a lung mask.

	Distance				P-Value
	0 – 3 mm	0 – 6 mm	0 – 9 mm	0 – 12	
Algorithm 1-Initial parameter 1 vs Initial parameter 2					
CCC < 0.75	0 (0)	0 (0)	1 (3.3)	9 (30.0)	
CCC ≥ 0.75 & CCC ≤ 0.95	7 (23.3)	1 (3.3)	25 (83.3)	6 (20.0)	
CCC > 0.95	23 (76.7)	29 (96.7)	4 (13.3)	15 (50.0)	<0.001
	P-value	0.052	<0.001	<0.001	
Algorithm 1-Initial parameter 1 vs Initial parameter 3					
CCC < 0.75	0 (0)	0 (0)	1 (3.33)	9 (30.0)	
CCC ≥ 0.75 & CCC ≤ 0.95	10 (33.3)	12 (40.0)	25 (83.3)	6 (20.0)	
CCC > 0.95	20 (66.7)	18 (60.0)	4 (13.3)	15 (50.0)	<0.001
	P-value	0.789	<0.001	<0.001	
Algorithm 1-Initial parameter 2 vs Initial parameter 3					
CCC < 0.75	0 (0)	0 (0)	0 (0)	0 (0)	
CCC ≥ 0.75 & CCC ≤ 0.95	0 (0)	5 (16.7)	0 (0)	0 (0)	
CCC > 0.95	30 (100)	-83.300	30 (100)	30 (100)	0.003
	P-value	0.052	0.052	1.000	
Algorithm 2-Initial parameter 1 vs Initial parameter 2					
CCC < 0.75	7 (23.3)	0 (0)	0 (0)	0 (0)	
CCC ≥ 0.75 & CCC ≤ 0.95	22 (73.3)	30 ()	28 (93.3)	2 (6.7)	
CCC > 0.95	1 (3.33)	0 (0)	2 (6.7)	28 (93.3)	<0.001
	P-value	0.005	0.492	<0.001	
Algorithm 2-Initial parameter 1 vs Initial parameter 3					
CCC < 0.75	7 (23.3)	0 (0)	0 (0)	1 (3.3)	
CCC ≥ 0.75 & CCC ≤ 0.95	22 (73.3)	29 (96.7)	29 (96.7)	8 (26.7)	
CCC > 0.95	1 (3.33)	1 (3.3)	1 (3.33)	21 (70.0)	<0.001
	P-value	0.011	1.000	<0.001	
Algorithm 2-Initial parameter 2 vs Initial parameter 3					
CCC < 0.75	0 (0)	0 (0)	0 (0)	0 (0)	
CCC ≥ 0.75 & CCC ≤ 0.95	4 (13.3)	3 (10.0)	0 (0)	2 (6.7)	
CCC > 0.95	26 (86.7)	27 (90.0)	30 (100)	28 (93.3)	0.242
	P-value	1.000	0.237	0.492	
Algorithm 3-Initial parameter 1 vs Initial parameter 2					
CCC < 0.75	0 (0)	0 (0)	0 (0)	0 (0)	
CCC ≥ 0.75 & CCC ≤ 0.95	19 (63.3)	15 (50)	15 (50)	30 (100)	
CCC > 0.95	11 (36.7)	15 (50)	15 (50)	0 (0)	<0.001
	P-value	<0.001	<0.001	<0.001	
Algorithm 3-Initial parameter 1 vs Initial parameter 3					
CCC < 0.75	2 (6.7)	3 (10.0)	1 (3.3)	0 (0)	
CCC ≥ 0.75 & CCC ≤ 0.95	23 (76.7)	27 (90.0)	10 (33.3)	24 (80.0)	
CCC > 0.95	5 (16.7)	0 (0)	19 (63.3)	6 (20.0)	<0.001
	P-value	0.101	<0.001	0.001	
Algorithm 3-Initial parameter 2 vs Initial parameter 3					
CCC < 0.75	1 (3.3)	16 (53.3)	0 (0)	0 (0)	
CCC ≥ 0.75 & CCC ≤ 0.95	28 (93.3)	14 (46.7)	0 (0)	18 (60.0)	
CCC > 0.95	1 (3.3)	0 (0)	30 (100)	12 (40.0)	<0.001
	P-value	<0.001	<0.001	<0.001	
Algorithm 1 vs Algorithm 2					
CCC < 0.75	23 (76.7)	26 (86.7)	16 (53.3)	25 (83.3)	
CCC ≥ 0.75 & CCC ≤ 0.95	7 (23.3)	4 (13.3)	14 (46.7)	5 (16.7)	
CCC > 0.95	0 (0)	0 (0)	0 (0)	0 (0)	0.018
	P-value	0.506	0.010	0.025	
Algorithm 1 vs Algorithm 3					
CCC < 0.75	15 (50.0)	15 (50.0)	15 (50.0)	7 ()	
CCC ≥ 0.75 & CCC ≤ 0.95	15 (50.0)	15 (50.0)	2 (6.7)	23 (76.7)	
CCC > 0.95	0 (0)	0 (0)	13 (43.3)	0 (0)	<0.001
	P-value	1.000	<0.001	<0.001	
Algorithm 2 vs Algorithm 3					
CCC < 0.75	17 (56.7)	15 (50.0)	17 (56.7)	15 (50.0)	
CCC ≥ 0.75 & CCC ≤ 0.95	13 (43.3)	15 (50.0)	13 (43.3)	15 (50.0)	
CCC > 0.95	0 (0)	0 (0)	0 (0)	0 (0)	0.930
	P-value	0.796	0.796	0.736	

Table A.10

Distribution of stability groups of all peritumoral radiomic features extracted from ROIs that are not bounded by a lung mask.

	Distance				P- Value
	0 – 3 mm	0 – 6 mm	0 – 9 mm	0 – 12	
Algorithm 1-Initial parameter 1 vs Initial parameter 2					
CCC < 0.75	30 (11.4)	0 (0)	10 (3.8)	23 (8.7)	
CCC ≥ 0.75 & CCC ≤ 0.95	68 (25.8)	18 (6.8)	23 (8.7)	17 (6.4)	
CCC > 0.95	166 (62.9)	246 (93.2)	231 (87.5)	224 (84.9)	<0.001
P-value		<0.001	0.002	0.045	
Algorithm 1-Initial parameter 1 vs Initial parameter 3					
CCC < 0.75	37 (14.0)	18 (6.8)	29 (11.0)	26 (9.9)	
CCC ≥ 0.75 & CCC ≤ 0.95	60 (22.7)	17 (6.4)	39 (14.8)	14 (5.3)	
CCC > 0.95	167 (63.3)	229 (86.8)	196 (74.2)	224 (84.8)	<0.001
P-value		<0.001	0.001	0.001	
Algorithm 1-Initial parameter 2 vs Initial parameter 3					
CCC < 0.75	32 (12.1)	19 (7.2)	33 (12.5)	20 (7.5)	
CCC ≥ 0.75 & CCC ≤ 0.95	38 (14.4)	18 (6.8)	10 (3.8)	6 (2.3)	
CCC > 0.95	194 (73.5)	227 (86.0)	221 (83.8)	238 (90.2)	<0.001
P-value		0.001	0.045	0.092	
Algorithm 2-Initial parameter 1 vs Initial parameter 2					
CCC < 0.75	48 (18.2)	27 (10.2)	35 (13.3)	42 (15.9)	
CCC ≥ 0.75 & CCC ≤ 0.95	123 (46.6)	71 (26.9)	53 (20.1)	14 (5.3)	
CCC > 0.95	93 (35.2)	166 (62.9)	176 (66.6)	208 (78.8)	<0.001
P-value		<0.001	0.142	<0.001	
Algorithm 2-Initial parameter 1 vs Initial parameter 3					
CCC < 0.75	40 (15.1)	22 (8.3)	40 (15.1)	46 (17.4)	
CCC ≥ 0.75 & CCC ≤ 0.95	118 (44.7)	61 (23.1)	35 (13.3)	9 (3.4)	
CCC > 0.95	106 (40.2)	181 (68.6)	189 (71.6)	209 (79.2)	<0.001
P-value		<0.001	0.002	<0.001	
Algorithm 2-Initial parameter 2 vs Initial parameter 3					
CCC < 0.75	29 (11.0)	1 (0.4)	33 (12.5)	11 (4.2)	
CCC ≥ 0.75 & CCC ≤ 0.95	110 (41.7)	39 (14.8)	12 (4.5)	42 (15.9)	
CCC > 0.95	125 (47.3)	224 (84.8)	219 (83.0)	211 (79.9)	<0.001
P-value		<0.001	<0.001	<0.001	
Algorithm 3-Initial parameter 1 vs Initial parameter 2					
CCC < 0.75	182 (68.9)	120 (45.4)	65 (24.6)	64 (24.2)	
CCC ≥ 0.75 & CCC ≤ 0.95	79 (29.9)	126 (47.7)	160 (60.6)	148 (56.1)	
CCC > 0.95	3 (1.2)	18 (6.8)	39 (14.8)	52 (19.7)	<0.001
P-value		<0.001	0.023	0.017	
Algorithm 3-Initial parameter 1 vs Initial parameter 3					
CCC < 0.75	43 (16.3)	32 (12.1)	33 (12.5)	2 (0.8)	
CCC ≥ 0.75 & CCC ≤ 0.95	128 (48.5)	46 (17.4)	27 (10.2)	49 (18.5)	
CCC > 0.95	93 (35.2)	186 (70.5)	204 (77.3)	213 (80.7)	<0.001
P-value		0.008	0.054	<0.001	
Algorithm 3-Initial parameter 2 vs Initial parameter 3					
CCC < 0.75	58 (22.0)	49 (18.6)	33 (12.5)	32 (12.1)	
CCC ≥ 0.75 & CCC ≤ 0.95	147 (55.7)	89 (33.7)	18 (6.8)	31 (11.7)	
CCC > 0.95	59 (22.3)	126 (47.7)	213 (80.7)	201 (76.2)	<0.001
P-value		<0.001	<0.001	0.156	
Algorithm 1 vs Algorithm 2					
CCC < 0.75	138 (52.3)	76 (28.8)	65 (24.6)	67 (25.4)	
CCC ≥ 0.75 & CCC ≤ 0.95	120 (45.4)	148 (56.1)	84 (31.8)	45 (17.0)	
CCC > 0.95	6 (2.3)	40 (15.1)	115 (43.6)	152 (57.6)	<0.001
P-value		<0.001	<0.001	<0.001	
Algorithm 1 vs Algorithm 3					
CCC < 0.75	169 (64.0)	119 (45.1)	46 (17.4)	50 (18.9)	
CCC ≥ 0.75 & CCC ≤ 0.95	80 (30.3)	102 (38.6)	150 (56.8)	153 (58.0)	
CCC > 0.95	15 (5.7)	43 (16.3)	68 (25.8)	61 (23.1)	<0.001
P-value		<0.001	<0.001	0.740	
Algorithm 2 vs Algorithm 3					
CCC < 0.75	174 (65.9)	122 (46.2)	70 (26.5)	60 (22.7)	
CCC ≥ 0.75 & CCC ≤ 0.95	86 (32.6)	123 (46.6)	158 (59.9)	154 (58.3)	
CCC > 0.95	4 (1.5)	19 (7.2)	36 (13.6)	50 (18.9)	<0.001
P-value		<0.001	<0.001	0.218	

Table A.11
Distribution of stability groups of statistical peritumoral radiomic features extracted from ROIs that are not bounded by a lung mask.

	Distance				P-value
	0 – 3 mm	0 – 6 mm	0 – 9 mm	0 – 12	
Algorithm 1-Initial parameter 1 vs Initial parameter 2					
CCC < 0.75	0 (0)	0 (0)	0 (0)	0 (0)	
CCC ≥ 0.75 & CCC ≤ 0.95	0 (0)	0 (0)	0 (0)	1 (5.3)	
CCC > 0.95	19 (100)	19 (100)	19 (100)	18 (94.7)	1.000
P-value		1.000	1.000	1.000	
Algorithm 1-Initial parameter 1 vs Initial parameter 3					
CCC < 0.75	0 (0)	0 (0)	0 (0)	0 (0)	
CCC ≥ 0.75 & CCC ≤ 0.95	1 (5.3)	0 (0)	0 (0)	1 (5.3)	
CCC > 0.95	18 (94.7)	19 (100)	19 (100)	18 (94.7)	1.000
P-value		1.000	1.000	1.000	
Algorithm 1-Initial parameter 2 vs Initial parameter 3					
CCC < 0.75	0 (0)	0 (0)	0 (0)	0 (0)	
CCC ≥ 0.75 & CCC ≤ 0.95	1 (5.3)	0 (0)	0 (0)	0 (0)	
CCC > 0.95	18 (94.7)	19 (100)	19 (100)	19 (100)	1.000
P-value		1.000	1.000	1.000	
Algorithm 2-Initial parameter 1 vs Initial parameter 2					
CCC < 0.75	0 (0)	0 (0)	0 (0)	0 (0)	
CCC ≥ 0.75 & CCC ≤ 0.95	2 (10.5)	2 (10.5)	2 (10.5)	1 (5.3)	
CCC > 0.95	17 (89.5)	17 (89.5)	17 (89.5)	18 (94.7)	1.000
P-value		1.000	1.000	1.000	
Algorithm 2-Initial parameter 1 vs Initial parameter 3					
CCC < 0.75	0 (0)	0 (0)	0 (0)	1 (5.3)	
CCC ≥ 0.75 & CCC ≤ 0.95	1 (5.3)	0 (0)	0 (0)	1 (5.3)	
CCC > 0.95	18 (94.7)	19 (100)	19 (100)	17 (89.4)	0.610
P-value		1.000	1.000	0.486	
Algorithm 2-Initial parameter 2 vs Initial parameter 3					
CCC < 0.75	0 (0)	0 (0)	0 (0)	1 (5.3)	
CCC ≥ 0.75 & CCC ≤ 0.95	2 (10.5)	2 (10.5)	2 (10.5)	1 (5.3)	
CCC > 0.95	17 (89.5)	17 (89.5)	17 (89.5)	17 (89.4)	1.000
P-value		1.000	1.000	1.000	
Algorithm 3-Initial parameter 1 vs Initial parameter 2					
CCC < 0.75	6 (31.6)	1 (5.3)	1 (5.3)	1 (5.3)	
CCC ≥ 0.75 & CCC ≤ 0.95	12 (63.2)	16 (84.2)	15 (79.0)	14 (73.7)	
CCC > 0.95	1 (5.3)	2 (10.5)	3 (15.8)	4 (21.0)	0.170
P-value		0.115	1.000	0.714	
Algorithm 3-Initial parameter 1 vs Initial parameter 3					
CCC < 0.75	0 (0)	0 (0)	0 (0)	0 (0)	
CCC ≥ 0.75 & CCC ≤ 0.95	8 (42.1)	0 (0)	0 (0)	1 (5.3)	
CCC > 0.95	11 (57.9)	19 (100)	19 (100)	18 (94.7)	<0.001
P-value		0.003	1.000	<0.001	
Algorithm 3-Initial parameter 2 vs Initial parameter 3					
CCC < 0.75	0 (0)	0 (0)	0 (0)	0 (0)	
CCC ≥ 0.75 & CCC ≤ 0.95	9 (47.4)	2 (10.5)	2 (10.5)	0 (0)	
CCC > 0.95	10 (52.6)	17 (89.5)	17 (89.5)	19 (100)	0.001
P-value		0.029	1.000	0.486	
Algorithm 1 vs Algorithm 2					
CCC < 0.75	4 (21.0)	0 (0)	0 (0)	1 (5.3)	
CCC ≥ 0.75 & CCC ≤ 0.95	13 (68.4)	10 (52.6)	6 (31.6)	3 (15.8)	
CCC > 0.95	2 (10.5)	9 (47.4)	13 (68.4)	15 (79.0)	<0.001
P-value		0.012	0.325	0.447	
Algorithm 1 vs Algorithm 3					
CCC < 0.75	2 (10.5)	0 (0)	0 (0)	0 (0)	
CCC ≥ 0.75 & CCC ≤ 0.95	13 (68.4)	13 (68.4)	11 (57.9)	12 (63.2)	
CCC > 0.95	4 (21.0)	6 (31.6)	8 (42.1)	7 (36.8)	0.461
P-value		0.484	0.737	0.500	
Algorithm 2 vs Algorithm 3					
CCC < 0.75	5 (26.3)	0 (0)	0 (0)	1 (5.3)	
CCC ≥ 0.75 & CCC ≤ 0.95	13 (68.4)	17 (89.5)	16 (84.2)	15 (79.0)	
CCC > 0.95	1 (5.3)	2 (10.5)	3 (15.8)	3 (15.8)	0.082
P-value		0.046	1.000	1.000	

Table A.12

Distribution of stability groups of histogram peritumoral radiomic features extracted from ROIs that are not bounded by a lung mask.

	Distance				P- Value
	0 – 3 mm	0 – 6 mm	0 – 9 mm	0 – 12	
Algorithm 1-Initial parameter 1 vs Initial parameter 2					
CCC < 0.75	0 (0)	0 (0)	1 (3.6)	0 (0)	0.753
CCC ≥ 0.75 & CCC ≤ 0.95	2 (7.1)	1 (3.6)	1 (3.6)	0 (0)	
CCC > 0.95	26 (92.9)	27 (96.4)	26 (92.9)	28 (100)	
P-value		1.000	1.000	0.491	
Algorithm 1-Initial parameter 1 vs Initial parameter 3					
CCC < 0.75	1 (3.6)	1 (3.6)	1 (3.6)	1 (3.6)	1.000
CCC ≥ 0.75 & CCC ≤ 0.95	1 (3.6)	1 (3.6)	1 (3.6)	1 (3.6)	
CCC > 0.95	26 (92.9)	26 (92.9)	26 (92.9)	26 (92.9)	
P-value		1.000	1.000	1.000	
Algorithm 1-Initial parameter 2 vs Initial parameter 3					
CCC < 0.75	0 (0)	1 (3.6)	1 (3.6)	0 (0)	0.874
CCC ≥ 0.75 & CCC ≤ 0.95	3 (10.7)	1 (3.6)	1 (3.6)	2 (7.1)	
CCC > 0.95	25 (89.3)	26 (92.8)	26 (92.8)	26 (92.8)	
P-value		0.611	1.000	1.000	
Algorithm 2-Initial parameter 1 vs Initial parameter 2					
CCC < 0.75	1 (3.6)	0 (0)	1 (3.6)	0 (0)	0.569
CCC ≥ 0.75 & CCC ≤ 0.95	4 (14.3)	3 (10.7)	2 (7.1)	1 (3.6)	
CCC > 0.95	23 (82.1)	25 (89.3)	25 (89.3)	27 (96.4)	
P-value		0.705	1.000	0.611	
Algorithm 2-Initial parameter 1 vs Initial parameter 3					
CCC < 0.75	1 (3.6)	0 (0)	0 (0)	0 (0)	0.459
CCC ≥ 0.75 & CCC ≤ 0.95	4 (14.3)	2 (7.1)	1 (3.6)	3 (10.7)	
CCC > 0.95	23 (82.1)	26 (92.9)	27 (96.4)	25 (89.3)	
P-value		0.422	1.000	0.305	
Algorithm 2-Initial parameter 2 vs Initial parameter 3					
CCC < 0.75	0 (0)	0 (0)	1 (3.6)	0 (0)	0.793
CCC ≥ 0.75 & CCC ≤ 0.95	6 (21.4)	4 (14.3)	3 (10.7)	4 (14.3)	
CCC > 0.95	22 (78.6)	24 (85.7)	24 (85.7)	24 (85.7)	
P-value		0.729	1.000	1.000	
Algorithm 3-Initial parameter 1 vs Initial parameter 2					
CCC < 0.75	12 (42.9)	3 (10.7)	2 (7.1)	2 (7.1)	0.002
CCC ≥ 0.75 & CCC ≤ 0.95	16 (57.1)	21 (75.0)	21 (75.0)	20 (71.5)	
CCC > 0.95	0 (0)	4 (14.3)	5 (17.9)	6 (21.4)	
P-value		0.005	1.000	1.000	
Algorithm 3-Initial parameter 1 vs Initial parameter 3					
CCC < 0.75	1 (3.6)	0 (0)	1 (3.6)	0 (0)	<0.001
CCC ≥ 0.75 & CCC ≤ 0.95	11 (39.3)	1 (3.6)	1 (3.6)	0 (0)	
CCC > 0.95	16 (57.1)	27 (96.4)	26 (92.8)	28 (100)	
P-value		0.001	1.000	0.491	
Algorithm 3-Initial parameter 2 vs Initial parameter 3					
CCC < 0.75	1 (3.6)	1 (3.6)	1 (3.6)	0 (0)	<0.001
CCC ≥ 0.75 & CCC ≤ 0.95	11 (39.3)	2 (7.1)	2 (7.1)	0 (0)	
CCC > 0.95	16 (57.1)	25 (89.3)	25 (89.3)	28 (100)	
P-value		0.010	1.000	0.236	
Algorithm 1 vs Algorithm 2					
CCC < 0.75	8 (28.6)	2 (7.1)	2 (7.1)	2 (7.1)	<0.001
CCC ≥ 0.75 & CCC ≤ 0.95	19 (67.8)	16 (57.1)	10 (35.7)	11 (39.3)	
CCC > 0.95	1 (3.6)	10 (35.7)	16 (57.1)	15 (53.6)	
P-value		0.003	0.309	1.000	
Algorithm 1 vs Algorithm 3					
CCC < 0.75	7 (25.0)	3 (10.7)	2 (7.1)	1 (3.6)	0.053
CCC ≥ 0.75 & CCC ≤ 0.95	18 (64.3)	16 (57.1)	15 (53.6)	15 (53.6)	
CCC > 0.95	3 (10.7)	9 (32.1)	11 (39.3)	12 (42.8)	
P-value		0.122	0.855	1.000	
Algorithm 2 vs Algorithm 3					
CCC < 0.75	9 (32.1)	3 (10.7)	2 (7.1)	2 (7.1)	0.022
CCC ≥ 0.75 & CCC ≤ 0.95	19 (67.9)	22 (78.6)	21 (75.0)	20 (71.4)	
CCC > 0.95	0 (0)	3 (10.7)	5 (17.9)	6 (21.5)	
P-value		0.040	0.807	1.000	

Table A.13

Distribution of stability groups of texture peritumoral radiomic features extracted from ROIs that are not bounded by a lung mask.

	Distance				P- Value
	0 – 3 mm	0 – 6 mm	0 – 9 mm	0 – 12	
Algorithm 1-Initial parameter 1 vs Initial parameter 2					
CCC < 0.75	0 (0)	0 (0)	0 (0)	0 (0)	1.000
CCC ≥ 0.75 & CCC ≤ 0.95	1 (1.6)	0 (0)	0 (0)	0 (0)	
CCC > 0.95	61 (98.4)	62 (100)	62 (100)	62 (100)	
P-value		1.000	1.000	1.000	
Algorithm 1-Initial parameter 1 vs Initial parameter 3					
CCC < 0.75	0 (0)	0 (0)	0 (0)	0 (0)	0.197
CCC ≥ 0.75 & CCC ≤ 0.95	3 (4.8)	0 (0)	0 (0)	1 (1.6)	
CCC > 0.95	59 (95.2)	62 (100)	62 (100)	61 (98.4)	
P-value		0.244	1.000	1.000	
Algorithm 1-Initial parameter 2 vs Initial parameter 3					
CCC < 0.75	0 (0)	0 (0)	0 (0)	0 (0)	0.060
CCC ≥ 0.75 & CCC ≤ 0.95	3 (4.8)	0 (0)	0 (0)	0 (0)	
CCC > 0.95	59 (95.2)	62 (100)	62 (100)	62 (100)	
P-value		0.244	1.000	1.000	
Algorithm 2-Initial parameter 1 vs Initial parameter 2					
CCC < 0.75	0 (0)	0 (0)	0 (0)	0 (0)	<0.001
CCC ≥ 0.75 & CCC ≤ 0.95	13 (21.0)	1 (1.6)	2 (3.2)	1 (1.6)	
CCC > 0.95	49 (79.0)	61 (98.4)	60 (96.8)	61 (98.4)	
P-value		0.001	1.000	1.000	
Algorithm 2-Initial parameter 1 vs Initial parameter 3					
CCC < 0.75	0 (0)	0 (0)	0 (0)	0 (0)	0.015
CCC ≥ 0.75 & CCC ≤ 0.95	4 (6.4)	0 (0)	0 (0)	0 (0)	
CCC > 0.95	58 (93.6)	62 (100)	62 (100)	62 (100)	
P-value		0.119	1.000	1.000	
Algorithm 2-Initial parameter 2 vs Initial parameter 3					
CCC < 0.75	0 (0)	0 (0)	0 (0)	0 (0)	<0.001
CCC ≥ 0.75 & CCC ≤ 0.95	13 (21.0)	1 (1.6)	2 (3.2)	1 (1.6)	
CCC > 0.95	49 (79.0)	61 (98.4)	60 (96.8)	61 (98.4)	
P-value		0.001	1.000	1.000	
Algorithm 3-Initial parameter 1 vs Initial parameter 2					
CCC < 0.75	23 (37.1)	4 (6.4)	3 (4.8)	3 (4.8)	<0.001
CCC ≥ 0.75 & CCC ≤ 0.95	37 (59.7)	46 (74.2)	28 (45.2)	24 (38.7)	
CCC > 0.95	2 (3.2)	10 (19.4)	31 (50.0)	35 (56.5)	
P-value		< 0.001	0.001	0.865	
Algorithm 3-Initial parameter 1 vs Initial parameter 3					
CCC < 0.75	0 (0)	0 (0)	0 (0)	0 (0)	<0.001
CCC ≥ 0.75 & CCC ≤ 0.95	10 (16.1)	1 (1.6)	0 (0)	0 (0)	
CCC > 0.95	52 (83.9)	61 (98.4)	62 (100)	62 (100)	
P-value		0.008	1.000	1.000	
Algorithm 3-Initial parameter 2 vs Initial parameter 3					
CCC < 0.75	0 (0)	0 (0)	0 (0)	0 (0)	<0.001
CCC ≥ 0.75 & CCC ≤ 0.95	30 (48.4)	2 (3.2)	1 (1.6)	0 (0)	
CCC > 0.95	32 (51.6)	60 (96.8)	61 (98.4)	62 (100)	
P-value		< 0.001	1.000	1.000	
Algorithm 1 vs Algorithm 2					
CCC < 0.75	0 (0)	0 (0)	0 (0)	0 (0)	<0.001
CCC ≥ 0.75 & CCC ≤ 0.95	30 (48.4)	2 (3.2)	1 (1.6)	0 (0)	
CCC > 0.95	32 (51.6)	60 (96.8)	61 (98.4)	62 (100)	
P-value		< 0.001	1.000	1.000	
Algorithm 1 vs Algorithm 3					
CCC < 0.75	20 (32.3)	2 (3.2)	1 (1.6)	1 (1.6)	<0.001
CCC ≥ 0.75 & CCC ≤ 0.95	34 (54.8)	32 (51.6)	22 (35.5)	19 (30.7)	
CCC > 0.95	8 (12.9)	28 (45.2)	39 (62.9)	42 (67.7)	
P-value		< 0.001	0.118	0.850	
Algorithm 2 vs Algorithm 3					
CCC < 0.75	18 (29.0)	3 (4.8)	2 (3.2)	2 (3.2)	<0.001
CCC ≥ 0.75 & CCC ≤ 0.95	41 (66.1)	45 (72.6)	32 (51.6)	24 (38.7)	
CCC > 0.95	3 (4.8)	14 (22.6)	28 (45.2)	36 (58.1)	
P-value		< 0.001	0.025	0.331	

Table A.14

Distribution of stability groups of 3D Laws peritumoral radiomic features extracted from ROIs that are not bounded by a lung mask.

	Distance				P- Value
	0 – 3 mm	0 – 6 mm	0 – 9 mm	0 – 12	
Algorithm 1-Initial parameter 1 vs Initial parameter 2					
CCC < 0.75	30 (24.0)	0 (0)	8 (6.4)	20 (16.0)	
CCC ≥ 0.75 & CCC ≤ 0.95	58 (46.4)	2 (1.6)	0 (0)	5 (4.0)	
CCC > 0.95	37 (29.6)	123 (98.4)	117 (93.6)	100 (80.0)	<0.001
	P-value	<0.001	0.003	0.002	
Algorithm 1-Initial parameter 1 vs Initial parameter 3					
CCC < 0.75	36 (28.8)	17 (13.6)	27 (21.6)	20 (16.0)	
CCC ≥ 0.75 & CCC ≤ 0.95	43 (34.4)	3 (2.4)	13 (10.4)	1 (0.8)	
CCC > 0.95	46 (36.8)	105 (84.0)	85 (68.0)	104 (83.2)	<0.001
	P-value	<0.001	0.005	0.001	
Algorithm 1-Initial parameter 2 vs Initial parameter 3					
CCC < 0.75	32 (25.6)	18 (14.4)	32 (25.6)	20 (16.0)	
CCC ≥ 0.75 & CCC ≤ 0.95	31 (24.8)	2 (1.6)	8 (6.4)	0 (0)	
CCC > 0.95	62 (49.6)	105 (84.0)	85 (68.0)	105 (84.0)	<0.001
	P-value	<0.001	0.007	0.001	
Algorithm 2-Initial parameter 1 vs Initial parameter 2					
CCC < 0.75	40 (32.0)	27 (21.6)	33 (26.4)	42 (33.6)	
CCC ≥ 0.75 & CCC ≤ 0.95	82 (65.6)	35 (28.0)	18 (14.4)	5 (4.0)	
CCC > 0.95	3 (2.4)	63 (50.4)	74 (59.2)	78 (62.4)	<0.001
	P-value	<0.001	0.035	0.013	
Algorithm 2-Initial parameter 1 vs Initial parameter 3					
CCC < 0.75	32 (25.6)	22 (17.6)	39 (31.2)	44 (35.2)	
CCC ≥ 0.75 & CCC ≤ 0.95	87 (69.6)	30 (24.0)	5 (4.0)	0 (0)	
CCC > 0.95	6 (4.8)	73 (58.4)	81 (64.8)	81 (64.8)	<0.001
	P-value	<0.001	<0.001	0.035	
Algorithm 2-Initial parameter 2 vs Initial parameter 3					
CCC < 0.75	29 (23.2)	1 (0.8)	32 (25.6)	10 (8.0)	
CCC ≥ 0.75 & CCC ≤ 0.95	85 (68.0)	29 (23.2)	3 (2.4)	34 (27.2)	
CCC > 0.95	11 (8.8)	95 (76.0)	90 (72.0)	81 (64.8)	<0.001
	P-value	<0.001	<0.001	<0.001	
Algorithm 3-Initial parameter 1 vs Initial parameter 2					
CCC < 0.75	123 (98.4)	98 (78.4)	44 (35.2)	32 (25.6)	
CCC ≥ 0.75 & CCC ≤ 0.95	2 (1.6)	27 (21.6)	81 (64.8)	86 (68.8)	
CCC > 0.95	0 (0)	0 (0)	0 (0)	7 (5.6)	<0.001
	P-value	<0.001	<0.001	0.008	
Algorithm 3-Initial parameter 1 vs Initial parameter 3					
CCC < 0.75	41 (32.8)	30 (24.0)	31 (24.8)	2 (1.6)	
CCC ≥ 0.75 & CCC ≤ 0.95	76 (60.8)	16 (12.8)	13 (10.4)	28 (22.4)	
CCC > 0.95	8 (6.4)	79 (63.2)	81 (64.8)	95 (76.0)	<0.001
	P-value	<0.001	0.861	<0.001	
Algorithm 3-Initial parameter 2 vs Initial parameter 3					
CCC < 0.75	57 (45.6)	32 (25.6)	32 (25.6)	32 (25.6)	
CCC ≥ 0.75 & CCC ≤ 0.95	68 (54.4)	69 (55.2)	12 (9.6)	6 (4.8)	
CCC > 0.95	0 (0)	24 (19.2)	81 (64.8)	87 (69.6)	<0.001
	P-value	<0.001	<0.001	0.332	
Algorithm 1 vs Algorithm 2					
CCC < 0.75	80 (64.0)	45 (36.0)	44 (35.2)	44 (35.2)	
CCC ≥ 0.75 & CCC ≤ 0.95	45 (36.0)	79 (63.2)	39 (31.2)	6 (4.8)	
CCC > 0.95	0 (0)	1 (0.8)	42 (33.6)	75 (60.0)	<0.001
	P-value	<0.001	<0.001	0.036	
Algorithm 1 vs Algorithm 3					
CCC < 0.75	125 (100)	99 (79.2)	28 (22.4)	44 (35.2)	
CCC ≥ 0.75 & CCC ≤ 0.95	0 (0)	26 (20.8)	97 (77.6)	81 (64.8)	
CCC > 0.95	0 (0)	0 (0)	0 (0)	0 (0)	<0.001
	P-value	<0.001	<0.001	<0.001	
Algorithm 2 vs Algorithm 3					
CCC < 0.75	125 (100)	102 (81.6)	44 (35.2)	44 (35.2)	
CCC ≥ 0.75 & CCC ≤ 0.95	0 (0)	23 (18.4)	81 (64.8)	76 (60.8)	
CCC > 0.95	0 (0)	0 (0)	0 (0)	5 (4.0)	<0.001
	P-value	<0.001	<0.001	0.088	

Table A.15

Distribution of stability groups of wavelet peritumoral radiomic features extracted from ROIs that are not bounded by a lung mask.

	Distance				P- Value
	0 – 3 mm	0 – 6 mm	0 – 9 mm	0 – 12	
Algorithm 1-Initial parameter 1 vs Initial parameter 2					
CCC < 0.75	0 (0)	0 (0)	1 (3.3)	3 (10.0)	
CCC ≥ 0.75 & CCC ≤ 0.95	7 (23.3)	15 (50.0)	22 (73.3)	11 (36.7)	
CCC > 0.95	23 (76.7)	15 (50.0)	7 (23.3)	16 (53.3)	<0.001
P-value		0.060	0.060	0.012	
Algorithm 1-Initial parameter 1 vs Initial parameter 3					
CCC < 0.75	0 (0)	0 (0)	1 (3.3)	5 (16.7)	
CCC ≥ 0.75 & CCC ≤ 0.95	12 (40.0)	13 (43.3)	25 (83.3)	10 (33.3)	
CCC > 0.95	18 (60.0)	17 (56.7)	4 (13.3)	15 (50.0)	<0.001
P-value		1.000	0.001	<0.001	
Algorithm 1-Initial parameter 2 vs Initial parameter 3					
CCC < 0.75	0 (0)	0 (0)	0 (0)	0 (0)	
CCC ≥ 0.75 & CCC ≤ 0.95	0 (0)	15 (50.0)	1 (3.3)	4 (13.3)	
CCC > 0.95	30 (100)	15 (50.0)	29 (96.7)	26 (86.7)	<0.001
P-value		<0.001	<0.001	0.353	
Algorithm 2-Initial parameter 1 vs Initial parameter 2					
CCC < 0.75	7 (23.3)	0 (0)	1 (3.3)	0 (0)	
CCC ≥ 0.75 & CCC ≤ 0.95	22 (73.3)	30 (100)	29 (96.7)	6 (20.0)	
CCC > 0.95	1 (3.3)	0 (0)	0 (0)	24 (80.0)	<0.001
P-value		0.005	1.000	<0.001	
Algorithm 2-Initial parameter 1 vs Initial parameter 3					
CCC < 0.75	7 (23.3)	0 (0)	1 (3.3)	1 (3.3)	
CCC ≥ 0.75 & CCC ≤ 0.95	22 (73.3)	29 (96.7)	29 (96.7)	5 (16.7)	
CCC > 0.95	1 (3.3)	1 (3.3)	0 (0)	24 (80.0)	<0.001
P-value		0.011	1.000	<0.001	
Algorithm 2-Initial parameter 2 vs Initial parameter 3					
CCC < 0.75	0 (0)	0 (0)	0 (0)	0 (0)	
CCC ≥ 0.75 & CCC ≤ 0.95	4 (13.3)	3 (10.0)	2 (6.7)	2 (6.7)	
CCC > 0.95	26 (86.7)	27 (90.0)	28 (93.3)	28 (93.3)	0.900
P-value		1.000	1.000	1.000	
Algorithm 3-Initial parameter 1 vs Initial parameter 2					
CCC < 0.75	18 (60.0)	14 (46.7)	15 (50.0)	26 (86.7)	
CCC ≥ 0.75 & CCC ≤ 0.95	12 (40.0)	16 (53.3)	15 (50.0)	4 (13.3)	
CCC > 0.95	0 (0)	0 (0)	0 (0)	0 (0)	0.004
P-value		0.438	1.000	0.005	
Algorithm 3-Initial parameter 1 vs Initial parameter 3					
CCC < 0.75	1 (3.3)	2 (6.7)	1 (3.3)	0 (0)	
CCC ≥ 0.75 & CCC ≤ 0.95	23 (76.7)	28 (93.3)	13 (43.3)	2 (66.7)	
CCC > 0.95	6 (20.0)	0 (0)	16 (53.3)	10 (33.3)	<0.001
P-value		0.024	<0.001	0.119	
Algorithm 3-Initial parameter 2 vs Initial parameter 3					
CCC < 0.75	0 (0)	16 (53.3)	0 (0)	0 (0)	
CCC ≥ 0.75 & CCC ≤ 0.95	29 (96.7)	14 (46.7)	1 (3.3)	25 (83.3)	
CCC > 0.95	1 (3.3)	0 (0)	29 (96.7)	5 (16.7)	<0.001
P-value		<0.001	<0.001	<0.001	
Algorithm 1 vs Algorithm 2					
CCC < 0.75	24 (80.0)	25 (83.3)	17 (56.7)	20 (66.7)	
CCC ≥ 0.75 & CCC ≤ 0.95	6 (20.0)	5 (16.7)	13 (43.3)	10 (33.3)	
CCC > 0.95	0 (0)	0 (0)	0 (0)	0 (0)	0.096
P-value		1.000	0.047	0.596	
Algorithm 1 vs Algorithm 3					
CCC < 0.75	15 (50.0)	15 (50.0)	15 (50.0)	4 (13.3)	
CCC ≥ 0.75 & CCC ≤ 0.95	15 (50.0)	15 (50.0)	5 (16.7)	26 (86.7)	
CCC > 0.95	0 (0)	0 (0)	10 (33.3)	0 (0)	<0.001
P-value		1.000	<0.001	<0.001	
Algorithm 2 vs Algorithm 3					
CCC < 0.75	17 (56.7)	14 (46.7)	22 (73.3)	11 (36.7)	
CCC ≥ 0.75 & CCC ≤ 0.95	13 (43.3)	16 (53.3)	8 (26.7)	19 (63.3)	
CCC > 0.95	0 (0)	0 (0)	0 (0)	0 (0)	0.032
P-value		0.606	0.064	0.009	

Table A.16
Distribution of stability groups of all intratumoral radiomic features.

Comparisons groups	Counts
Algorithm 1-Initial parameter 1 vs Initial parameter 2	
CCC < 0.75	12 (3.9)
CCC ≥ 0.75 & CCC ≤ 0.95	62 (20.4)
CCC > 0.95	230 (75.7)
Algorithm 1-Initial parameter 1 vs Initial parameter 3	
CCC < 0.75	14 (4.6)
CCC ≥ 0.75 & CCC ≤ 0.95	63 (20.7)
CCC > 0.95	227 (74.7)
Algorithm 1-Initial parameter 2 vs Initial parameter 3	
CCC < 0.75	8 (2.6)
CCC ≥ 0.75 & CCC ≤ 0.95	51 (16.8)
CCC > 0.95	245 (80.6)
Algorithm 2-Initial parameter 1 vs Initial parameter 2	
CCC < 0.75	83 (27.3)
CCC ≥ 0.75 & CCC ≤ 0.95	160 (52.6)
CCC > 0.95	61 (20.1)
Algorithm 2-Initial parameter 1 vs Initial parameter 3	
CCC < 0.75	58 (19.1)
CCC ≥ 0.75 & CCC ≤ 0.95	160 (52.6)
CCC > 0.95	86 (28.3)
Algorithm 2-Initial parameter 2 vs Initial parameter 3	
CCC < 0.75	55 (18.1)
CCC ≥ 0.75 & CCC ≤ 0.95	134 (44.1)
CCC > 0.95	115 (37.8)
Algorithm 3-Initial parameter 1 vs Initial parameter 2	
CCC < 0.75	219 (72.0)
CCC ≥ 0.75 & CCC ≤ 0.95	78 (25.7)
CCC > 0.95	7 (2.3)
Algorithm 3-Initial parameter 1 vs Initial parameter 3	
CCC < 0.75	33 (10.9)
CCC ≥ 0.75 & CCC ≤ 0.95	70 (23.0)
CCC > 0.95	201 (66.1)
Algorithm 3-Initial parameter 2 vs Initial parameter 3	
CCC < 0.75	36 (11.8)
CCC ≥ 0.75 & CCC ≤ 0.95	110 (36.2)
CCC > 0.95	158 (52.0)
Algorithm 1 vs Algorithm 2	
CCC < 0.75	157 (51.6)
CCC ≥ 0.75 & CCC ≤ 0.95	140 (46.1)
CCC > 0.95	7 (2.3)
Algorithm 1 vs Algorithm 3	
CCC < 0.75	134 (44.1)
CCC ≥ 0.75 & CCC ≤ 0.95	150 (49.3)
CCC > 0.95	20 (6.6)
Algorithm 2 vs Algorithm 3	
CCC < 0.75	192 (63.2)
CCC ≥ 0.75 & CCC ≤ 0.95	108 (35.5)
CCC > 0.95	4 (1.3)

Numbers inside parenthesis are the percentage values.

Table A.17
Distribution of stability groups of intratumoral statistical features.

Comparisons groups	Counts
Algorithm 1-Initial parameter 1 vs Initial parameter 2	
CCC < 0.75	0 (0)
CCC ≥ 0.75 & CCC ≤ 0.95	0 (0)
CCC > 0.95	19 (100)
Algorithm 1-Initial parameter 1 vs Initial parameter 3	
CCC < 0.75	0 (0)
CCC ≥ 0.75 & CCC ≤ 0.95	0 (0)
CCC > 0.95	19 (100)
Algorithm 1-Initial parameter 2 vs Initial parameter 3	
CCC < 0.75	0 (0)
CCC ≥ 0.75 & CCC ≤ 0.95	0 (0)
CCC > 0.95	19 (100)
Algorithm 2-Initial parameter 1 vs Initial parameter 2	
CCC < 0.75	5 (26.3)
CCC ≥ 0.75 & CCC ≤ 0.95	8 (42.1)
CCC > 0.95	6 (31.6)
Algorithm 2-Initial parameter 1 vs Initial parameter 3	
CCC < 0.75	1 (5.3)
CCC ≥ 0.75 & CCC ≤ 0.95	10 (52.6)
CCC > 0.95	8 (42.1)
Algorithm 2-Initial parameter 2 vs Initial parameter 3	
CCC < 0.75	5 (26.3)
CCC ≥ 0.75 & CCC ≤ 0.95	8 (42.1)
CCC > 0.95	6 (31.6)
Algorithm 3-Initial parameter 1 vs Initial parameter 2	
CCC < 0.75	15 (79.0)
CCC ≥ 0.75 & CCC ≤ 0.95	4 (21.0)
CCC > 0.95	0 (0)
Algorithm 3-Initial parameter 1 vs Initial parameter 3	
CCC < 0.75	0 (0)
CCC ≥ 0.75 & CCC ≤ 0.95	10 (52.6)
CCC > 0.95	9 (47.4)
Algorithm 3-Initial parameter 2 vs Initial parameter 3	
CCC < 0.75	0 (0)
CCC ≥ 0.75 & CCC ≤ 0.95	12 (63.2)
CCC > 0.95	7 (36.8)
Algorithm 1 vs Algorithm 2	
CCC < 0.75	12 (63.2)
CCC ≥ 0.75 & CCC ≤ 0.95	7 (36.8)
CCC > 0.95	0 (0)
Algorithm 1 vs Algorithm 3	
CCC < 0.75	12 (63.2)
CCC ≥ 0.75 & CCC ≤ 0.95	6 (31.6)
CCC > 0.95	1 (5.3)
Algorithm 2 vs Algorithm 3	
CCC < 0.75	14 (73.7)
CCC ≥ 0.75 & CCC ≤ 0.95	5 (26.3)
CCC > 0.95	0 (0)

Numbers inside parenthesis are the percentage values.

Table A.18
Distribution of stability groups of intratumoral histogram features.

Comparisons groups	Counts
Algorithm 1-Initial parameter 1 vs Initial parameter 2	
CCC < 0.75	2 (7.1)
CCC ≥ 0.75 & CCC ≤ 0.95	3 (10.7)
CCC > 0.95	23 (82.1)
Algorithm 1-Initial parameter 1 vs Initial parameter 3	
CCC < 0.75	2 (7.1)
CCC ≥ 0.75 & CCC ≤ 0.95	2 (7.1)
CCC > 0.95	24 (85.7)
Algorithm 1-Initial parameter 2 vs Initial parameter 3	
CCC < 0.75	1 (3.6)
CCC ≥ 0.75 & CCC ≤ 0.95	4 (14.3)
CCC > 0.95	23 (82.1)
Algorithm 2-Initial parameter 1 vs Initial parameter 2	
CCC < 0.75	10 (35.7)
CCC ≥ 0.75 & CCC ≤ 0.95	14 (50.0)
CCC > 0.95	4 (14.3)
Algorithm 2-Initial parameter 1 vs Initial parameter 3	
CCC < 0.75	3 (10.7)
CCC ≥ 0.75 & CCC ≤ 0.95	17 (60.7)
CCC > 0.95	8 (28.6)
Algorithm 2-Initial parameter 2 vs Initial parameter 3	
CCC < 0.75	5 (17.9)
CCC ≥ 0.75 & CCC ≤ 0.95	19 (67.9)
CCC > 0.95	4 (14.3)
Algorithm 3-Initial parameter 1 vs Initial parameter 2	
CCC < 0.75	24 (85.7)
CCC ≥ 0.75 & CCC ≤ 0.95	4 (14.3)
CCC > 0.95	0 (0)
Algorithm 3-Initial parameter 1 vs Initial parameter 3	
CCC < 0.75	0 (0)
CCC ≥ 0.75 & CCC ≤ 0.95	13 (46.4)
CCC > 0.95	15 (53.6)
Algorithm 3-Initial parameter 2 vs Initial parameter 3	
CCC < 0.75	1 (3.6)
CCC ≥ 0.75 & CCC ≤ 0.95	16 (57.1)
CCC > 0.95	11 (39.3)
Algorithm 1 vs Algorithm 2	
CCC < 0.75	0 (0)
CCC ≥ 0.75 & CCC ≤ 0.95	22 (78.6)
CCC > 0.95	6 (21.4)
Algorithm 1 vs Algorithm 3	
CCC < 0.75	13 (46.4)
CCC ≥ 0.75 & CCC ≤ 0.95	14 (50.0)
CCC > 0.95	1 (3.6)
Algorithm 2 vs Algorithm 3	
CCC < 0.75	24 (85.7)
CCC ≥ 0.75 & CCC ≤ 0.95	4 (14.3)
CCC > 0.95	0 (0)

Numbers inside parenthesis are the percentage values.

Table A.19
Distribution of stability groups of features intratumoral GLCM, GLRLM, GLSZM and NGTDM texture features.

Comparisons groups	Counts
Algorithm 1-Initial parameter 1 vs Initial parameter 2	
CCC < 0.75	1 (1.6)
CCC ≥ 0.75 & CCC ≤ 0.95	6 (9.7)
CCC > 0.95	55 (88.7)
Algorithm 1-Initial parameter 1 vs Initial parameter 3	
CCC < 0.75	0 (0)
CCC ≥ 0.75 & CCC ≤ 0.95	8 (12.9)
CCC > 0.95	24 (87.1)
Algorithm 1-Initial parameter 2 vs Initial parameter 3	
CCC < 0.75	0 (0)
CCC ≥ 0.75 & CCC ≤ 0.95	6 (9.7)
CCC > 0.95	56 (90.3)
Algorithm 2-Initial parameter 1 vs Initial parameter 2	
CCC < 0.75	19 (30.7)
CCC ≥ 0.75 & CCC ≤ 0.95	26 (41.9)
CCC > 0.95	17 (27.4)
Algorithm 2-Initial parameter 1 vs Initial parameter 3	
CCC < 0.75	12 (19.3)
CCC ≥ 0.75 & CCC ≤ 0.95	21 (33.9)
CCC > 0.95	29 (46.8)
Algorithm 2-Initial parameter 2 vs Initial parameter 3	
CCC < 0.75	18 (29.0)
CCC ≥ 0.75 & CCC ≤ 0.95	22 (35.5)
CCC > 0.95	22 (35.5)
Algorithm 3-Initial parameter 1 vs Initial parameter 2	
CCC < 0.75	44 (71.0)
CCC ≥ 0.75 & CCC ≤ 0.95	11 (17.7)
CCC > 0.95	7 (11.3)
Algorithm 3-Initial parameter 1 vs Initial parameter 3	
CCC < 0.75	0 (0)
CCC ≥ 0.75 & CCC ≤ 0.95	10 (16.1)
CCC > 0.95	52 (83.9)
Algorithm 3-Initial parameter 2 vs Initial parameter 3	
CCC < 0.75	2 (3.2)
CCC ≥ 0.75 & CCC ≤ 0.95	31 (50.0)
CCC > 0.95	29 (46.8)
Algorithm 1 vs Algorithm 2	
CCC < 0.75	27 (43.5)
CCC ≥ 0.75 & CCC ≤ 0.95	28 (45.2)
CCC > 0.95	7 (11.3)
Algorithm 1 vs Algorithm 3	
CCC < 0.75	22 (35.5)
CCC ≥ 0.75 & CCC ≤ 0.95	28 (45.2)
CCC > 0.95	12 (19.3)
Algorithm 2 vs Algorithm 3	
CCC < 0.75	45 (72.6)
CCC ≥ 0.75 & CCC ≤ 0.95	13 (21.0)
CCC > 0.95	4 (6.4)

Numbers inside parenthesis are the percentage values.

Table A.20
Distribution of stability groups of intratumoral 3D Laws texture features.

Comparisons groups	Counts
Algorithm 1-Initial parameter 1 vs Initial parameter 2	
CCC < 0.75	9 (7.2)
CCC ≥ 0.75 & CCC ≤ 0.95	35 (28.0)
CCC > 0.95	81 (64.8)
Algorithm 1-Initial parameter 1 vs Initial parameter 3	
CCC < 0.75	12 (9.6)
CCC ≥ 0.75 & CCC ≤ 0.95	38 (30.4)
CCC > 0.95	75 (60.0)
Algorithm 1-Initial parameter 2 vs Initial parameter 3	
CCC < 0.75	7 (5.6)
CCC ≥ 0.75 & CCC ≤ 0.95	31 (24.8)
CCC > 0.95	87 (69.6)
Algorithm 2-Initial parameter 1 vs Initial parameter 2	
CCC < 0.75	40 (32.0)
CCC ≥ 0.75 & CCC ≤ 0.95	73 (58.4)
CCC > 0.95	12 (9.6)
Algorithm 2-Initial parameter 1 vs Initial parameter 3	
CCC < 0.75	41 (32.8)
CCC ≥ 0.75 & CCC ≤ 0.95	71 (56.8)
CCC > 0.95	13 (10.4)
Algorithm 2-Initial parameter 2 vs Initial parameter 3	
CCC < 0.75	24 (19.2)
CCC ≥ 0.75 & CCC ≤ 0.95	60 (48.0)
CCC > 0.95	41 (32.8)
Algorithm 3-Initial parameter 1 vs Initial parameter 2	
CCC < 0.75	95 (76.0)
CCC ≥ 0.75 & CCC ≤ 0.95	30 (24.0)
CCC > 0.95	0 (0)
Algorithm 3-Initial parameter 1 vs Initial parameter 3	
CCC < 0.75	32 (25.6)
CCC ≥ 0.75 & CCC ≤ 0.95	11 (8.8)
CCC > 0.95	82 (65.6)
Algorithm 3-Initial parameter 2 vs Initial parameter 3	
CCC < 0.75	32 (25.6)
CCC ≥ 0.75 & CCC ≤ 0.95	23 (18.4)
CCC > 0.95	70 (56.0)
Algorithm 1 vs Algorithm 2	
CCC < 0.75	59 (47.2)
CCC ≥ 0.75 & CCC ≤ 0.95	66 (52.8)
CCC > 0.95	0 (0)
Algorithm 1 vs Algorithm 3	
CCC < 0.75	60 (48.0)
CCC ≥ 0.75 & CCC ≤ 0.95	65 (52.0)
CCC > 0.95	0 (0)
Algorithm 2 vs Algorithm 3	
CCC < 0.75	66 (52.8)
CCC ≥ 0.75 & CCC ≤ 0.95	59 (47.2)
CCC > 0.95	0 (0)

Numbers inside parenthesis are the percentage values.

Table A.21
Distribution of stability groups of intratumoral wavelet texture features.

Comparisons groups	Counts
Algorithm 1-Initial parameter 1 vs Initial parameter 2	
CCC < 0.75	0 (0)
CCC ≥ 0.75 & CCC ≤ 0.95	11 (36.7)
CCC > 0.95	19 (63.3)
Algorithm 1-Initial parameter 1 vs Initial parameter 3	
CCC < 0.75	0 (0)
CCC ≥ 0.75 & CCC ≤ 0.95	7 (23.3)
CCC > 0.95	23 (76.7)
Algorithm 1-Initial parameter 2 vs Initial parameter 3	
CCC < 0.75	0 (0)
CCC ≥ 0.75 & CCC ≤ 0.95	5 (16.7)
CCC > 0.95	25 (83.3)
Algorithm 2-Initial parameter 1 vs Initial parameter 2	
CCC < 0.75	6 (20.0)
CCC ≥ 0.75 & CCC ≤ 0.95	23 (76.7)
CCC > 0.95	1 (3.3)
Algorithm 2-Initial parameter 1 vs Initial parameter 3	
CCC < 0.75	0 (0)
CCC ≥ 0.75 & CCC ≤ 0.95	30 (100)
CCC > 0.95	0 (0)
Algorithm 2-Initial parameter 2 vs Initial parameter 3	
CCC < 0.75	0 (0)
CCC ≥ 0.75 & CCC ≤ 0.95	14 (46.7)
CCC > 0.95	16 (53.3)
Algorithm 3-Initial parameter 1 vs Initial parameter 2	
CCC < 0.75	19 (63.3)
CCC ≥ 0.75 & CCC ≤ 0.95	11 (36.7)
CCC > 0.95	0 (0)
Algorithm 3-Initial parameter 1 vs Initial parameter 3	
CCC < 0.75	0 (0)
CCC ≥ 0.75 & CCC ≤ 0.95	19 (63.3)
CCC > 0.95	11 (36.7)
Algorithm 3-Initial parameter 2 vs Initial parameter 3	
CCC < 0.75	0 (0)
CCC ≥ 0.75 & CCC ≤ 0.95	19 (63.3)
CCC > 0.95	11 (36.7)
Algorithm 1 vs Algorithm 2	
CCC < 0.75	21 (70.0)
CCC ≥ 0.75 & CCC ≤ 0.95	9 (30.0)
CCC > 0.95	0 (0)
Algorithm 1 vs Algorithm 3	
CCC < 0.75	14 (46.7)
CCC ≥ 0.75 & CCC ≤ 0.95	12 (40.0)
CCC > 0.95	4 (13.3)
Algorithm 2 vs Algorithm 3	
CCC < 0.75	21 (70.0)
CCC ≥ 0.75 & CCC ≤ 0.95	9 (30.0)
CCC > 0.95	0 (0)

Numbers inside parenthesis are the percentage values.

Table A.22
Distribution of stability groups of intratumoral size features.

Comparisons groups	Counts
Algorithm 1-Initial parameter 1 vs Initial parameter 2	
CCC < 0.75	0 (0)
CCC ≥ 0.75 & CCC ≤ 0.95	0 (0)
CCC > 0.95	10 (100)
Algorithm 1-Initial parameter 1 vs Initial parameter 3	
CCC < 0.75	0 (0)
CCC ≥ 0.75 & CCC ≤ 0.95	0 (0)
CCC > 0.95	10 (100)
Algorithm 1-Initial parameter 2 vs Initial parameter 3	
CCC < 0.75	0 (0)
CCC ≥ 0.75 & CCC ≤ 0.95	0 (0)
CCC > 0.95	10 (100)
Algorithm 2-Initial parameter 1 vs Initial parameter 2	
CCC < 0.75	0 (0)
CCC ≥ 0.75 & CCC ≤ 0.95	0 (0)
CCC > 0.95	10 (100)
Algorithm 2-Initial parameter 1 vs Initial parameter 3	
CCC < 0.75	0 (0)
CCC ≥ 0.75 & CCC ≤ 0.95	0 (0)
CCC > 0.95	10 (100)
Algorithm 2-Initial parameter 2 vs Initial parameter 3	
CCC < 0.75	0 (0)
CCC ≥ 0.75 & CCC ≤ 0.95	0 (0)
CCC > 0.95	10 (100)
Algorithm 3-Initial parameter 1 vs Initial parameter 2	
CCC < 0.75	5 (50.0)
CCC ≥ 0.75 & CCC ≤ 0.95	5 (50.0)
CCC > 0.95	0 (0)
Algorithm 3-Initial parameter 1 vs Initial parameter 3	
CCC < 0.75	0 (0)
CCC ≥ 0.75 & CCC ≤ 0.95	0 (0)
CCC > 0.95	10 (100)
Algorithm 3-Initial parameter 2 vs Initial parameter 3	
CCC < 0.75	0 (0)
CCC ≥ 0.75 & CCC ≤ 0.95	0 (0)
CCC > 0.95	10 (100)
Algorithm 1 vs Algorithm 2	
CCC < 0.75	1 (10.0)
CCC ≥ 0.75 & CCC ≤ 0.95	9 (90.0)
CCC > 0.95	0 (0)
Algorithm 1 vs Algorithm 3	
CCC < 0.75	0 (0)
CCC ≥ 0.75 & CCC ≤ 0.95	8 (80.0)
CCC > 0.95	2 (20.0)
Algorithm 2 vs Algorithm 3	
CCC < 0.75	5 (50.0)
CCC ≥ 0.75 & CCC ≤ 0.95	5 (50.0)
CCC > 0.95	0 (0)

Numbers inside parenthesis are the percentage values.

Table A.23

Distribution of stability groups of intratumoral shape features.

Comparisons groups	Counts
Algorithm 1-Initial parameter 1 vs Initial parameter 2	
CCC < 0.75	0 (0)
CCC ≥ 0.75 & CCC ≤ 0.95	7 (23.3)
CCC > 0.95	23 (76.7)
Algorithm 1-Initial parameter 1 vs Initial parameter 3	
CCC < 0.75	0 (0)
CCC ≥ 0.75 & CCC ≤ 0.95	8 (26.7)
CCC > 0.95	22 (73.3)
Algorithm 1-Initial parameter 2 vs Initial parameter 3	
CCC < 0.75	0 (0)
CCC ≥ 0.75 & CCC ≤ 0.95	5 (16.7)
CCC > 0.95	25 (83.3)
Algorithm 2-Initial parameter 1 vs Initial parameter 2	
CCC < 0.75	3 (10.0)
CCC ≥ 0.75 & CCC ≤ 0.95	16 (53.3)
CCC > 0.95	11 (36.7)
Algorithm 2-Initial parameter 1 vs Initial parameter 3	
CCC < 0.75	1 (3.3)
CCC ≥ 0.75 & CCC ≤ 0.95	11 (36.7)
CCC > 0.95	18 (60.0)
Algorithm 2-Initial parameter 2 vs Initial parameter 3	
CCC < 0.75	3 (10.0)
CCC ≥ 0.75 & CCC ≤ 0.95	11 (36.7)
CCC > 0.95	16 (53.3)
Algorithm 3-Initial parameter 1 vs Initial parameter 2	
CCC < 0.75	17 (56.7)
CCC ≥ 0.75 & CCC ≤ 0.95	13 (43.3)
CCC > 0.95	0 (0)
Algorithm 3-Initial parameter 1 vs Initial parameter 3	
CCC < 0.75	1 (3.3)
CCC ≥ 0.75 & CCC ≤ 0.95	7 (23.3)
CCC > 0.95	22 (73.3)
Algorithm 3-Initial parameter 2 vs Initial parameter 3	
CCC < 0.75	1 (3.3)
CCC ≥ 0.75 & CCC ≤ 0.95	9 (30.0)
CCC > 0.95	20 (66.7)
Algorithm 1 vs Algorithm 2	
CCC < 0.75	15 (50.0)
CCC ≥ 0.75 & CCC ≤ 0.95	15 (50.0)
CCC > 0.95	0 (0)
Algorithm 1 vs Algorithm 3	
CCC < 0.75	13 (43.3)
CCC ≥ 0.75 & CCC ≤ 0.95	17 (56.7)
CCC > 0.95	0 (0)
Algorithm 2 vs Algorithm 3	
CCC < 0.75	17 (56.7)
CCC ≥ 0.75 & CCC ≤ 0.95	13 (43.3)
CCC > 0.95	0 (0)

Numbers inside parenthesis are the percentage values.

Table A.24
Distribution of reproducibility groups of intratumoral radiomic features.

Feature Groups	Counts
All features	
CCC < 0.75	95 (31.2)
CCC ≥ 0.75 & CCC ≤ 0.95	154 (50.7)
CCC > 0.95	55 (18.1)
Statistical features	
CCC < 0.75	1 (5.3)
CCC ≥ 0.75 & CCC ≤ 0.95	13 (68.4)
CCC > 0.95	5 (26.3)
Histogram features	
CCC < 0.75	3 (10.7)
CCC ≥ 0.75 & CCC ≤ 0.95	20 (71.4)
CCC > 0.95	5 (17.9)
Texture features¹	
CCC < 0.75	3 (4.8)
CCC ≥ 0.75 & CCC ≤ 0.95	20 (69.4)
CCC > 0.95	5 (25.8)
3D Laws texture features	
CCC < 0.75	66 (52.8)
CCC ≥ 0.75 & CCC ≤ 0.95	59 (47.2)
CCC > 0.95	0 (0)
Wavelet texture features	
CCC < 0.75	15 (50)
CCC ≥ 0.75 & CCC ≤ 0.95	0 (0)
CCC > 0.95	15 (50)
Size features	
CCC < 0.75	0 (0)
CCC ≥ 0.75 & CCC ≤ 0.95	0 (0)
CCC > 0.95	10 (100)
Shape features	
CCC < 0.75	7 (23.2)
CCC ≥ 0.75 & CCC ≤ 0.95	19 (63.3)
CCC > 0.95	4 (13.3)

Numbers inside parenthesis are the percentage values.

¹Features consist GLCM, GLRLM, GLSZM and NGTDM texture features.



Figure A.3 Concordance correlation coefficient groups of peritumoral features not bounded by lung parenchyma. The green boxes represent higher ($CCC > 0.95$), yellow boxes represent moderate ($CCC \geq 0.75$ and $CCC \leq 0.95$) and red boxes represent lower ($CCC < 0.75$) CCCs.

TRAINING COHORT

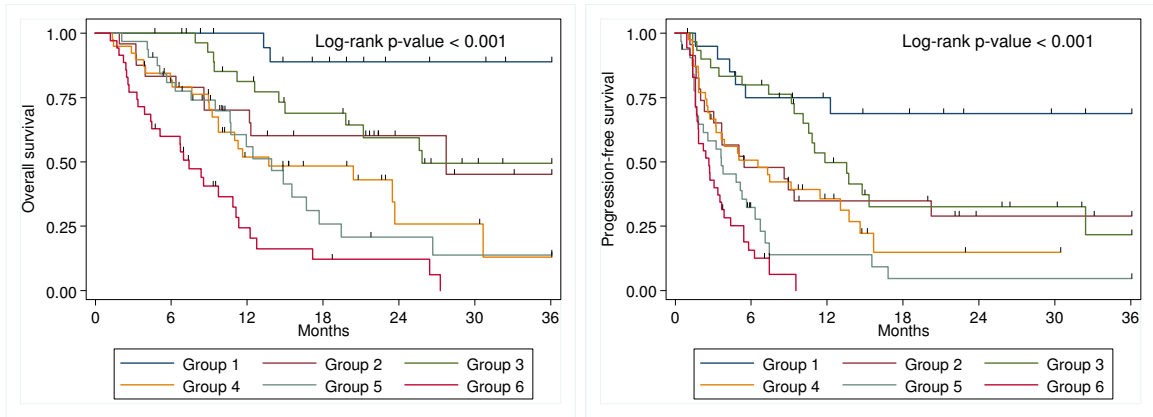


Figure A.4 Overall survival and progression-free survival for the six risk groups identified by CART in the training cohort. Groups 2, 3 and groups 4, 5 were combined for the analysis in Figure 6.4.

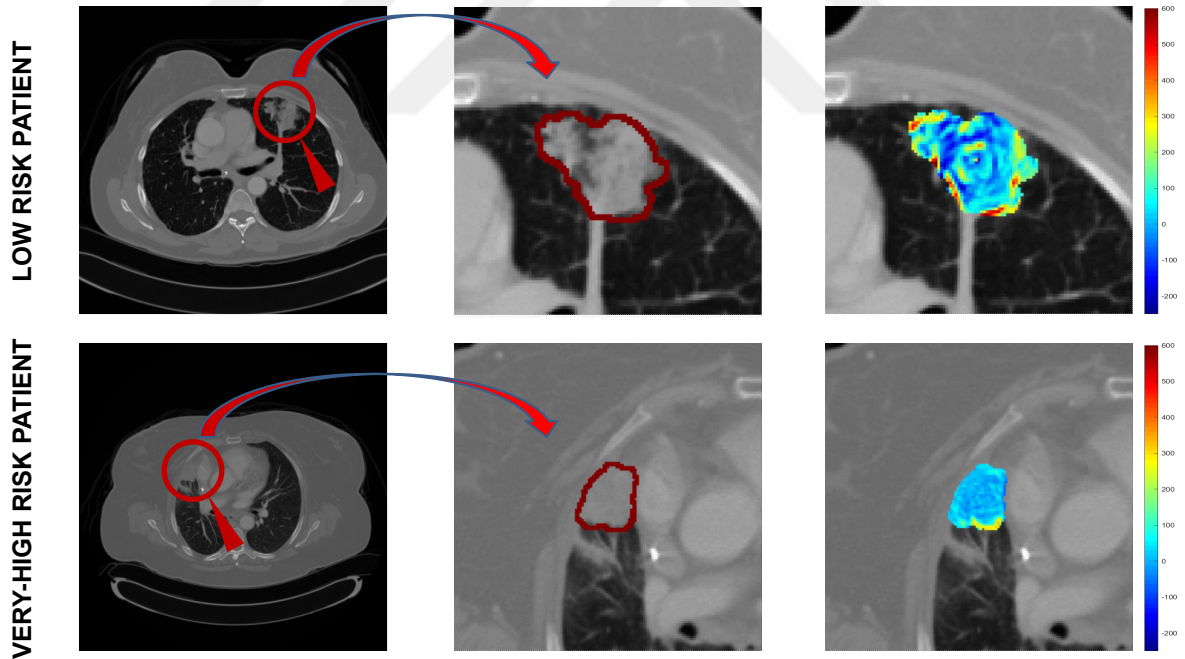


Figure A.5 Patients on low and very-high risk groups. First column represents the primary target lesion CT scan. Second column represents the tumor segmentation. Third column represents a gradient image of the segmented area for visualization of the tumor texture. Patient on the top was identified as a low risk patient to immunotherapy and had a less dense tumor phenotype with lower GLCM inverse difference score. Patient on the bottom was identified as a very-high risk patient and had a dense tumor phenotype with higher GLCM inverse score.

APPENDIX B. LIST OF PUBLICATIONS PRODUCED FROM THE THESIS

B.1 Scientific Journals

- (i) Radial gradient and radial deviation radiomic features from pre-surgical CT scans are associated with survival among lung adenocarcinoma patients, **I. Tunali**, O. Stringfield, A. Guvenis, H. Wang, Y. Liu, Y. Balagurunathan, P. Lambin, R. J. Gillies, M. B. Schabath, *Oncotarget*, Vol. 4, no. 56, pp. 96013-96026, 2017.
- (ii) Novel Clinical and Radiomic Predictors of Rapid Disease Progression Phenotypes among Lung Cancer Patients Treated with Immunotherapy: An Early Report, **I. Tunali**, J. E. Gray, J. Qi, D. K. Jeong, A. Guvenis, R. J. Gillies, M. B. Schabath, *Lung Cancer*, Vol. 129, pp. 75-79, 2019.
- (iii) Stability and reproducibility of computed tomography radiomic features extracted from peritumoral regions of lung cancer lesions, **I. Tunali**, L. O. Hall, S. Napel, D. Cherezov, A. Guvenis, R. J. Gillies, M. B. Schabath, *Medical Physics*, e-pub ahead of print, 2019.
- (iv) Hypoxia-related radiomics predict immunotherapy response: A multi-cohort study of NSCLC , **I. Tunali**, Y. Tan, J. E. Gray, E. Kasoulakis, S. A. Eschrich, T. A. Boyle, J. Saller, A. Guvenis, R. J. Gillies, M. B. Schabath, *Nature Communications*, *UNDER REVIEW*, 2019.

B.2 Conference Abstracts

- (i) A Fusion Method for Pulmonary Nodule Segmentation in Chest CT Exams, **I. Tunali**, A. Guvenis, *IEEE International Conference on Biomedical and Health Informatics 2016*, Las Vegas, USA, 24-27 Feb.

- (ii) Radiomics and clinical predictors of disease progression among non-small cell lung cancer patients treated with checkpoint inhibitors, **I. Tunali**, J. E. Gray, J. Qi, D. K. Jeong, A. Guvenis, R. J. Gillies, M. B. Schabath, *WMIC 2018*, Seattle, USA, 12-15 Sep.
- (iii) Clinical-radiomic models predict overall survival among non-small cell lung cancer patients treated with immunotherapy, **I. Tunali**, Y. Tan, J. E. Gray, S. Eschrich, A. Guvenis, R. J. Gillies, M. B. Schabath *NACLC 2019*, Chicago, USA, 09-12 Oct.



APPENDIX C. OTHER CONTRIBUTIONS TO LITERATURE

- (i) Quantitative imaging features predict response of immunotherapy in non-small cell lung cancer patients, **I. Tunali**, J. E. Gray, J. Qi, M. Abdallah, Y. Balagurunathan, R. J. Gillies, M. B. Schabath, *IASLC 17th World Conference on Lung Cancer 2016*, Vienna, Austria, 4-7 Dec.
- (ii) Quantitative imaging biomarkers predict immunotherapy response on non-small cell lung cancer patients, **I. Tunali**, J. E. Gray, J. Qi, M. Abdallah, Y. Balagurunathan, R. J. Gillies, M. B. Schabath, *EMBL-Cancer Core Europe Conference: Cancer Immunotherapy 2017*, Heidelberg, Germany, 2-4 Feb.
- (iii) Combining radiomics and mathematical modeling to elucidate mechanisms of resistance to immune checkpoint blockade in non-small cell lung cancer, D. Saeed-Vafa, R. Bravo, J.A. Dean, A. El-Kenawi, N.M. Pere, M. Strobl, D. Daniels, O. Stringfield, M. Damaghi, **I. Tunali**, L.V. Brown, L. Curtin, D. Nichol, H. Peck, R.J. Gillies, J.A. Gallaher, *bioRxiv*, Sep 2017, doi: <https://doi.org/10.1101/190561>.
- (iv) PET/CT imaging prediction of response to checkpoint blockade in advanced non-small cell lung cancer patients, W. Mu, J. Qi, H. Lu, M.B. Schabath, Y. Balagurunathan, **I. Tunali**, S. Pilon Thomas, R.J. Gillies, *Proceedings: AACR Annual Meeting 2018*, Chicago, USA, 14-18 Apr.
- (v) Radiomic biomarkers from PET/CT multi-modality fusion images for the prediction of immunotherapy response in advanced non-small cell lung cancer patients, W. Mu, J. Qi, H. Lu, M.B. Schabath, Y. Balagurunathan, **I. Tunali**, R.J. Gillies, *Medical Imaging: Computer-Aided Diagnosis 2018*, 105753S.
- (vi) Radiomics of 18F-FDG PET/CT images predicts clinical benefit of advanced NSCLC patients to checkpoint blockade immunotherapy, W. Mu, **I. Tunali**, J. E. Gray, J. Qi, M.B. Schabath, R.J. Gillies, *European Journal of Nuclear Medicine and Molecular Imaging*, 2019, pp 1-15.

- (vii) Radiomics of ^{18}F FDG PET/CT images predict immune-related severe adverse events in NSCLC patients, W. Mu, **I. Tunali**, J. Qi, M.B. Schabath, R.J. Gillies, *Radiology: Artificial Intelligence*, 2019, *ARTICLE IN PRESS*.
- (viii) Predicting PD-L1 expression and immunotherapy response using deep learning of PET/CT images: A multi-institutional study, W. Mu, L. Jiang, Chunyan W., Y. Shi, **I. Tunali**, J. E. Gray, J. Qi, J. Tian, M.B. Schabath, R.J. Gillies, *Nature Precision Medicine*, 2019, *TO BE SUBMITTED*.
- (ix) Peritumoral and intratumoral radiomic features predict survival outcomes among patients diagnosed in lung cancer screening, J. Perez-Morales, **I. Tunali**, O. Stringfield, S. Eschrich, Y. Balagurunathan, R.J. Gillies, M.B. Schabath, *Nature Scientific Reports*, 2019, *UNDER REVIEW*.
- (x) Predictors of disease progression and treatment response among lung cancer patients treated with immunotherapy, M. B. Schabath, **I. Tunali**, J. E. Gray, R. J. Gillies, *AACR Annual Meeting 2019*, Atlanta, USA, Mar 29- Apr 3.

REFERENCES

1. Siegel, R. L., K. D. Miller, and A. Jemal, "Cancer statistics, 2019," *CA Cancer J Clin*, Vol. 69, no. 1, pp. 7–34, 2019.
2. Borghaei, H., L. Paz-Ares, and L. Horn et al, "Nivolumab versus docetaxel in advanced nonsquamous non-small-cell lung cancer," *N Engl J Med*, Vol. 373, no. 17, pp. 1627–39, 2015.
3. Brahmer, J., K. L. Reckamp, and P. Baas et al, "Nivolumab versus docetaxel in advanced squamous-cell non-small-cell lung cancer," *N Engl J Med*, Vol. 373, no. 2, pp. 123–35, 2015.
4. Herbst, R. S., P. Baas, and D. W. Kim et al, "Pembrolizumab versus docetaxel for previously treated, pd-l1-positive, advanced non-small-cell lung cancer (keynote-010): a randomised controlled trial," *Lancet*, Vol. 387, no. 10027, pp. 1540–50, 2016.
5. Reck, M., D. Rodriguez-Abreu, A. G. Robinson, R. Hui, T. Csoszi, A. Fulop, M. Gottfried, N. Peled, A. Tafreshi, S. Cuffe, M. O'Brien, S. Rao, K. Hotta, M. A. Leiby, and G. M. Lubiniecki, "Pembrolizumab versus chemotherapy for pd-l1-positive non-small-cell lung cancer," *N Engl J Med*, Vol. 375, no. 19, pp. 1823–1833, 2016.
6. Rittmeyer, A., F. Barlesi, and D. Waterkamp et al, "Atezolizumab versus docetaxel in patients with previously treated non-small-cell lung cancer (oak): a phase 3, open-label, multicentre randomised controlled trial," *Lancet*, Vol. 389, no. 10066, pp. 255–265, 2017.
7. Gandhi, L., D. Rodriguez-Abreu, and S. Gadgeel et al, "Pembrolizumab plus chemotherapy in metastatic non-small-cell lung cancer," *N Engl J Med*, Vol. 378, no. 22, pp. 2078–2092, 2018.
8. Shukuya, T., and D. P. Carbone, "Predictive markers for the efficacy of anti-pd-1/pd-l1 antibodies in lung cancer," *J Thorac Oncol*, Vol. 11, no. 7, pp. 976–88, 2016.
9. Patel, S. P., and R. Kurzrock, "Pd-l1 expression as a predictive biomarker in cancer immunotherapy," *Molecular Cancer Therapeutics*, Vol. 14, no. 4, pp. 847–856, 2015.
10. Antonia, S. J., A. Villegas, and D. Daniel et al, "Durvalumab after chemoradiotherapy in stage iii non-small-cell lung cancer," *N Engl J Med*, Vol. 377, no. 20, pp. 1919–1929, 2017.
11. Seymour, L., J. Bogaerts, and A. Perrone, "irecist: guidelines for response criteria for use in trials testing immunotherapeutics (vol 18, pg e143, 2017)," *Lancet Oncology*, Vol. 20, no. 5, pp. E242–E242, 2019.
12. Gillies, R. J., P. E. Kinahan, and H. Hricak, "Radiomics: Images are more than pictures, they are data," *Radiology*, Vol. 278, no. 2, pp. 563–577, 2016.
13. Ganeshan, B., E. Panayiotou, K. Burnand, S. Dizdarevic, and K. Miles, "Tumour heterogeneity in non-small cell lung carcinoma assessed by ct texture analysis: a potential marker of survival," *Eur Radiol*, Vol. 22, no. 4, pp. 796–802, 2012.

14. Grove, O., A. E. Berglund, M. B. Schabath, H. J. Aerts, A. Dekker, H. Wang, E. R. Velazquez, P. Lambin, Y. Gu, Y. Balagurunathan, E. Eikman, R. A. Gatenby, S. Eschrich, and R. J. Gillies, "Quantitative computed tomographic descriptors associate tumor shape complexity and intratumor heterogeneity with prognosis in lung adenocarcinoma," *PLoS One*, Vol. 10, no. 3, p. e0118261, 2015.
15. Wang, H., M. B. Schabath, Y. Liu, A. E. Berglund, G. C. Bloom, J. Kim, O. Stringfield, E. A. Eikman, D. L. Klippenstein, J. J. Heine, S. A. Eschrich, Z. Ye, and R. J. Gillies, "Semiquantitative computed tomography characteristics for lung adenocarcinoma and their association with lung cancer survival," *Clin Lung Cancer*, Vol. 16, no. 6, pp. e141–63, 2015.
16. Aerts, H. J., 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 nslc," *Sci Rep*, Vol. 6, p. 33860, 2016.
17. Coroller, T. P., V. Agrawal, V. Narayan, Y. Hou, P. Grossmann, S. W. Lee, R. H. Mak, and H. J. Aerts, "Radiomic phenotype features predict pathological response in non-small cell lung cancer," *Radiother Oncol*, Vol. 119, no. 3, pp. 480–6, 2016.
18. Wang, H., M. B. Schabath, Y. Liu, Y. Han, Q. Li, R. J. Gillies, and Z. Ye, "Clinical and ct characteristics of surgically resected lung adenocarcinomas harboring alk rearrangements or egfr mutations," *Eur J Radiol*, Vol. 85, no. 11, pp. 1934–1940, 2016.
19. Kalpathy-Cramer, J., B. Zhao, D. Goldgof, Y. Gu, X. Wang, H. Yang, Y. Tan, R. Gillies, and S. Napel, "A comparison of lung nodule segmentation algorithms: Methods and results from a multi-institutional study," *J Digit Imaging*, Vol. 29, no. 4, pp. 476–87, 2016.
20. Zhao, B., L. P. James, C. S. Moskowitz, P. Guo, M. S. Ginsberg, R. A. Lefkowitz, Y. Qin, G. J. Riely, M. G. Kris, and L. H. Schwartz, "Evaluating variability in tumor measurements from same-day repeat ct scans of patients with non-small cell lung cancer," *Radiology*, Vol. 252, no. 1, pp. 263–72, 2009.
21. Howlader, N., A. Noone, M. Krapcho, D. Miller, A. Brest, M. Yu, J. Ruhl, Z. Tatalovich, A. Mariotto, D. Lewis, H. Chen, E. Feuer, and K. Cronin, "Seer cancer statistics review, 1975-2016, national cancer institute.," report, 2017.
22. Goldstraw, P., K. Chansky, and J. Crowley et al, "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," *J Thorac Oncol*, Vol. 11, no. 1, pp. 39–51, 2016.
23. Garraway, L. A., and E. S. Lander, "Lessons from the cancer genome," *Cell*, Vol. 153, no. 1, pp. 17–37, 2013.
24. Parsa, N., "Environmental factors inducing human cancers," *Iran J Public Health*, Vol. 41, no. 11, pp. 1–9, 2012.
25. Tomasetti, C., L. Li, and B. Vogelstein, "Stem cell divisions, somatic mutations, cancer etiology, and cancer prevention," *Science*, Vol. 355, no. 6331, pp. 1330–1334, 2017.
26. National Lung Screening Trial Research, T., D. R. Aberle, A. M. Adams, C. D. Berg, W. C. Black, J. D. Clapp, R. M. Fagerstrom, I. F. Gareen, C. Gatsonis, P. M. Marcus, and J. D. Sicks, "Reduced lung-cancer mortality with low-dose computed tomographic screening," *N Engl J Med*, Vol. 365, no. 5, pp. 395–409, 2011.

27. De Koning, H., C. Van der Aalst, K. Ten Haaf, and M. Oudkerk, "Effects of volume ct lung cancer screening: Mortality results of the nelson randomised-controlled population based trial," *Journal of Thoracic Oncology*, Vol. 13, no. 10, pp. S185–S185, 2018.
28. Sordella, R., D. W. Bell, D. A. Haber, and J. Settleman, "Gefitinib-sensitizing egfr mutations in lung cancer activate anti-apoptotic pathways," *Science*, Vol. 305, no. 5687, pp. 1163–7, 2004.
29. Riely, G. J., W. Pao, D. Pham, A. R. Li, N. Rizvi, E. S. Venkatraman, M. F. Zakowski, M. G. Kris, M. Ladanyi, and V. A. Miller, "Clinical course of patients with non-small cell lung cancer and epidermal growth factor receptor exon 19 and exon 21 mutations treated with gefitinib or erlotinib," *Clin Cancer Res*, Vol. 12, no. 3 Pt 1, pp. 839–44, 2006.
30. Solomon, B. J., D. W. Kim, Y. L. Wu, K. Nakagawa, T. Mekhail, E. Felip, F. Cappuzzo, J. Paolini, T. Usari, Y. Y. Tang, K. D. Wilner, F. Blackhall, and T. S. Mok, "Final overall survival analysis from a study comparing first-line crizotinib versus chemotherapy in alk-mutation-positive non-small-cell lung cancer," *Journal of Clinical Oncology*, Vol. 36, no. 22, pp. 2251–+, 2018.
31. Teng, F., X. Meng, L. Kong, and J. Yu, "Progress and challenges of predictive biomarkers of anti pd-1/pd-l1 immunotherapy: A systematic review," *Cancer Lett*, Vol. 414, pp. 166–173, 2018.
32. Mirsadraee, S., D. Oswal, Y. Alizadeh, A. Caulo, and J. van Beek, E., "The 7th lung cancer tnm classification and staging system: Review of the changes and implications," *World J Radiol*, Vol. 4, no. 4, pp. 128–34, 2012.
33. Birim, O., A. P. Kappetein, R. J. van Klaveren, and A. J. Bogers, "Prognostic factors in non-small cell lung cancer surgery," *Eur J Surg Oncol*, Vol. 32, no. 1, pp. 12–23, 2006.
34. Kachroo, S., L. Tong, M. R. Spitz, Y. Xing, K. Merriman, D. K. Zhu, J. Fueger, C. I. Amos, and C. J. Etzel, "Trends in prevalence of prognostic factors and survival in lung cancer patients from 1985 to 2004 at a tertiary care center," *Cancer Detect Prev*, Vol. 32, no. 2, pp. 101–8, 2008.
35. Dela Cruz, C. S., L. T. Tanoue, and R. A. Matthay, "Lung cancer: epidemiology, etiology, and prevention," *Clin Chest Med*, Vol. 32, no. 4, pp. 605–44, 2011.
36. Pao, W., and N. Girard, "New driver mutations in non-small-cell lung cancer," *Lancet Oncol*, Vol. 12, no. 2, pp. 175–80, 2011.
37. Schabath, M. B., Z. J. Thompson, and J. E. Gray, "Temporal trends in demographics and overall survival of non-small-cell lung cancer patients at moffitt cancer center from 1986 to 2008," *Cancer Control*, Vol. 21, no. 1, pp. 51–6, 2014.
38. Gatenby, R. A., O. Grove, and R. J. Gillies, "Quantitative imaging in cancer evolution and ecology," *Radiology*, Vol. 269, no. 1, pp. 8–15, 2013.
39. Yip, S. S., and H. J. Aerts, "Applications and limitations of radiomics," *Phys Med Biol*, Vol. 61, no. 13, pp. R150–66, 2016.

40. Kalpathy-Cramer, J., A. Mamomov, B. Zhao, L. Lu, D. Cherezov, S. Napel, S. Echeagaray, D. Rubin, M. McNitt-Gray, P. Lo, J. C. Sieren, J. Uthoff, S. K. Dilger, B. Driscoll, I. Yeung, L. Hadjiiski, K. Cha, Y. Balagurunathan, R. Gillies, and D. Goldgof, "Radiomics of lung nodules: A multi-institutional study of robustness and agreement of quantitative imaging features," *Tomography*, Vol. 2, no. 4, pp. 430–437, 2016.
41. Hadjiiski, L., B. Sahiner, H. P. Chan, N. Petrick, and M. Helvie, "Classification of malignant and benign masses based on hybrid art2lda approach," *IEEE Trans Med Imaging*, Vol. 18, no. 12, pp. 1178–87, 1999.
42. Wei, L. Y., Y. Y. Yang, R. M. Nishikawa, and Y. L. Jiang, "A study on several machine-learning methods for classification of malignant and benign clustered microcalcifications," *Ieee Transactions on Medical Imaging*, Vol. 24, no. 3, pp. 371–380, 2005.
43. Hawkins, S., H. Wang, Y. Liu, A. Garcia, O. Stringfield, H. Krewer, Q. Li, D. Cherezov, R. A. Gatenby, Y. Balagurunathan, D. Goldgof, M. B. Schabath, L. Hall, and R. J. Gillies, "Predicting malignant nodules from screening ct scans," *J Thorac Oncol*, Vol. 11, no. 12, pp. 2120–2128, 2016.
44. Peikert, T., F. Duan, S. Rajagopalan, R. A. Karwoski, R. Clay, R. A. Robb, Z. Qin, J. Sicks, B. J. Bartholmai, and F. Maldonado, "Novel high-resolution computed tomography-based radiomic classifier for screen-identified pulmonary nodules in the national lung screening trial," *PLoS One*, Vol. 13, no. 5, p. e0196910, 2018.
45. Huang, P., S. Park, R. Yan, J. Lee, L. C. Chu, C. T. Lin, A. Hussien, J. Rathmell, B. Thomas, C. Chen, R. Hales, D. S. Ettinger, M. Brock, P. Hu, E. K. Fishman, E. Gabrielson, and S. Lam, "Added value of computer-aided ct image features for early lung cancer diagnosis with small pulmonary nodules: A matched case-control study," *Radiology*, Vol. 286, no. 1, pp. 286–295, 2018.
46. Cherezov, D., S. H. Hawkins, D. B. Goldgof, L. O. Hall, Y. Liu, Q. Li, Y. Balagurunathan, R. J. Gillies, and M. B. Schabath, "Delta radiomic features improve prediction for lung cancer incidence: A nested case-control analysis of the national lung screening trial," *Cancer Med*, Vol. 7, no. 12, pp. 6340–6356, 2018.
47. Chae, H. D., C. M. Park, S. J. Park, S. M. Lee, K. G. Kim, and J. M. Goo, "Computerized texture analysis of persistent part-solid ground-glass nodules: differentiation of preinvasive lesions from invasive pulmonary adenocarcinomas," *Radiology*, Vol. 273, no. 1, pp. 285–93, 2014.
48. Liu, Y., Y. Balagurunathan, T. Atwater, S. Antic, Q. Li, R. C. Walker, G. T. Smith, P. P. Massion, M. B. Schabath, and R. J. Gillies, "Radiological image traits predictive of cancer status in pulmonary nodules," *Clin Cancer Res*, Vol. 23, no. 6, pp. 1442–1449, 2017.
49. Dhara, A. K., S. Mukhopadhyay, A. Dutta, M. Garg, and N. Khandelwal, "A combination of shape and texture features for classification of pulmonary nodules in lung ct images," *J Digit Imaging*, Vol. 29, no. 4, pp. 466–75, 2016.
50. Aerts, H. J., E. R. Velazquez, R. T. Leijenaar, C. Parmar, P. Grossmann, S. Carvalho, J. Bussink, R. Monshouwer, B. Haibe-Kains, D. Rietveld, F. Hoebbers, 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,” *Nat Commun*, Vol. 5, p. 4006, 2014.
51. Tunali, I., O. Stringfield, A. Guvenis, H. Wang, Y. Liu, Y. Balagurunathan, P. Lambin, R. J. Gillies, and M. B. Schabath, “Radial gradient and radial deviation radiomic features from pre-surgical ct scans are associated with survival among lung adenocarcinoma patients,” *Oncotarget*, Vol. 8, no. 56, pp. 96013–96026, 2017.
 52. Coroller, T. P., P. Grossmann, Y. Hou, E. Rios Velazquez, R. T. Leijenaar, G. Hermann, P. Lambin, B. Haibe-Kains, R. H. Mak, and H. J. Aerts, “Ct-based radiomic signature predicts distant metastasis in lung adenocarcinoma,” *Radiother Oncol*, Vol. 114, no. 3, pp. 345–50, 2015.
 53. Wu, J., T. Aguilera, D. Shultz, M. Gudur, D. L. Rubin, J. Loo, B. W., M. Diehn, and R. Li, “Early-stage non-small cell lung cancer: Quantitative imaging characteristics of (18)f fluorodeoxyglucose pet/ct allow prediction of distant metastasis,” *Radiology*, Vol. 281, no. 1, pp. 270–8, 2016.
 54. Huang, Y. Q., Z. Y. Liu, L. He, X. Chen, D. Pan, Z. L. Ma, C. S. Liang, J. Tian, and C. H. 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, no. 3, pp. 947–957, 2016.
 55. Huynh, E., T. P. Coroller, V. Narayan, V. Agrawal, J. Romano, I. Franco, C. Parmar, Y. Hou, R. H. Mak, and H. J. W. L. Aerts, “Associations of radiomic data extracted from static and respiratory-gated ct scans with disease recurrence in lung cancer patients treated with sbrt,” *Plos One*, Vol. 12, no. 1, 2017.
 56. Li, Q., J. Kim, Y. Balagurunathan, Y. Liu, K. Latifi, O. Stringfield, A. Garcia, E. G. Moros, T. J. Dilling, M. B. Schabath, Z. Ye, and R. J. Gillies, “Imaging features from pretreatment ct scans are associated with clinical outcomes in nonsmall-cell lung cancer patients treated with stereotactic body radiotherapy,” *Med Phys*, Vol. 44, no. 8, pp. 4341–4349, 2017.
 57. Oikonomou, A., F. Khalvati, P. N. Tyrrell, M. A. Haider, U. Tarique, L. Jimenez-Juan, M. C. Tjong, I. Poon, A. Eilaghi, L. Ehrlich, and P. Cheung, “Radiomics analysis at pet/ct contributes to prognosis of recurrence and survival in lung cancer treated with stereotactic body radiotherapy,” *Scientific Reports*, Vol. 8, 2018.
 58. Win, T., K. A. Miles, S. M. Janes, B. Ganeshan, M. Shastry, R. Endozo, M. Meagher, R. I. Shortman, S. Wan, I. Kayani, P. J. Ell, and A. M. Groves, “Tumor heterogeneity and permeability as measured on the ct component of pet/ct predict survival in patients with non-small cell lung cancer,” *Clinical Cancer Research*, Vol. 19, no. 13, pp. 3591–3599, 2013.
 59. She, Y. L., L. Zhang, H. Y. Zhu, C. Y. Dai, D. Xie, H. K. Xie, W. Zhang, L. L. Zhao, L. L. Zou, K. Fei, X. W. 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*, Vol. 28, no. 12, pp. 5121–5128, 2018.
 60. Coroller, T. P., 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*, Vol. 12, no. 3, pp. 467–476, 2017.

61. Yu, W., C. Tang, B. P. Hobbs, X. Li, E. J. Koay, I. I. Wistuba, B. Sepesi, C. Behrens, J. R. Canales, E. R. P. Cuentas, J. J. Erasmus, L. E. Court, and J. Y. Chang, "Development and validation of a predictive radiomics model for clinical outcomes in stage i non-small cell lung cancer," *International Journal of Radiation Oncology Biology Physics*, Vol. 102, no. 4, pp. 1090–1097, 2018.
62. Mattonen, S. A., D. A. Palma, C. Johnson, A. V. Louie, M. Landis, G. Rodrigues, I. Chan, R. Etemad-Rezai, T. P. Yeung, S. Senan, and A. D. Ward, "Detection of local cancer recurrence after stereotactic ablative radiation therapy for lung cancer: Physician performance versus radiomic assessment," *Int J Radiat Oncol Biol Phys*, Vol. 94, no. 5, pp. 1121–8, 2016.
63. Khorrami, M., 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, no. 2, 2019.
64. Fave, X., L. F. Zhang, J. Z. Yang, D. Mackin, P. Balter, D. Gomez, D. Followill, A. K. Jones, F. Stingo, Z. X. Liao, R. Mohan, and L. Court, "Delta-radiomics features for the prediction of patient outcomes in non-small cell lung cancer," *Scientific Reports*, Vol. 7, 2017.
65. Tunali, I., 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*, Vol. 129, pp. 75–79, 2019.
66. Trebeschi, S., S. G. Drago, N. J. Birkbak, I. Kurilova, A. M. Calin, A. D. Pizzi, F. Lalezari, D. M. J. Lambregts, M. Rohaan, C. Parmar, K. J. Hartemink, C. Swanton, J. Haanen, C. U. Blank, E. F. Smit, R. G. H. Beets-Tan, and H. Aerts, "Predicting response to cancer immunotherapy using non-invasive radiomic biomarkers," *Ann Oncol*, 2019.
67. Wu, W., C. Parmar, P. Grossmann, J. Quackenbush, P. Lambin, J. Bussink, R. Mak, and H. J. Aerts, "Exploratory study to identify radiomics classifiers for lung cancer histology," *Front Oncol*, Vol. 6, p. 71, 2016.
68. Rios Velazquez, E., C. Parmar, Y. Liu, T. P. Coroller, G. Cruz, O. Stringfield, Z. Ye, M. Makrigiorgos, F. Fennessy, R. H. Mak, R. Gillies, J. Quackenbush, and H. Aerts, "Somatic mutations drive distinct imaging phenotypes in lung cancer," *Cancer Res*, Vol. 77, no. 14, pp. 3922–3930, 2017.
69. Gevaert, O., S. Echegaray, A. Khuong, C. D. Hoang, J. B. Shrager, K. C. Jensen, G. J. Berry, H. H. Guo, C. Lau, S. K. Plevritis, D. L. Rubin, S. Napel, and A. N. Leung, "Predictive radiogenomics modeling of egfr mutation status in lung cancer," *Sci Rep*, Vol. 7, p. 41674, 2017.
70. Liu, Y., J. Kim, Y. Balagurunathan, Q. Li, A. L. Garcia, O. Stringfield, Z. Ye, and R. J. Gillies, "Radiomic features are associated with egfr mutation status in lung adenocarcinomas," *Clin Lung Cancer*, Vol. 17, no. 5, pp. 441–448 e6, 2016.
71. Weiss, G. J., B. Ganeshan, K. A. Miles, D. H. Campbell, P. Y. Cheung, S. Frank, and R. L. Korn, "Noninvasive image texture analysis differentiates k-ras mutation from pan-wildtype nslc and is prognostic," *PLoS One*, Vol. 9, no. 7, p. e100244, 2014.

72. Yamamoto, S., R. L. Korn, R. Oklu, C. Migdal, M. B. Gotway, G. J. Weiss, A. J. Iafrate, D. W. Kim, and M. D. Kuo, "Alk molecular phenotype in non-small cell lung cancer: Ct radiogenomic characterization," *Radiology*, Vol. 272, no. 2, pp. 568–76, 2014.
73. Yoon, H. J., I. Sohn, J. H. Cho, H. Y. Lee, J. H. Kim, Y. L. Choi, H. Kim, G. Lee, K. S. Lee, and J. Kim, "Decoding tumor phenotypes for alk, ros1, and ret fusions in lung adenocarcinoma using a radiomics approach," *Medicine (Baltimore)*, Vol. 94, no. 41, p. e1753, 2015.
74. Zhou, M., A. Leung, S. Echegaray, A. Gentles, J. B. Shrager, K. C. Jensen, G. J. Berry, S. K. Plevritis, D. L. Rubin, S. Napel, and O. Gevaert, "Non-small cell lung cancer radiogenomics map identifies relationships between molecular and imaging phenotypes with prognostic implications," *Radiology*, Vol. 286, no. 1, pp. 307–315, 2018.
75. Shafiq-Ul-Hassan, M., G. G. Zhang, K. Latifi, G. Ullah, D. C. Hunt, Y. Balagurunathan, M. A. Abdalah, M. B. Schabath, D. G. Goldgof, D. Mackin, L. E. Court, R. J. Gillies, and E. G. Moros, "Intrinsic dependencies of ct radiomic features on voxel size and number of gray levels," *Med Phys*, Vol. 44, no. 3, pp. 1050–1062, 2017.
76. Zwanenburg, A., S. Leger, M. Vallières, and S. Löck, "Image biomarker standardisation initiative," <https://arxiv.org/abs/1612.07003>, 2018.
77. Balagurunathan, Y., Y. Gu, H. Wang, V. Kumar, O. Grove, S. Hawkins, J. Kim, D. B. Goldgof, L. O. Hall, R. A. Gatenby, and R. J. Gillies, "Reproducibility and prognosis of quantitative features extracted from ct images," *Transl Oncol*, Vol. 7, no. 1, pp. 72–87, 2014.
78. Zhang, H., S. Tan, W. G. Chen, S. Kligerman, G. Kim, W. D. D'Souza, M. Suntharalingam, and W. Lu, "Modeling pathologic response of esophageal cancer to chemoradiation therapy using spatial-temporal f-18-fdg pet features, clinical parameters, and demographics," *International Journal of Radiation Oncology Biology Physics*, Vol. 88, no. 1, pp. 195–203, 2014.
79. Bundschuh, R. A., J. Dinges, L. Neumann, M. Seyfried, N. Zsoter, L. Papp, R. Rosenberg, K. Becker, S. T. Astner, M. Henninger, K. Herrmann, S. I. Ziegler, M. Schwaiger, and M. Essler, "Textural parameters of tumor heterogeneity in f-18-fdg pet/ct for therapy response assessment and prognosis in patients with locally advanced rectal cancer," *Journal of Nuclear Medicine*, Vol. 55, no. 6, pp. 891–897, 2014.
80. Tan, S., S. Kligerman, W. G. Chen, M. Lu, G. Kim, S. Feigenberg, W. D. D'Souza, M. Suntharalingam, and W. Lu, "Spatial-temporal [f-18]fdg-pet features for predicting pathologic response of esophageal cancer to neoadjuvant chemoradiation therapy," *International Journal of Radiation Oncology Biology Physics*, Vol. 85, no. 5, pp. 1375–1382, 2013.
81. Chalkidou, A., M. J. O'Doherty, and P. K. Marsden, "False discovery rates in pet and ct studies with texture features: A systematic review," *PLoS One*, Vol. 10, no. 5, p. e0124165, 2015.
82. Bland, J. M., and D. G. Altman, "Multiple significance tests: the bonferroni method," *BMJ*, Vol. 310, no. 6973, p. 170, 1995.
83. Holm, M., "A simple sequentially rejective multiple test procedure," *Scandinavian Journal of Statistics*, Vol. 6, pp. 65–70, 1979.

84. Benjamini, Y., and Y. Hochberg, "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *Journal of the Royal Statistical Society Series B-Statistical Methodology*, Vol. 57, pp. 289–300, 1995.
85. Lambin, P., R. T. H. Leijenaar, and T. M. Deist et al, "Radiomics: the bridge between medical imaging and personalized medicine," *Nature Reviews Clinical Oncology*, Vol. 14, no. 12, pp. 749–762, 2017.
86. Park, J. E., D. Kim, H. S. Kim, S. Y. Park, J. Y. Kim, S. J. Cho, J. H. Shin, and J. H. Kim, "Quality of science and reporting of radiomics in oncologic studies: room for improvement according to radiomics quality score and tripod statement," *Eur Radiol*, 2019.
87. Paul, R., S. H. Hawkins, M. B. Schabath, R. J. Gillies, L. O. Hall, and D. B. Goldgof, "Predicting malignant nodules by fusing deep features with classical radiomics features," *J Med Imaging (Bellingham)*, Vol. 5, no. 1, p. 011021, 2018.
88. Kazuhiro, K., R. A. Werner, F. Toriumi, M. S. Javadi, M. G. Pomper, L. B. Solnes, F. Verde, T. Higuchi, and S. P. Rowe, "Generative adversarial networks for the creation of realistic artificial brain magnetic resonance images," *Tomography*, Vol. 4, no. 4, pp. 159–163, 2018.
89. Paul, R., S. H. Hawkins, Y. Balagurunathan, M. B. Schabath, R. J. Gillies, L. O. Hall, and D. B. Goldgof, "Deep feature transfer learning in combination with traditional features predicts survival among patients with lung adenocarcinoma," *Tomography*, Vol. 2, no. 4, pp. 388–395, 2016.
90. Nishio, M., O. Sugiyama, M. Yakami, S. Ueno, T. Kubo, T. Kuroda, and K. Togashi, "Computer-aided diagnosis of lung nodule classification between benign nodule, primary lung cancer, and metastatic lung cancer at different image size using deep convolutional neural network with transfer learning," *Plos One*, Vol. 13, no. 7, 2018.
91. Wang, C., N. Tyagi, A. Rimner, Y. C. Hu, H. Veeraraghavan, G. Li, M. Hunt, G. Mageras, and P. Zhang, "Segmenting lung tumors on longitudinal imaging studies via a patient-specific adaptive convolutional neural network," *Radiother Oncol*, Vol. 131, pp. 101–107, 2019.
92. Liu, Y., Y. Balagurunathan, T. Atwater, S. Antic, Q. Li, R. C. Walker, G. T. Smith, P. P. Massion, M. B. Schabath, and R. J. Gillies, "Radiological image traits predictive of cancer status in pulmonary nodules," *Clin Cancer Res*, 2016.
93. Wang, H., M. B. Schabath, Y. Liu, O. Stringfield, Y. Balagurunathan, J. J. Heine, S. A. Eschrich, Z. Ye, and R. J. Gillies, "Association between computed tomographic features and kirsten rat sarcoma viral oncogene mutations in patients with stage i lung adenocarcinoma and their prognostic value," *Clin Lung Cancer*, Vol. 17, no. 4, pp. 271–8, 2016.
94. Hardie, R. C., S. K. Rogers, T. Wilson, and A. Rogers, "Performance analysis of a new computer aided detection system for identifying lung nodules on chest radiographs," *Med Image Anal*, Vol. 12, no. 3, pp. 240–58, 2008.
95. Messay, T., R. C. Hardie, and S. K. Rogers, "A new computationally efficient cad system for pulmonary nodule detection in ct imagery," *Med Image Anal*, Vol. 14, no. 3, pp. 390–406, 2010.

96. Messay, T., R. C. Hardie, and T. R. Tuinstra, "Segmentation of pulmonary nodules in computed tomography using a regression neural network approach and its application to the lung image database consortium and image database resource initiative dataset," *Med Image Anal*, Vol. 22, no. 1, pp. 48–62, 2015.
97. Gu, Y., V. Kumar, L. O. Hall, D. B. Goldgof, C. Y. Li, R. Korn, C. Bendtsen, E. R. Velazquez, A. Dekker, H. Aerts, P. Lambin, X. Li, J. Tian, R. A. Gatenby, and R. J. Gillies, "Automated delineation of lung tumors from ct images using a single click ensemble segmentation approach," *Pattern Recognit*, Vol. 46, no. 3, pp. 692–702, 2013.
98. Hansell, D. M., A. A. Bankier, H. MacMahon, T. C. McLoud, N. L. Muller, and J. Remy, "Fleischner society: glossary of terms for thoracic imaging," *Radiology*, Vol. 246, no. 3, pp. 697–722, 2008.
99. Zwirwich, C. V., S. Vedal, R. R. Miller, and N. L. Muller, "Solitary pulmonary nodule: high-resolution ct and radiologic-pathologic correlation," *Radiology*, Vol. 179, no. 2, pp. 469–76, 1991.
100. Onn, A., D. H. Choe, R. S. Herbst, A. M. Correa, R. F. Munden, M. T. Truong, A. A. Vaporciyan, T. Isobe, M. Z. Gilcrease, and E. M. Marom, "Tumor cavitation in stage i non-small cell lung cancer: Epidermal growth factor receptor expression and prediction of poor outcome 1," *Radiology*, Vol. 237, no. 1, pp. 342–347, 2005.
101. Gevaert, O., J. J. Xu, C. D. Hoang, A. N. Leung, Y. Xu, A. Quon, D. L. Rubin, S. Napel, and S. K. Plevritis, "Non-small cell lung cancer: Identifying prognostic imaging biomarkers by leveraging public gene expression microarray data-methods and preliminary results," *Radiology*, Vol. 264, no. 2, pp. 387–396, 2012.
102. Shultz, D. B., N. Trakul, J. A. Abelson, J. D. Murphy, P. G. Maxim, Q. T. Le, J. Loo, B. W., and M. Diehn, "Imaging features associated with disease progression after stereotactic ablative radiotherapy for stage i non-small-cell lung cancer," *Clin Lung Cancer*, Vol. 15, no. 4, pp. 294–301 e3, 2014.
103. Kumar, V., Y. Gu, S. Basu, A. Berglund, S. A. Eschrich, M. B. Schabath, K. Forster, H. J. Aerts, A. Dekker, D. Fenstermacher, D. B. Goldgof, L. O. Hall, P. Lambin, Y. Balagurunathan, R. A. Gatenby, and R. J. Gillies, "Radiomics: the process and the challenges," *Magn Reson Imaging*, Vol. 30, no. 9, pp. 1234–48, 2012.
104. Champiat, S., L. Derle, S. Ammari, C. Massard, A. Hollebecque, S. Postel-Vinay, N. Chaput, A. Eggermont, A. Marabelle, J. C. Soria, and C. Ferte, "Hyperprogressive disease is a new pattern of progression in cancer patients treated by anti-pd-1/pd-l1," *Clinical Cancer Research*, Vol. 23, no. 8, pp. 1920–1928, 2017.
105. Lambin, P., E. Rios-Velazquez, R. Leijenaar, S. Carvalho, R. G. van Stiphout, P. Granton, C. M. Zegers, R. Gillies, R. Boellard, A. Dekker, and H. J. Aerts, "Radiomics: extracting more information from medical images using advanced feature analysis," *Eur J Cancer*, Vol. 48, no. 4, pp. 441–6, 2012.
106. Arkenau, H. T., D. Olmos, J. E. Ang, J. de Bono, I. Judson, and S. Kaye, "Clinical outcome and prognostic factors for patients treated within the context of a phase i study: the royal marsden hospital experience," *Br J Cancer*, Vol. 98, no. 6, pp. 1029–33, 2008.

107. Wheler, J., A. M. Tsimberidou, D. Hong, A. Naing, G. Falchook, S. Piha-Paul, S. Fu, S. Moulder, B. Stephen, S. Wen, and R. Kurzrock, "Survival of 1,181 patients in a phase i clinic: the md anderson clinical center for targeted therapy experience," *Clin Cancer Res*, Vol. 18, no. 10, pp. 2922–9, 2012.
108. Kato, S., A. Goodman, V. Walavalkar, D. A. Barkauskas, A. Sharabi, and R. Kurzrock, "Hyperprogressors after immunotherapy: Analysis of genomic alterations associated with accelerated growth rate," *Clin Cancer Res*, Vol. 23, no. 15, pp. 4242–4250, 2017.
109. Saada-Bouزيد, E., C. Defaucheux, A. Karabajakian, V. Palomar 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," *Ann Oncol*, 2017.
110. Breiman, L., "Classification and regression trees," *New York, NY: Kluwer Academic Publishers*, 1984.
111. Chawla, N. V., K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: Synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, Vol. 16, pp. 321–357, 2002.
112. Chen, D. S., and I. Mellman, "Elements of cancer immunity and the cancer-immune set point," *Nature*, Vol. 541, no. 7637, pp. 321–330, 2017.
113. Bagley, S. J., S. Kothari, and C. Aggarwal et al, "Pretreatment neutrophil-to-lymphocyte ratio as a marker of outcomes in nivolumab-treated patients with advanced non-small-cell lung cancer," *Lung Cancer*, Vol. 106, pp. 1–7, 2017.
114. Reinstein, Z. Z., S. Pamarthy, V. Sagar, R. Costa, S. A. Abdulkadir, F. J. Giles, and B. A. Carneiro, "Overcoming immunosuppression in bone metastases," *Crit Rev Oncol Hematol*, Vol. 117, pp. 114–127, 2017.
115. Mezquita, L., E. Auclin, and R. Ferrara et al, "Association of the lung immune prognostic index with immune checkpoint inhibitor outcomes in patients with advanced non-small cell lung cancer," *JAMA Oncol*, 2018.
116. Whiteside, T. L., "The tumor microenvironment and its role in promoting tumor growth," *Oncogene*, Vol. 27, no. 45, pp. 5904–12, 2008.
117. Laws, K., "Texture image segmentation," *University of South California, Los Angeles, CA*, 1980.
118. Garrido-Laguna, I., F. Janku, C. Vaklavas, G. S. Falchook, S. Fu, D. S. Hong, A. Naing, A. M. Tsimberidou, S. Wen, and R. Kurzrock, "Validation of the royal marsden hospital prognostic score in patients treated in the phase i clinical trials program at the md anderson cancer center," *Cancer*, Vol. 118, no. 5, pp. 1422–8, 2012.
119. Amadasun, M., and R. King, "Textural features corresponding to textural properties," *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 19, p. 1264–1274, 1989.
120. Grossmann, P., 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, 2017.

121. Shien, K., V. A. Papadimitrakopoulou, and I. Wistuba, "Predictive biomarkers of response to pd-1/pd-l1 immune checkpoint inhibitors in non-small cell lung cancer," *Lung Cancer*, Vol. 99, pp. 79–87, 2016.
122. Grigg, C., and N. A. Rizvi, "Pd-l1 biomarker testing for non-small cell lung cancer: truth or fiction?," *J Immunother Cancer*, Vol. 4, p. 48, 2016.
123. Parmar, C., R. T. Leijenaar, P. Grossmann, E. Rios Velazquez, J. Bussink, D. Rietveld, M. M. Rietbergen, B. Haibe-Kains, P. Lambin, and H. J. Aerts, "Radiomic feature clusters and prognostic signatures specific for lung and head neck cancer," *Sci Rep*, Vol. 5, p. 11044, 2015.
124. Bremnes, R. M., C. Camps, and R. Sirera, "Angiogenesis in non-small cell lung cancer: the prognostic impact of neoangiogenesis and the cytokines vegf and bfgf in tumours and blood," *Lung Cancer*, Vol. 51, no. 2, pp. 143–58, 2006.
125. Christiansen, A., and M. Detmar, "Lymphangiogenesis and cancer," *Genes Cancer*, Vol. 2, no. 12, pp. 1146–58, 2011.
126. Grivennikov, S. I., F. R. Greten, and M. Karin, "Immunity, inflammation, and cancer," *Cell*, Vol. 140, no. 6, pp. 883–99, 2010.
127. Pages, F., J. Galon, M. C. Dieu-Nosjean, E. Tartour, C. Sautes-Fridman, and W. H. Fridman, "Immune infiltration in human tumors: a prognostic factor that should not be ignored," *Oncogene*, Vol. 29, no. 8, pp. 1093–102, 2010.
128. Way, T. W., L. M. Hadjiiski, B. Sahiner, H. P. Chan, P. N. Cascade, E. A. Kazerooni, N. Bogot, and C. Zhou, "Computer-aided diagnosis of pulmonary nodules on ct scans: segmentation and classification using 3d active contours," *Med Phys*, Vol. 33, no. 7, pp. 2323–37, 2006.
129. Way, T. W., B. Sahiner, H. P. Chan, L. Hadjiiski, P. N. Cascade, A. Chughtai, N. Bogot, and E. Kazerooni, "Computer-aided diagnosis of pulmonary nodules on ct scans: improvement of classification performance with nodule surface features," *Med Phys*, Vol. 36, no. 7, pp. 3086–98, 2009.
130. Shah, S. K., M. F. McNitt-Gray, S. R. Rogers, J. G. Goldin, R. D. Suh, J. W. Sayre, I. Petkovska, H. J. Kim, and D. R. Aberle, "Computer aided characterization of the solitary pulmonary nodule using volumetric and contrast enhancement features," *Academic Radiology*, Vol. 12, no. 10, pp. 1310–1319, 2005.
131. Sun, R., E. J. Limkin, M. Vakalopoulou, L. Dercle, S. Champiat, S. R. Han, L. Verlingue, D. Brandao, A. Lancia, S. Ammari, A. Hollebecque, J. Y. Scoazec, A. Marabelle, C. Massard, J. C. Soria, C. Robert, N. Paragios, E. Deutsch, and C. Ferte, "A radiomics approach to assess tumour-infiltrating cd8 cells and response to anti-pd-1 or anti-pd-l1 immunotherapy: an imaging biomarker, retrospective multicohort study," *Lancet Oncol*, Vol. 19, no. 9, pp. 1180–1191, 2018.
132. Braman, N. M., M. Etesami, P. Prasanna, C. Dubchuk, H. Gilmore, P. Tiwari, D. Plecha, and A. Madabhushi, "Intratumoral and peritumoral radiomics for the pre-treatment prediction of pathological complete response to neoadjuvant chemotherapy based on breast dce-mri," *Breast Cancer Res*, Vol. 19, no. 1, p. 57, 2017.

133. Prasanna, P., J. Patel, S. Partovi, A. Madabhushi, and P. Tiwari, "Radiomic features from the peritumoral brain parenchyma on treatment-naive multi-parametric mr imaging predict long versus short-term survival in glioblastoma multiforme: Preliminary findings," *Eur Radiol*, Vol. 27, no. 10, pp. 4188–4197, 2017.
134. Dou, T. H., T. P. Coroller, J. J. M. van Griethuysen, R. H. Mak, and H. Aerts, "Peritumoral radiomics features predict distant metastasis in locally advanced nsclc," *PLoS One*, Vol. 13, no. 11, p. e0206108, 2018.
135. Mattonen, S., G. Davidzon, S. Bakr, S. Echegaray, A. N. C. Leung, M. Vasanawala, G. Horng, S. Napel, and V. S. Nair, "Utility of 18f-fdg positron emission tomography (pet) tumor and penumbra texture features for recurrence prediction in non-small cell lung cancer," *Tomography*, Vol. IN PRESS, November 2018.
136. Balagurunathan, Y., V. Kumar, Y. Gu, J. Kim, H. Wang, Y. Liu, D. B. Goldgof, L. O. Hall, R. Korn, B. Zhao, L. H. Schwartz, S. Basu, S. Eschrich, R. A. Gatenby, and R. J. Gillies, "Test-retest reproducibility analysis of lung ct image features," *J Digit Imaging*, Vol. 27, no. 6, pp. 805–23, 2014.
137. Balagurunathan, Y., A. Beers, J. Kalpathy-Cramer, M. McNitt-Gray, L. Hadjiiski, B. Zhao, J. Zhu, H. Yang, S. S. F. Yip, H. Aerts, S. Napel, D. Cherezov, K. Cha, H. P. Chan, C. Flores, A. Garcia, R. Gillies, and D. Goldgof, "Semi-automated pulmonary nodule interval segmentation using the nlst data," *Med Phys*, Vol. 45, no. 3, pp. 1093–1107, 2018.
138. Schabath, M. B., E. A. Welsh, and W. J. Fulp et al, "Differential association of stk11 and tp53 with kras mutation-associated gene expression, proliferation and immune surveillance in lung adenocarcinoma," *Oncogene*, Vol. 35, no. 24, pp. 3209–16, 2016.
139. Tian, D., Y. Pei, Q. Zheng, J. Zhang, S. Li, X. Wang, D. Lin, and Y. Yang, "Effect of visceral pleural invasion on the prognosis of patients with lymph node negative non-small cell lung cancer," *Thorac Cancer*, Vol. 8, no. 2, pp. 97–105, 2017.
140. Borcoman, E., A. Nandikolla, G. Long, S. Goel, and C. Le Tourneau, "Patterns of response and progression to immunotherapy," *Am Soc Clin Oncol Educ Book*, no. 38, pp. 169–178, 2018.
141. Sharma, P., S. Hu-Lieskovan, J. A. Wargo, and A. Ribas, "Primary, adaptive, and acquired resistance to cancer immunotherapy," *Cell*, Vol. 168, no. 4, pp. 707–723, 2017.
142. Yarchoan, M., A. Hopkins, and E. M. Jaffee, "Tumor mutational burden and response rate to pd-1 inhibition," *N Engl J Med*, Vol. 377, no. 25, pp. 2500–2501, 2017.
143. Cristescu, R., R. Mogg, and M. Ayers et al, "Pan-tumor genomic biomarkers for pd-1 checkpoint blockade-based immunotherapy," *Science*, Vol. 362, no. 6411, 2018.
144. Hellmann, M. D., T. E. Ciuleanu, and A. Pluzanski et al, "Nivolumab plus ipilimumab in lung cancer with a high tumor mutational burden," *N Engl J Med*, Vol. 378, no. 22, pp. 2093–2104, 2018.
145. Hellmann, M. D., T. Nathanson, and H. Rizvi et al, "Genomic features of response to combination immunotherapy in patients with advanced non-small-cell lung cancer," *Cancer Cell*, Vol. 33, no. 5, pp. 843–852 e4, 2018.

146. Goodman, A. M., S. Kato, L. Bazhenova, S. P. Patel, G. M. Frampton, V. Miller, P. J. Stephens, G. A. Daniels, and R. Kurzrock, "Tumor mutational burden as an independent predictor of response to immunotherapy in diverse cancers," *Molecular Cancer Therapeutics*, Vol. 16, no. 11, pp. 2598–2608, 2017.
147. Gerlinger, M., A. J. Rowan, and S. Horswell et al, "Intratumor heterogeneity and branched evolution revealed by multiregion sequencing," *N Engl J Med*, Vol. 366, no. 10, pp. 883–892, 2012.
148. Rizzardi, A. E., A. T. Johnson, R. I. Vogel, S. E. Pambuccian, J. Henriksen, A. P. Skubitz, G. J. Metzger, and S. C. Schmechel, "Quantitative comparison of immunohistochemical staining measured by digital image analysis versus pathologist visual scoring," *Diagn Pathol*, Vol. 7, p. 42, 2012.
149. Liu, Y., H. Wang, Q. Li, M. J. McGettigan, Y. Balagurunathan, A. L. Garcia, Z. J. Thompson, J. J. Heine, Z. Ye, R. J. Gillies, and M. B. Schabath, "Radiologic features of small pulmonary nodules and lung cancer risk in the national lung screening trial: A nested case-control study," *Radiology*, Vol. 286, no. 1, pp. 298–306, 2018.
150. Traverso, A., L. Wee, A. Dekker, and R. Gillies, "Repeatability and reproducibility of radiomic features: A systematic review," *Int J Radiat Oncol Biol Phys*, 2018.
151. Harris, A. L., "Hypoxia—a key regulatory factor in tumour growth," *Nat Rev Cancer*, Vol. 2, no. 1, pp. 38–47, 2002.
152. Chan, D. A., and A. J. Giaccia, "Hypoxia, gene expression, and metastasis," *Cancer Metastasis Rev*, Vol. 26, no. 2, pp. 333–9, 2007.
153. Ilie, M., V. Hofman, J. Zangari, J. Chiche, J. Mouroux, N. M. Mazure, J. Pouyssegur, P. Brest, and P. Hofman, "Response of caix and caxii to in vitro re-oxygenation and clinical significance of the combined expression in nslc patients," *Lung Cancer*, Vol. 82, no. 1, pp. 16–23, 2013.
154. Pastorek, J., and S. Pastorekova, "Hypoxia-induced carbonic anhydrase ix as a target for cancer therapy: From biology to clinical use," *Seminars in Cancer Biology*, Vol. 31, pp. 52–64, 2015.
155. Muz, B., P. de la Puente, F. Azab, and A. K. Azab, "The role of hypoxia in cancer progression, angiogenesis, metastasis, and resistance to therapy," *Hypoxia (Auckl)*, Vol. 3, pp. 83–92, 2015.
156. Espinosa, E., J. Feliu, P. Zamora, M. Gonzalez Baron, J. J. Sanchez, A. Ordon ez, and J. Espinosa, "Serum albumin and other prognostic factors related to response and survival in patients with advanced non-small cell lung cancer," *Lung Cancer*, Vol. 12, no. 1-2, pp. 67–76, 1995.
157. Miura, K., K. Hamanaka, T. Koizumi, Y. Kitaguchi, Y. Terada, D. Nakamura, H. Kumeda, H. Agatsuma, A. Hyogotani, S. Kawakami, A. Yoshizawa, S. Asaka, and K. I. Ito, "Clinical significance of preoperative serum albumin level for prognosis in surgically resected patients with non-small cell lung cancer: Comparative study of normal lung, emphysema, and pulmonary fibrosis," *Lung Cancer*, Vol. 111, pp. 88–95, 2017.

158. Wheler, J., A. M. Tsimberidou, D. Hong, A. Naing, G. Falchook, S. Piha-Paul, S. Fu, S. Moulder, B. Stephen, S. Wen, and R. Kurzrock, "Survival of 1,181 patients in a phase i clinic: the md anderson clinical center for targeted therapy experience," *Clin Cancer Res*, Vol. 18, no. 10, pp. 2922–9, 2012.
159. Aerts, H. J. W. L., P. Grossmann, Y. Q. Tan, G. R. Oxnard, N. Rizvi, L. H. Schwartz, and B. S. Zhao, "Defining a radiomic response phenotype: A pilot study using targeted therapy in nslc (vol 6, 33860, 2016)," *Scientific Reports*, Vol. 7, 2017.
160. Jia, T. Y., J. F. Xiong, X. Y. Li, W. Yu, Z. Y. Xu, X. W. Cai, J. C. Ma, Y. C. Ren, R. Larsson, J. Zhang, J. Zhao, and X. L. Fu, "Identifying egfr mutations in lung adenocarcinoma by noninvasive imaging using radiomics features and random forest modeling," *Eur Radiol*, 2019.
161. Bibault, J. E., P. Giraud, M. Housset, C. Durdux, J. Taieb, A. Berger, R. Coriat, S. Chaussade, B. Dousset, B. Nordlinger, and A. Burgun, "Deep learning and radiomics predict complete response after neo-adjuvant chemoradiation for locally advanced rectal cancer (vol 8, 12611, 2018)," *Scientific Reports*, Vol. 8, 2018.
162. Innocenti, A., S. Pastorekova, J. Pastorek, A. Scozzafava, G. De Simone, and C. T. Supuran, "The proteoglycan region of the tumor-associated carbonic anhydrase isoform ix acts as anintrinsic buffer optimizing co2 hydration at acidic ph values characteristic of solid tumors," *Bioorg Med Chem Lett*, Vol. 19, no. 20, pp. 5825–8, 2009.