

Article

Impact of Interobserver Variability in Manual Segmentation of Non-Small Cell Lung Cancer (NSCLC) Applying Low-Rank Radiomic Representation on Computed Tomography

Michelle Hershman ^{1,*}, Bardia Yousefi ^{1,2,†}, Lacey Serletti ³, Maya Galperin-Aizenberg ¹, Leonid Roshkovan ¹, José Marcio Luna ^{1,2}, Jeffrey C. Thompson ⁴, Charu Aggarwal ⁵, Erica L. Carpenter ⁵, Despina Kontos ^{1,2} and Sharyn I. Katz ^{1,*}

- ¹ Department of Radiology, University of Pennsylvania, Philadelphia, PA 19104, USA; bardia.yousefi@gmail.com (B.Y.); maya.galperin-aizenberg@pennmedicine.upenn.edu (M.G.-A.); Leonid.roshkovan@pennmedicine.upenn.edu (L.R.); jose.luna@pennmedicine.upenn.edu (J.M.L.); despina.kontos@pennmedicine.upenn.edu (D.K.)
 - ² Center for Biomedical Image Computing and Analytics, University of Pennsylvania, Philadelphia, PA 19104, USA
 - ³ Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA; lacey.serletti@pennmedicine.upenn.edu
 - ⁴ Section of Interventional Pulmonology, Department of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA; jeffrey.thompson@pennmedicine.upenn.edu
 - ⁵ Division of Hematology and Oncology, Department of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA; charu.aggarwal@pennmedicine.upenn.edu (C.A.); erical@upenn.edu (E.L.C.)
- * Correspondence: Michelle.hershman@pennmedicine.upenn.edu (M.H.); Sharyn.katz@pennmedicine.upenn.edu (S.I.K.)
† Denotes co-first authors with equal contributions.

Citation: Hershman, M.; Yousefi, B.; Serletti, L.; Galperin-Aizenberg, M.; Roshkovan, L.; Luna, J.M.; Thompson, J.C.; Aggarwal, C.; Carpenter, E.L.; Kontos, D.; et al. Impact of Interobserver Variability in Manual Segmentation of Non-Small Cell Lung Cancer (NSCLC) Applying Low-Rank Radiomic Representation on Computed Tomography. *Cancers* **2021**, *13*, 5985. <https://doi.org/10.3390/cancers13235985>

Academic Editor: Mario Mascali

Received: 16 August 2021

Accepted: 25 November 2021

Published: 28 November 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Simple Summary: Discovery of predictive and prognostic radiomic features in cancer is currently of great interest to the radiologic and oncologic community. Tumor phenotypic and prognostic information can be obtained by extracting features on tumor segmentations, and it is typically imaging analysts, physician trainees, and attending physicians who provide these labeled datasets for analysis. The potential impact of level and type of specialty training on interobserver variability in manual segmentation of NSCLC was examined. Although there was some variability in segmentation between readers, the subsequently extracted radiomic features were overall well correlated. High fidelity radiomic feature extraction relies on accurate feature extraction from imaging that produce robust prognostic and predictive radiomic NSCLC biomarkers. This study concludes that this goal can be obtained using segmenters of different levels of training and clinical experience.

Abstract: This study tackles interobserver variability with respect to specialty training in manual segmentation of non-small cell lung cancer (NSCLC). Four readers included for segmentation are: a data scientist (BY), a medical student (LS), a radiology trainee (MH), and a specialty-trained radiologist (SK) for a total of 295 patients from two publicly available databases. Sørensen–Dice (SD) coefficients and low rank Pearson correlation coefficients (CC) of 429 radiomics were calculated to assess interobserver variability. Cox proportional hazard (CPH) models and Kaplan–Meier (KM) curves of overall survival (OS) prediction for each dataset were also generated. SD and CC for segmentations demonstrated high similarities, yielding, SD: 0.79 and CC: 0.92 (BY–SK), SD: 0.81 and CC: 0.83 (LS–SK), and SD: 0.84 and CC: 0.91 (MH–SK) in average for both databases, respectively. OS through the maximal CPH model for the two datasets yielded c-statistics of 0.7 (95% CI) and 0.69 (95% CI), while adding radiomic and clinical variables (sex, stage/morphological status, and histology) together. KM curves also showed significant discrimination between high- and low-risk patients (p -value < 0.005). This supports that readers' level of training and clinical experience may not significantly influence the ability to extract accurate radiomic features for NSCLC on CT. This potentially allows flexibility in the training required to produce robust prognostic imaging biomarkers for potential clinical translation.

Keywords: radiomics; interobserver variability; non-small cell lung cancer; computed tomography (CT)

1. Introduction

Lung cancer is the leading cause of cancer-related death in the United States [1]. Non-small cell lung cancer (NSCLC) represents the majority of primary lung cancers and carries a poor prognosis and low overall survival [2]. Computed tomography (CT) is a routinely used diagnostic imaging tool in clinical management in oncology due to the ability of CT to noninvasively provide anatomic information for detection, staging, and therapy response assessment. Over the past decade it has become evident that quantitative features are embedded in conventional medical imaging data, not appreciable to the human eye [3]. These radiomics features are a reflection of tissue architecture, heterogeneity, and pericellular environment and can be harnessed to construct tissue signatures that correlate with clinically relevant biomarkers, including tumor histologic subtype, mutational status, degree of infiltration with tumor infiltrating lymphocytes, as well as therapeutic endpoints such as overall survival [4–9]. These imaging “phenotypes” provide valuable data that may enhance personalization of medical care in oncology [10].

It is well known that repeatability and reproducibility of radiomic features on CT are sensitive to various image details such as image acquisition settings, processing, reconstruction algorithm, and specific software used for radiomic feature extraction [5,7,9,11–17]. Furthermore, certain radiomic features are more sensitive to these variations than others, with first order features, specifically entropy, consistently reported as being very stable while other texture features, such as coarseness and contrast, being the least reproducible [18].

Discovery of predictive and prognostic radiomic features in cancer is currently of great interest to the radiologic community; however, there is no reliable fully automated means of segmenting lung cancer. Tumor delineation and contouring are often performed by scientists with a range of training in anatomical imaging including imaging analysts, students, physician trainees, and attending physicians using either manual or semi-automated techniques. In addition to being time consuming, 3-dimensional manual and semi-automated contouring are subject to interobserver variability. This variability has been shown to be particularly challenging with segmented lesions when associated with ground glass components and postobstructive atelectasis [4]. In order to generate high fidelity phenotypic radiomic signatures, tumor segmentations must be reproducible across different readers [17]. Performing quality segmentations is an important task. Although the ability to anticipate tumor histology, mutational status, and therapeutic consequences are all ultimate goals of radiomics, interobserver variability between readers should be thoroughly investigated before subsequent feature analysis is tested, given that these segmentations form the basis of the analyses.

To our knowledge, no study has examined how both the level and type of specialty training in manual or semi-automated segmentations affects the subsequent extraction of radiomic features. Thus, our purpose in this study is to examine how the level of specialty training impacts interobserver variability in manual segmentation and radiomic feature extraction of NSCLC on CT.

2. Materials and Methods

The proposed approach presents a comparative assessment of interobserver variability in segmenting NSCLC tumors on chest CTs and its effect on subsequent extraction of radiomic features and survival analysis (see Figure 1 for study schema).

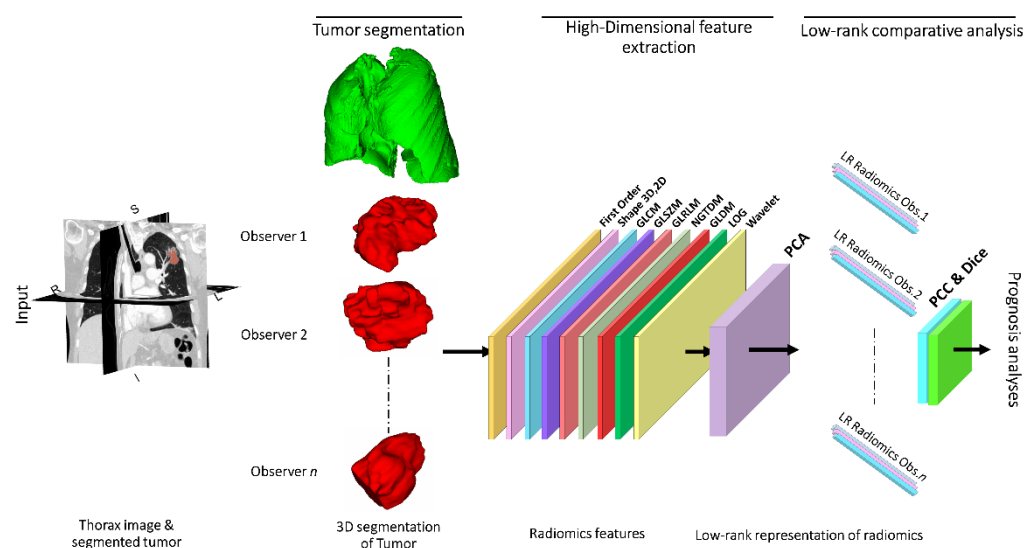


Figure 1. Workflow of the approach. The NSCLC tumor is segmented from the original CT images by four segmenters ($n = 4$) with different backgrounds, yielding radiomics features and tumor masks as inputs. Next, PCA categorizes features based on their maximum variance in radiomics. For every group, three principal components of feature sets are selected and used for correlative analysis and prediction of survival.

2.1. Patient Population and Study Data

This was a single-center study with segmentations performed at our institution between July 2018 and December 2019. The CT images included in this study had slice thicknesses between 1 and 5 mm, and both contrast and non-contrast enhanced studies were included. No pre-processing methods of the CT images were employed. Two publicly available datasets containing CT images from patients with NSCLC were analyzed. The NSCLC-Radiomics-Genomics-Lung3 (also known as Harvard) dataset (Table 1) [11,19,20] contains pre-treatment CT images from 89 patients with NSCLC and the NSCLC-Radiogenomics (also known as Stanford) dataset [20–22] contains pre-treatment CT images from 211 patients with NSCLC and both are publicly available from the National Institutes of Health (NIH) mentioned in The Cancer Imaging Archive (TCIA) [20–23]. Patients without available imaging in the online dataset were excluded.

Table 1. Clinical and demographic data including gender, type of NSCLC, and stage of cancer for collected patients in NSCLC-Radiomics-Genomics (Harvard) lung dataset is presented.

| NSCLC-Radiomics-Genomics | | |
|------------------------------------|-------------------------|------------|
| Gender | Male | 61 (68.5%) |
| | Female | 28 (31.5%) |
| Clinical combined stage curated | Stage I | 39 (43.8%) |
| | Stage II | 25 (28.1%) |
| | Stage III | 12 (13.5%) |
| | Unknown | 11 (12.4%) |
| | Adenocarcinoma, | 42 (47.2%) |
| Non-small cell lung cancer (NSCLC) | Squamous cell carcinoma | 33 (37.1%) |
| | Other or unknown | 12 (13.5%) |
| Event | Recurrence or death | 46 (51.7%) |

2.2. Radiomic Feature Extraction and Statistical Analysis

Radiomic features can be divided into categories, for example: first-order features, which include tissue density, shape features (i.e., volume and surface area) and texture features, describing spatial patterns of voxel intensities [5,7,9,11–17]. The proposed

approach employs 429 radiomics features in nine categories: first-order statistics (FO) (18 features), shape-based expression (SB) (13 features), gray level co-occurrence matrix (GLCM) (23 features), gray level dependence matrix (GLDM) (14 features), gray level run length matrix (GLRLM) (16 features), gray level size zone matrix (GLSZM) (16 features), neighboring gray tone difference matrix (NGTDM) (5 features), Laplacian of Gaussian (LOG) (180 features), and three-layer filtering wavelet (144 features) features (Supplementary Materials Table S10).

Four readers with different levels of training performed manual segmentations on Neuroimaging Informatics Technology Initiative (NIFTI) format images and included a data scientist (BY) with no formal medical experience, a medical student (LS), a radiology trainee (MH) with 5 years of clinical radiology experience, and a specialty-trained thoracic radiologist (SK) with 18 years of experience. The data scientist (BY) used the snake feature of ITkSnap region growing tool, while he manually selected the region of tumors in the CT images, adjusted the contrast, set initial bubbles, controlled them to grow to a substantial size, and manually with a brush tool cleaned the areas that were not in the boundaries or exceeded them. The reader with the most experience (SK) was defined as the reference standard (RS) used for benchmarking. Prior to performing segmentations, each reader performed a NSCLC tumor segmentation in a training set of 10 cases from a different source (institution PACS system) supervised by the specialty-trained radiologist (SK) and received feedback on segmentation methods. After completing the training set, each observer completed segmentations of tumors for the complete data set of CT exams. The tumors were labeled in 3D on standard lung windows using ITkSnap (version 3.6.0) [24] by each reader. Segmentations were only performed once per patient per reader, taking breaks between segmentations at the discretion of the reader. A total of 429 radiomic features were extracted within the tumor volume of each image using the *Pyradiomics* library (v2.2.0) and analyzed using low-rank representations of radiomics using principal component analysis and selecting the first principal component (PC) corresponding to the maximum variance in the radiomics. The radiomic analyses were carried out in Python programming language (3.6.8), while the survival analyses were conducted in R programming software (4.0.1). Correlation between the extracted features and agreement between 3D segmentations were analyzed using a Pearson correlation coefficient and Sørensen-Dice coefficient [25], respectively. Dice coefficient measures variabilities of the segmented regions, and low-rank correlation shows its corresponding effect on radiomics by calculating correlation for direction of the maximum variances. In other words, correlation among three first PCs represent the correlation of the entire radiomics (all 429 radiomics). Appendix A provides additional information regarding principal component analysis (PCA). The proposed approach involves using machine learning to reduce the radiomic dimensionality and predict survival using PCA and Cox regression models, which increases the importance of applying unsupervised and supervised models' integration.

Cox regression modeling was performed for each dataset, incorporating radiomic phenotypes, and clinical and demographic data (i.e., sex, stage status, and histology). Kaplan-Meier curves of overall survival were generated for each dataset to determine if contributing radiomic signatures were able to stratify high- and low-risk patients.

3. Results

3.1. Patient Population

A total of 89 patients were in the NSCLC- Radiomics-Genomics-Lung3 dataset, 3 of whom did not have available data and were excluded from the study. There were 42 patients with adenocarcinoma, 32 patients with squamous cell carcinoma, and 12 patients with another type of NSCLC. Thirty-nine patients had stage I disease, 26 patients had stage II disease, 10 patients had stage III disease, and 11 patients had an unknown stage. Of the NSCLC-Radiogenomics data in the NIH-TCIA dataset, 4 patients were excluded from the study for a total of 207 patients included. Of the included tumors in the Harvard

dataset, all were solid, and of the included tumors in the Stanford dataset, 134 were solid, 68 were subsolid, and 5 were unknown.

The total number of patients included in the study is described in Figure 2. Clinical information and demographics of patients are provided in Tables 1 and 2.

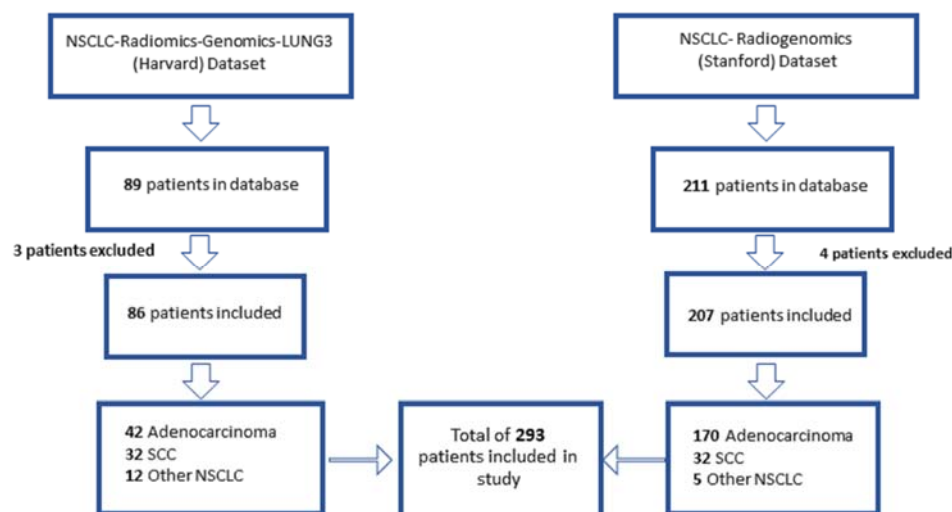


Figure 2. Number of patients included in study. Two publicly available datasets were analyzed in the study, the NSCLC-Radiomics-Genomics-Lung3 (Harvard) dataset and the NSCLC-Radiogenomics (Stanford dataset). Eighty-nine patients and 211 patients are part of the Harvard and Stanford datasets, respectively. A total of 3 patients were excluded from the Harvard dataset and 4 patients were excluded from the Stanford dataset due to lack of available data. Tumor types consisted of adenocarcinoma (Adeno), squamous cell carcinoma (SCC), and other types of NSCLC. A total of 293 patients were segmented as part of the study.

Table 2. Clinical and demographic data including age, race, type of NSCLC, EGFR, and KRAS receptor status, and smoking status for collected patients NSCLC-Radiogenomics (Stanford) is presented.

| NSCLC-Radiogenomics | | |
|----------------------|-------------------------------------|-------------|
| Age | Median (±IQR) | 69 (43,87) |
| Gender | Male | 133 (64.2%) |
| | Female | 74 (35.8%) |
| Race | Caucasian | 120 (57.4%) |
| | Asian | 24 (11.8%) |
| | Hispanic/Latino | 5 (2.4%) |
| | African-American | 6 (2.9%) |
| | Native Hawaiian/Pacific Islander | 3 (1.5%) |
| | Unknown | 48(23.2) |
| Smoking Status | Non-smoking | 47 (22.7%) |
| | Smoking | 34 (16.4%) |
| | Former smoking | 126 (60.9%) |
| EGFR-Mutation Status | Wildtype | 128 (61.8%) |
| | Mutant | 42 (20.2%) |
| | Unknown | 37 (17.8%) |
| KRAS Mutation Status | Wildtype | 130 (62.8%) |
| | Mutant | 38 (18.3%) |
| | Unknown | 39 (18.8%) |
| Histology | Adenocarcinoma | 170 (82.1%) |
| | Squamous cell carcinoma | 32 (15.5%) |
| | NSCLC NOS (not otherwise specified) | 5 (2.4%) |

| | | |
|--------------------------------|---------------------|-------------|
| Solid-Subsolid (Morphology) | Solid | 134 (64.7%) |
| | Subsolid | 68 (32.8%) |
| | Unknown | 5 (2.4%) |
| Event | Recurrence or death | 41(21.1%) |

3.2. Analysis of Interobserver Variability on Radiomic Feature Extraction

From the 429 radiomic features initially extracted from the tumors on CT images, the feature-level was reduced to 3 radiomic signatures (three first PCs) for all the segmenters (Figure 1). The correlation coefficient among the low rank radiomic signatures showed significant correlation among the segmenters with a correlation of greater than 0.7 for all the cases (Table 3).

Table 3. Similarity of the radiomic signatures using multiple scoring methods among different segmenters are presented.

| NSCLC Dataset | Similarity among Segmenters | | | | | | |
|------------------------------------------|-----------------------------|-------------------|---------------------|--------------------|--------------------|--------------------|-------------------|
| | Segmenters ID | Correlation score | Dice Score | Precision (%) | Recall (%) | Boundary Distance | Volume Difference |
| LUNG3 | BY | 0.92 | 0.89 (± 0.25) | 81.8(± 21.8) | 86.1(± 24.5) | 1.2(± 2.7) | 1.1(± 0.5) |
| NSCLC-Radiomics-Genomics Harvard Dataset | LS | 0.94 | 0.82 (± 0.14) | 81.2(± 2.7) | 69.6(± 24.5) | 6.5(± 26.4) | 2.3(± 21.1) |
| | MH | 0.95 | 0.84 (± 0.20) | 72.3(± 22.4) | 88.7(± 18.9) | 4.2(± 15.1) | 0.6(± 1.9) |
| NSCLC-Radiogenomics Stanford Dataset | BY | 0.93 | 0.69 (± 0.28) | 77.8(± 25.1) | 87.3(± 25.2) | 2.92(± 10.7) | 0.3(± 0.8) |
| | LS | 0.72 | 0.80 (± 0.27) | 84.2(± 31.5) | 47.8(± 29.9) | 16.6(± 52.6) | 0.3(± 1.2) |
| | MH | 0.87 | 0.83 (± 0.23) | 80(± 24.3) | 77.1(± 24.7) | 6.2(± 26.1) | 1.4(± 16.9) |

Corr coefficients using the first principal component between BY-SK (RS), LS-SK (RS), and MH-SK (RS) were 0.92, 0.94, and 0.95 (all having p -value < 0.005) for NSCLC-Radiomics-Genomics, and were 0.93, 0.72, and 0.87 (all having p -value < 0.005) for NSCLC-Radiogenomics, respectively, all indicating a strong correlation. The comparison of three significant radiomic descriptors corresponding with each group of segmentations showed 88.9% and 92.7% correlation of radiomics of each set with RS. Principal component analysis of the first three principal components demonstrates that, in some cases, there is a large standard deviation (STD), but the medians of the principal component analyses for the extracted features are similar and still have good correlation (Figure 3).

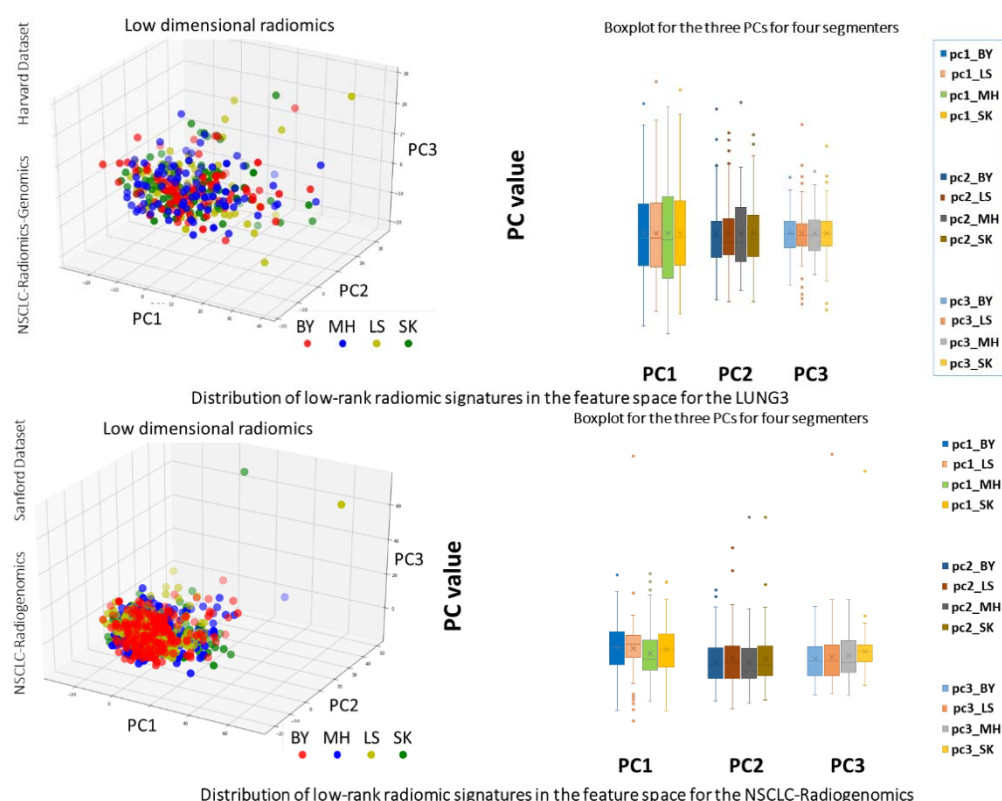


Figure 3. Two visual comparisons of low-rank radiomics representation with their boxplots relation for labels provided by BY, LS, MH, and SK for two different NSCLC Radiogenomics datasets.

The Dice coefficients for the 3D masks for Harvard NSCLC-Radiomics-Genomics and Stanford NSCLC-Radiogenomics for each segmenter (Table 3) was 0.894 (STD: ± 0.25)–0.71 (STD: ± 0.28) for the image scientist (BY)—Reference Standard (SK), 0.82 (STD: ± 0.14)–0.80 (STD: ± 0.27) between the medical student (LS)—Reference Standard (SK), and 0.839 (STD: ± 0.20)–0.83 (STD: ± 0.23) between the radiology trainee (MH)—Reference Standard (SK), respectively. Although the SD coefficients indicate a moderately high spatial agreement of the segmentations, there was some variability between segmentations for BY-SK (RS), LS-SK(RS), and MH-SK (RS) (Figure 4). Precision of the analyses for all segmenters for both NSCLC datasets showed relatively similar precision in segmenting the tumors, where BY-SK(RS) in Harvard and LS-SK(RS) in Stanford datasets have the highest precision yielded to 81.8% ($\pm 21.8\%$) and 84.2% ($\pm 31.5\%$), respectively. MH-SK(RS) and BY-SK(RS) showed the highest recall with 88.7% ($\pm 18.9\%$) and 87.3% ($\pm 25.2\%$), respectively. This pattern showed consistency with the minimum volume difference for MH-SK(RS), 0.6(± 1.9), in Harvard dataset, and BY-SK(RS), 0.3(± 0.8), and LS-SK(RS), 0.3(± 1.2), shared minimum volume difference in Stanford (See Table 3). We conducted in-depth correlation analysis for individual radiomics and showed the results based on radiomics' categories (Supplementary Materials Table S8). Moreover, we presented some radiomics that showed lesser stability among the segmenters in this study (Supplementary Materials Table S9).

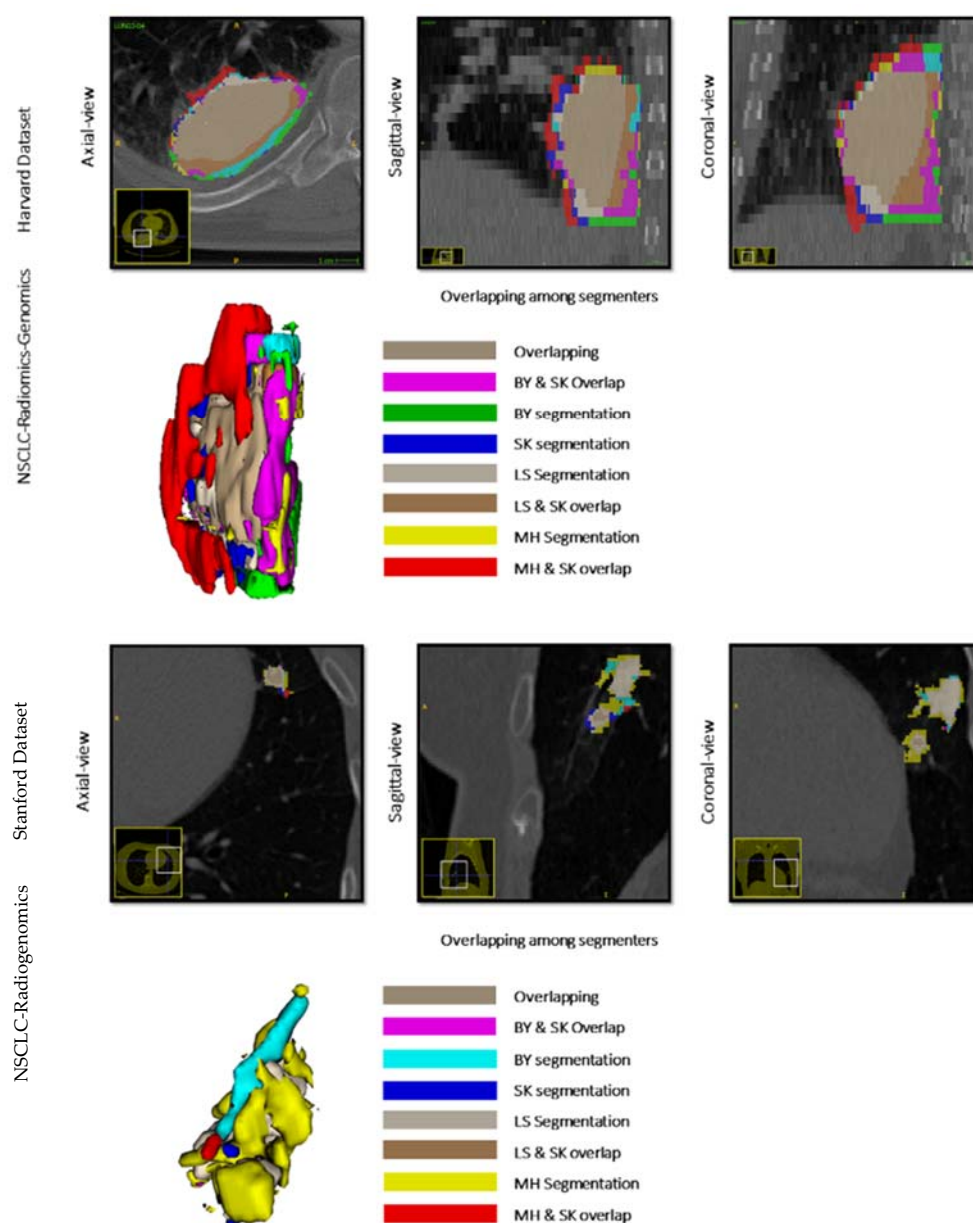


Figure 4. 3D tumor volume. 3D tumor volumes for four segmentation cases and two different NSCLC Radiogenomics datasets.

Cox regression modeling of overall survival for the NSCLC-Radiomics-Genomics-Lung3 (Harvard) and NSCLC-Radiogenomics (Stanford) datasets yielded a c-statistic of 0.64 (95% CI) and 0.6 (95% CI), respectively, for the model including only the clinical (sex, smoking status, and histology) and demographic covariates, which increased when adding radiomic signatures, having of c-statistic of 0.7 (95% CI) and 0.69 (95% CI), respectively. Adding clinical and demographic data to this model yielded an increase in c-statistic, although with slightly increased variability: 0.05–0.02 and 0.01–0.02 for NSCLC-Radiomics-Genomics-Lung3 and NSCLC-Radiogenomic datasets, respectively (Table 4).

Table 4. Overall survival, Cox regression. Using the low-rank representation of the radiomic signatures survival prediction is measured for each segmenter.

| NSCLC Datasets | Modeling Covariates | Prediction Survival | | | | | | | |
|-------------------------------------------------------|-----------------------------------------------|----------------------|-----------------------------------|----------------------|-----------------------------------|----------------------|-----------------------------------|----------------------|-----------------------------------|
| | | BY | | LS | | MH | | SK-RS | |
| | | c-Statistic (95% CI) | <i>p</i> versus Null ¹ | c-Statistic (95% CI) | <i>p</i> versus Null ¹ | c-Statistic (95% CI) | <i>p</i> versus Null ¹ | c-Statistic (95% CI) | <i>p</i> versus Null ¹ |
| LUNG3 NSCLC-Radiomics-Genomics Harvard Dataset | clinical and demographic ² | | | | | | | 0.64 | 0.2 |
| | Three PC radiomic signatures | 0.6 | 0.5 | 0.62 | 0.08 | 0.59 | 0.2 | 0.65 | 0.03 |
| | Radiomic signatures, clinical and demographic | 0.65 | 0.3 | 0.68 | 0.04 | 0.66 | 0.2 | 0.7 | 0.03 |
| NSCLC-Radiogenomics Stanford Dataset | clinical and demographic ³ | | | | | | | 0.6 | 0.007 |
| | Three PC radiomic signatures | 0.65 | 0.001 | 0.64 | 0.04 | 0.67 | 0.003 | 0.65 | 0.003 |
| | Radiomic signatures, clinical and demographic | 0.71 | <0.005 | 0.68 | 0.003 | 0.71 | <0.005 | 0.69 | <0.005 |

CI: confidence interval. 1. *p*-value by likelihood ratio test versus the hypothesis that the model is no better than the null model. 2. Clinical and demographic covariates for LUNG3-NSCLC-Radiomics-Genomics Harvard Dataset: sex, stage status, and histology. 3. Clinical and demographic covariates for NSCLC-Radiogenomics Stanford Dataset: sex, morphological status, and histology.

Additional Cox regression analysis data are presented in the supplemental materials. Kaplan-Meier curves of survival prediction for each dataset showed significant discrimination between high- and low-risk patients using extracted radiomic signatures ($p < 0.01$) and are presented in Figure 5. Median risk score was used as a distinguishing criterion for signifying high- and low-risk groups. The hazard ratio for each covariate in the maximal model is fully reported in the Supplementary Materials Table S11.

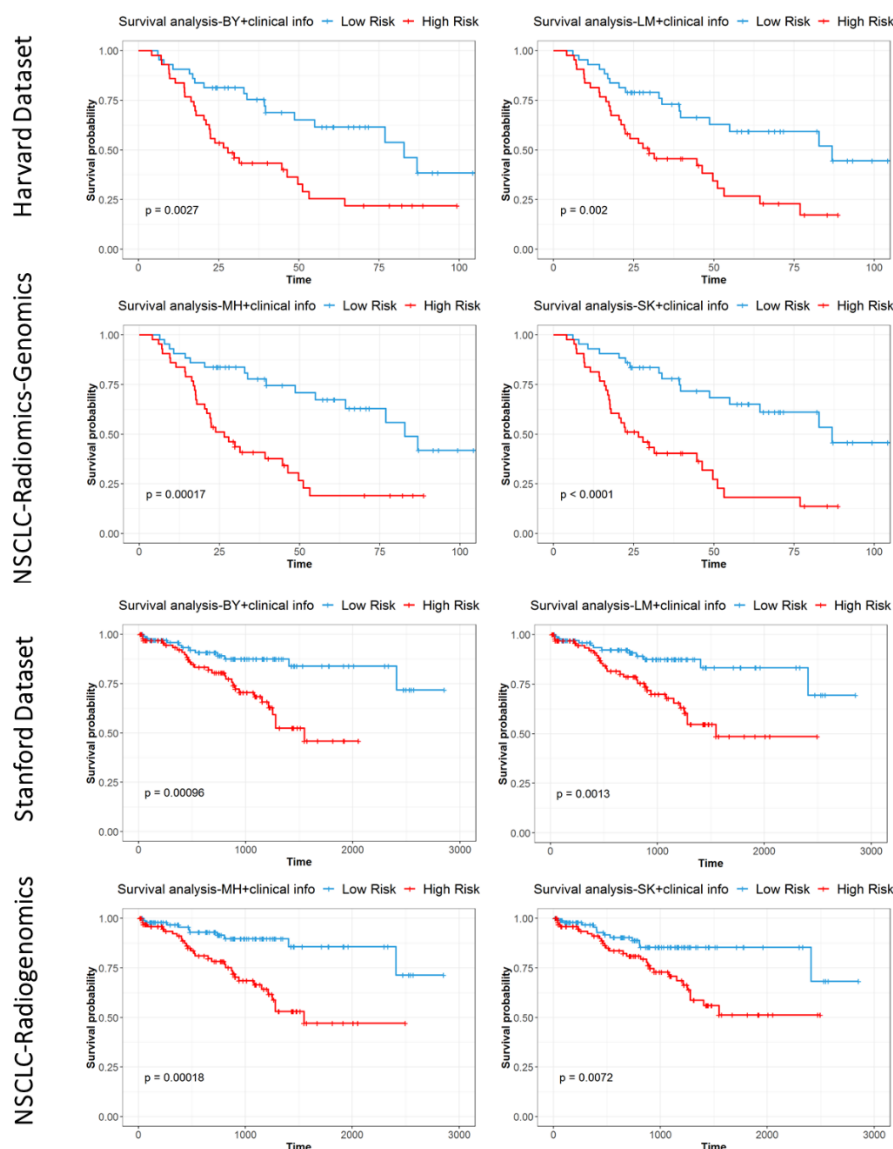


Figure 5. Kaplan-Meier curves for multivariate models of overall survival using low-rank radiomics show significant differences between high- and low-risk patients for each segmenter and NSCLC dataset using median risk score in the model.

4. Discussion

CT imaging is the workhorse of oncology staging and treatment response assessment. However, we now know that conventional imaging has imbedded “radiomic” features that are not appreciable by the eye but contain information on tumor heterogeneity that are reflections of the underlying tumor structure and can be harnessed to generate prognostic and predictive biomarkers. In addition, the morphologic qualitative descriptors used in conventional reporting of radiologic assessments of tumors on CT, such as “spiculated”, “heterogeneous”, and “necrotic”, while clinically useful, are subject to inter and intraobserver variability [10] due to their subjective nature; radiomic signatures may allow for more quantitative and precise measure of tumor description, potentially enhancing the clinical value of these interpretations.

In addition to providing a more quantitative approach to conventional morphologic descriptors, radiomics offers the potential to reveal aspects of tumor phenotype not discernable by the human eye, providing another layer of valuable information that can be extracted from conventional imaging for clinical management. Several studies have described the significance of these additional imaging features and radiomics in cancer

imaging [26–37] and have hypothesized that tumor genetic and cellular characteristics and phenotypes can be represented with medical imaging [38–40]. For example, studies by Ganeshan et al. [41–43] reported an association of extracted NSCLC CT tumor features with patient survival, tumor stage, metabolism, angiogenesis, and hypoxia. The importance of imaging in treatment planning and outcomes was demonstrated by El Naga et al. [44] for head and neck and cervical cancers, and Vaidya et al. [45] for lung cancer. Huang et al. [4] concluded that EGFR mutation status can be determined using quantitative imaging from extracted tumor phenotypes in NSCLC. Similarly, Bardia et al. [46] found that combining radiomic phenotypes, clinical variables, and circulating tumor DNA (ctDNA), enhanced prediction of EGFR-targeted therapy outcomes for NSCLC.

However, while the use of extracted radiomic features from conventional imaging poses exciting possibilities for precision medicine, there are challenges to clinical translation that must be overcome before the use of these novel techniques can become a reality in routine practice. There is variability introduced in the acquisition of imaging, for example the use of different imaging protocols, reconstruction algorithms, and scanner types. In addition, variability is introduced through choice of imaging processing techniques, such as choice of segmentation and feature extraction software, and degree of skill of the reader performing 3D segmentation. Variability is a particular concern with manual segmentations [47], and several studies have reported significant inter-clinician variation in contouring of tumors in radiation treatment planning, including head and neck, lung, prostate, and esophageal cancers [48–52]. In this study, we did find some variability between segmentations performed by the data scientist (BY), the medical student (LS), the radiology trainee (MH), and the most experienced reader, reference standard (SK). However, the SD coefficients suggest an overall moderate to high degree of spatial agreement of the segmentations and good overlap of tumor segmentations between readers.

Interobserver variability between readers in this study may have been introduced by several factors. One factor is differentiating between the boundaries of tumor and adjacent post-obstructive atelectasis [53,54] or pneumonia, a known problem with tumor delineation. In non-contrast CT examinations, it may also be difficult to delineate tumor and adjacent vascular structures that course in and adjacent to lung cancer, especially if the tumor abuts the hilum or mediastinum. Some lung cancers also demonstrated both a solid and a ground glass component, which can introduce variability in the choice of where to draw the boundary around faint ground glass components. Huang et al. [4] discovered that trained radiologists tended to focus on the solid component of a tumor as opposed to the ground glass component, whereas junior radiologists tended to include more of the ground glass component in their segmentations. The inclusion of more ground glass component would increase overall tumor volume and impact the spectrum of radiomic features extracted, thus a risk factor for variation. Window width and level settings on CT may also influence segmentations and gross tumor volumes [54–57]. ITkSnap software allows the reader to choose the window width and level settings in addition to an automatic window width/level selection. While some of our readers manually and arbitrarily adjusted the window width/level based on preference and ability to differentiate tumor from adjacent structures, other readers chose the automated window width/level setting chosen by the software.

Radiomic features used in this study follow imaging features defined by the Imaging Biomarker Standardization Initiative (IBSI). However, differences in CT exam parameters may also introduce segmentation variability between readers. This is particularly true with certain texture features such as coarseness and contrast, which tend to be the least reproducible. First order features, particularly entropy, are found to be the most reproducible [18]. Leijenaar et al. [58] found that radiomic features with high test-retest repeatability suffered less from interobserver differences. A few studies have confirmed that tube current (mAs) or tube voltage (kVp) had no influence on feature reproducibility [59,60]. Varying slice thicknesses of CT scans can also introduce variability in the extracted features, with 1–2.5 mm being the recommended slice thickness when contouring tumors

[17,61]. Our study used a publicly available online dataset with slice thickness varying from 1–5 mm (Supplemental Tables S1, and S2). We conducted in-depth analyses on the effect of CT parameters on the outcome of the selected features using the proposed approach and their final survival outcomes (Supplement Tables S3, S4, S5, S6, and S7). Our supplemental analyses testing the potential effects of CT parameters indicated that there was an overall similarity among segmentations between readers when considering contrast-enhancement, CT kernel, and slice thickness.

The degree of medical specialty training has been a concern for the introduction of variability in segmentations of tumors. Logue et al. [62] reported that radiologists tended to contour smaller gross tumor volumes compared to radiation oncologists in the segmentation of bladder cancers and concluded that a more correct anatomic gross tumor volume was provided by radiologists likely due to clinical practice differences, since radiation oncologists typically select more inclusive volumes around tumors in practice so as not to underestimate tumor extent radiation treatment planning [63]. Similar results were observed in NSCLC by Giraud et al. [64], who noted major discordances between radiation oncologists' and radiologists' tumor delineations, radiologists tending to delineate smaller volumes. In this same study, junior physicians included as readers tended to delineate smaller and more homogeneous volumes compared to senior physicians regardless of their specialty. Van de Steene et al. [63] looked at specialty dependence between junior and senior radiation oncologists, one pulmonologist, and one radiologist, on contouring lung cancer gross tumor volumes and noticed that the radiologist ended up with the smallest tumor volume. They also noted good agreement between the senior radiation oncologist and radiologist. Haga et al. [65] concluded that NSCLC tumor volumes should be contoured by a specialist, such as a radiation oncologist, in order to decrease tumor delineation uncertainty and overestimation of prognostic power in radiomic feature analysis. In this study we compared tumor segmentations between level of training (i.e., medical student, radiology trainee, and radiology attending), and specialty type (i.e., data scientist). Interestingly, the 3D masks in the Harvard Dataset for BY-SK (RS) had an overall higher correlation compared to the masks for MH-SK (RS) and LS-SK (RS) in the segmentation analysis. However, the 3D masks in the Stanford dataset for MH-SK (RS) had an overall higher correlation compared to the masks for BY-SK (RS) and LS-SK (RS). The Pearson correlation coefficients, comparing three significant radiomic phenotypes for PCA, were all relatively equal amongst segmenters in the Harvard dataset, although the correlation coefficients were slightly more variable in the Stanford dataset. Overall, these differences are small and can probably be overlooked given overall high correlation of segmentations amongst all segmenters in the principal component analysis. It should be noted, however, that all readers in this study participated in a training set of cases supervised by the reference standard (SK) to ensure a standard approach to contouring.

Our study had several limitations. The CT scans in the dataset had varying slice thicknesses, ranging from 1–5 mm, which is known to introduce some variability as described above. Additionally, while all the readers used ITKSnap software for segmentation, there was some variability in methods of tumor contouring, such as choice of purely manual or semi-automated tools and the exact window and level used to perform the contouring. However, while there was interobserver variability in contouring, the extracted radiomic features of both the medical student, radiology trainee, and data scientist were overall well correlated with the experienced reader (RS). Another limitation is that the readers were all trained by the expert reader; however, the number of training cases was small and consisted of feedback of the segmentations. Additionally, the training cases were from a different source than the databases that were used for analysis. Despite the limitations, overall correlation of extracted features between readers supports the inclusion of readers of various levels of training in performing segmentations for NSCLC.

Future research would include testing interobserver variability based on level and type of experience against other publicly and readily available datasets and testing intraobserver variability. Other future directions should include determining how factors

such as slice thickness, pixel spacing, window width/level, contrast enhancement, and pre- and post-processing of CT imaging affect interobserver variability between readers of different experience.

5. Conclusions

Although there is some variability in tumor contouring for imaging segmentations between readers, the extracted radiomic features were overall well correlated in observers. Therefore, level of training and clinical experience of the reader may not have a substantial impact on extracted radiomic features of NSCLC on CT, noting that all readers did have a supervised training set prior to contouring cases. Having more readers to perform tumor segmentations may accelerate the development of radiomic signatures in NSCLC that can provide added value to cancer management and precision medicine. This study shows that a greater degree of inclusion of personnel is allowable to perform these tumor segmentations.

Supplementary Materials: The following are available online at www.mdpi.com/article/10.3390/cancers13235985/s1, Table S1. The CT parameters for Stanford NSCLC Radiogenomics dataset, Table S2. The CT parameters for Harvard NSCLC Radiomics-Genomics dataset, Table S3. Similarity of the radiomic signatures using Pearson correlation among different segmenters are presented for different stratifications based on CT, Table S4. Overall survival, Cox regression. Using the low-rank representation of the radiomic signatures survival prediction is measured for each segmenter while there is Contrast-Enhancement (CE), Table S5. Overall survival, Cox regression. Using the low-rank representation of the radiomic signatures survival prediction is measured for each segmenter while there is Non-Contrast-Enhanced (UN), Table S6. Overall survival, Cox regression. Using the low-rank representation of the radiomic signatures survival prediction is measured for each segmenter for higher convolutional kernel (CKh), Table S7. Overall survival, Cox regression. Using the low-rank representation of the radiomic signatures survival prediction is measured for each segmenter for slice thickness between 2mm and 4mm, Table S8. Itraiclass correlation coefficient based on radiomics categories and with the respect of different group means. For each segmenter, mean and standard deviation of correlation coefficient is calculated for every radiomics' category, Table S9. Radiomic features with lesser stability with the respect to different segmenters. Means and standard deviations of these radiomics are presented. Table S10. More detailed information about the Radiomic features used in this study., Table S11. The hazard ratio for each covariate in the maximal cox proportional hazard model.

Author Contributions: Conceptualization: M.H., B.Y., S.I.K., and D.K.; methodology: M.H., B.Y., L.S., J.C.T., C.A., E.L.C., M.G.-A., L.R., J.M.L., S.I.K., and D.K. validation: M.H., L.S., B.Y. and S.I.K.; formal analysis: B.Y. and D.K.; funding acquisition: none; investigation: M.H., L.S., B.Y., S.I.K., and D.K.; resources: B.Y., D.K., J.C.T., C.A., and E.L.C.; software: B.Y. and D.K. data curation: B.Y. and D.K.; writing—original draft preparation: M.H. and B.Y.; writing—review and editing: M.H., B.Y., L.S., M.G.-A., L.R., J.M.L., J.C.T., C.A., E.L.C., D.K., and S.I.K.; visualization: M.H., B.Y., J.M.L., S.I.K., D.K., and E.L.C.; supervision: S.I.K. and D.K.; project administration: S.I.K. and D.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted using two publicly available data according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board (IRB) of: 1. Stanford University School of Medicine (69) and Palo Alto Veterans Affairs Healthcare System (93), between 7 April 2008 and 15 September 2012 (for Stanford NSCLC cohort). These data in TCIA comply with HIPAA de-identification standards using the Safe Harbor Method as defined in section 164.514(b)(2) of the HIPPA Privacy Rule. 2. Maastricht University Medical Center (MUMC), Maastricht, The Netherlands. This study was conducted according to national laws and guidelines and approved by the appropriate local trial committee at Maastricht University Medical Center (MUMC1), Maastricht, The Netherlands and MAASTRO Clinic, The Netherlands (Data release date: 07-02-2014).

Informed Consent Statement: Patient consent was waived due to the research posing no more than minimal risk to subjects and the waiver does not adversely affect the rights and welfare of the subjects who are involved in the research.

Data Availability Statement: Information on the publicly available datasets used in this study [19–21].

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Considering that $X_{Ob1} = \{x_1, x_2, x_3, \dots, x_n\}$ is the result of radiomic features extracted from the first segmenter, where $x_1, x_2 \in \mathbb{R}^p$, is a zero-mean (z-scored) vector (Z-score of a vector \bar{x}_1 defined by $x_1 = \frac{\bar{x}_1 - \mu}{\sigma}$, where μ is mean of \bar{x}_1 , $\mu = \frac{1}{n} \sum_{i=1}^n \bar{x}_i$, and σ is standard deviation of \bar{x}_1), with the size of our radiomic features ($p = 429$) and n was 86 and 207 for the Harvard and Stanford NSCLC Radiogenomic datasets, respectively. There were four segmenters, hence there exists a set of different observers $\{X_{Ob1}, X_{Ob2}, X_{Ob3}, X_{Ob4}\}$. The problem is to reduce the collinearity among the features in each X_{Ob} . For that, we propose using low-rank representation of each vector in the direction of maximum variance, using eigen decomposition method presented in the following section.

Appendix A.1. Low-Rank Representation of Radiomics

Principal component analysis (PCA) [66,67] is used for many applications such as dimension reduction, noise elimination, and classification, amongst others. The PCA can be performed by using a covariance matrix calculation with singular value decomposition (SVD) [68]. The decomposition matrix is performed for the input matrix (heat matrix) X which is $p \times n$, where n is the vectorized thermal image in every sequence and p corresponds to the number of observations, and decomposes to:

$$X = U\Sigma V^T \quad (A1)$$

where $k > p$ and Σ is a diagonal matrix with a dimension of $p \times p$ and either zero or positive elements. It is considered as the singular value of matrix X and U is the $p \times n$ matrix denotes as eigenvector or basis matrix of X . The data are arranged column-wise based on the observation variation. Spatial variations are mapped in the row direction (input data located in columns and rows show the observations). The PCA is a linear transformation method, which applies a decomposition of the input zero-mean data matrix into the basis U and coefficient matrix Σ . The basis matrix carries the orthonormal property that also maximizes the variance of projected data which leads to the principal components (PCs) of the input matrix.

Selecting the $k = 3$ to reduce dimensionality from 429 to 2 for each segmenter we would use Equation (A1), to convert X_{Ob} to U_{Ob} , where $U_{Ob} \in \mathbb{R}^{3 \times n}$, and $\{U_{Ob1}, U_{Ob2}, U_{Ob3}, U_{Ob4}\}$. The resulting comparison of the radiomic signatures is thus facilitated by this dimensionality reduction. The three initial PCs used to measure the correlation of radiomics by each segmenter corresponding to their Dice-scores, while survival analysis uses only the initial PCs. PCA selects the initial predominant eigenvectors, known as bases of analysis, and provides the highest variance among the radiomics. In other words, PCA finds the best signatures exist in the radiomics, we compared the best representative of radiomics for each segmenter with our reference to find overall correlation of radiomics.

Appendix A.2. Low-Rank Correlation of Interobserver's Radiomics

PCA between the four reader segmentations was performed on the extracted features with a Pearson correlation coefficient (*corr*) using the first principal components. A high degree of correlation between the extracted features was defined between ± 0.50 to ± 1 , a moderate degree of correlation was defined between ± 0.30 to ± 0.49 and a low degree of correlation was defined as $< \pm 0.29$. The Pearson correlation coefficient (PCC), or the bivariate correlation, [69] allows measurement of the linear correlation between two variables or vectors, X and Y . PCC calculates covariance of two variables divided by standard

divisions of both variables, involving the product moment. The Pearson's correlation coefficient, r_{xy} , is measured using the following formula:

$$\text{Corr}(x, y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (\text{A2})$$

where n is the sample size, x_i, y_i corresponds to the individual sample points with i , and $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ and analogously for \bar{y} .

Interobserver variability in the 3D segmentations between the readers and Reference Standard was performed also using a Sørensen-Dice (SD) coefficient to evaluate spatial agreement of the segmentations. High spatial agreement was defined as a SD between 0.7–1.0, moderate was defined by a SD between 0.5–0.7, and low spatial agreement was defined as a SD < 0.5. The results of Dice also indicate high spatial agreement among the segmenters by having SD > 0.7. The Sørensen–Dice coefficient (DSC) [25,70] is calculated by the following formula:

$$\text{DSC}_{xy} = \frac{2|X \cap Y|}{|X| + |Y|} \quad (\text{A3})$$

By use of these two methods of measuring the similarity, two coefficients are produced to gauge the pair-wised similarity between each two segmenters. Using Equations (A2) and (A3), we calculate $\text{Corr}_{i,j}(U_{\text{obj}}, U_{\text{obj}})$, and $\text{DSC}_{i,j}(U_{\text{obj}}, U_{\text{obj}})$ where $i \neq j$, respectively. The results of these two measures indicate the variability among the low-rank radiomic signatures. Table 3 shows the results of such correlation among the segmenters.

References

1. Aberle, D.R.; Adams, A.M.; Berg, C.D.; Black, W.C.; Clapp, J.D.; Fagerstrom, R.M.; Gareen, I.F.; Gastonis, C.; Marcus, P.M.; Sicks, J.D.; et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N. Engl. J. Med.* **2011**, *365*, 395–409.
2. van Baardwijk, A.; Wanders, S.; Boersma, L.; Borger, J.; Ollers, M.; Dingemans, A.M.; Bootsma, G.; Geraedts, W.; Pitz, C.; Lunde, R.; et al. Mature results of an individualized radiation dose prescription study based on normal tissue constraints in stages I to III non-small-cell lung cancer. *J. Clin. Oncol.* **2010**, *28*, 1380–1386.
3. Parmar, C.; Rios Velazquez, E.; Leijenaar, R.; Jermoumi, M.; Carvalho, S.; Mak, R.H.; Mitra, S.; Shankar, B.U.; Kikinis, R.; Haibe-Kains, B.; et al. Robust Radiomics feature quantification using semiautomatic volumetric segmentation. *PLoS ONE* **2014**, *9*, e102107.
4. Huang, Q.; Lu, L.; Dercle, L.; Lichtenstein, P.; Li, Y.; Yin, Q.; Zong, M.; Schwartz, L.; Zhao, B.; et al. Interobserver variability in tumor contouring affects the use of radiomics to predict mutational status. *J. Med. Imaging (Bellingham)* **2018**, *5*, 011005.
5. Balagurunathan, Y.; Kumar, V.; Gu, Y.; Kim, J.; Wang, H.; Liu, Y.; Goldgof, D.B.; Hall, L.O.; Korn, R.; Zhao, B.; et al. Test-retest reproducibility analysis of lung CT image features. *J. Digit. Imaging* **2014**, *27*, 805–823.
6. Wu, W.; Parmar, C.; Grossmann, P.; Quakenbush, J.; Lambin, P.; Bussink, J.; Mak, R.; Aerts, H.J. Exploratory study to identify radiomics classifiers for lung cancer histology. *Front. Oncol.* **2016**, *6*, 71.
7. Fried, D.V.; Tucker, S.L.; Zhou, S.; Liao, Z.; Mawlawi, O.; Ibbott, G.; Court, L.E. Prognostic value and reproducibility of pre-treatment CT texture features in stage III non-small cell lung cancer. *Int. J. Radiat. Oncol. Biol. Phys.* **2014**, *90*, 834–842.
8. Aerts, H.J.; Grossmann, P.; Tan, Y.; Oxnard, G.R.; Rizvi, N.; Schwartz, L.H.; Zhao, B. Defining a radiomic response phenotype: A pilot study using targeted therapy in NSCLC. *Sci. Rep.* **2016**, *6*, 33860.
9. Coroller, T.P.; Grossmann, P.; Hou, Y.; Rios Velasquez, E.; Leijenaar, R.T.; Hermann, G.; Lambin, P.; Haibe-Kains, B.; Mak, R.H.; Aerts, H.J. CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma. *Radiother. Oncol.* **2015**, *114*, 345–350.
10. Yip, S.S.; Aerts, H.J. Applications and limitations of radiomics. *Phys. Med. Biol.* **2016**, *61*, R150–R166.
11. Aerts, H.J.; Velazquez, E.R.; Leijenaar, R.T.; Parmar, C.; Grossmann, P.; Carvalho, S.; Cavalho, S.; Bussink, J.; Monshouwer, R.; Haibe-Kains, B.; et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat. Commun.* **2014**, *5*, 4006.
12. Desseroit, M.C.; Visvikis, D.; Tixier, F.; Majdoub, M.; Perdrisot, R.; Guillemin, R.; Cheze le Rest, C.; Hatt, M. Development of a nomogram combining clinical staging with (18)F-FDG PET/CT image features in non-small-cell lung cancer stage I–III. *Eur. J. Nucl. Med. Mol. Imaging* **2016**, *43*, 1477–1485.
13. Fave, X.; Cook, M.; Frederick, A.; Zhang, L.; Yang, J.; Fried, D.; Stingo, F.; Court, L. Preliminary investigation into sources of uncertainty in quantitative imaging features. *Comput. Med. Imaging Graph.* **2015**, *44*, 54–61.
14. Huynh, E.; Coroller, T.P.; Narayan, V.; Agrawal, V.; Romano, J.; Franco, I.; Parmar, C.; Hou, Y.; Mak, R.H.; Aerts, H.J. Associations of radiomic data extracted from static and respiratory-gated CT scans with disease recurrence in lung cancer patients treated with SBRT. *PLoS ONE* **2017**, *12*, e0169172.

15. Kalpathy-Cramer, J.; Mamomov, A.; Zhao, B.; Lu, L.; Cherezov, D.; Napel, S.; Echegaray, S.; Rubin, D.; McNitt-Gray, M.; Lo, P.; et al. Radiomics of lung nodules: A multi-institutional study of robustness and agreement of quantitative imaging features. *Tomography* **2016**, *2*, 430–437.
16. Mackin, D.; Fave, X.; Zhang, L.; Fried, D.; Yang, J.; Taylor, B.; Rodriguez-Rivera, E.; Dodge, C.; Jones, A.K.; Court, L. Measuring computed tomography scanner variability of radiomics features. *Invest. Radiol.* **2015**, *50*, 757–765.
17. Zhao, B.; Tan, Y.; Tsai, W.Y.; Qi, J.; Xie, C.; Lu, L.; Schwartz, L.H. Reproducibility of radiomics for deciphering tumor phenotype with imaging. *Sci. Rep.* **2016**, *6*, 23428.
18. Traverso, A.; Wee, L.; Dekker, A.; Gillies, R. Repeatability and reproducibility of radiomic features: A systematic review. *Int. J. Radiat. Oncol. Biol. Phys.* **2018**, *102*, 1143–1158.
19. Aerts, H.J.; Wee, L.; Rios Velasquez, E.; Leijenaar, R.T.; Parmar, C.; Grossmann, P.; Carvalho, S.; Lambin, P. Data from NSCLC-radiomics. *Cancer Imaging Arch.* **2019**, doi:10.7937/K9/TCIA.2015.L4FRET6Z.
20. Clark, K.; Vendt, B.; Smith, K.; Freymann, J.; Kirby, J.; Koppel, P.; Moore, S.; Phillips, S.; Maffitt, D.; Pringle, M.; et al. The Cancer Imaging Archive (TCIA): Maintaining and operating a public information repository. *J. Digit. Imaging* **2013**, *26*, 1045–1057.
21. Bakr, S.; Gevaert, O.; Echegaray, S.; Ayers, K.; Zhou, M.; Shafiq, M.; Zheng, H.; Zhang, W.; Leung, A.; Kadoch, M.; et al. Data for NSCLC radiogenomics collection. *Cancer Imaging Arch.* **2017**, doi:10.7937/K9/TCIA.2017.7hs46erv.
22. Bakr, S.; Gevaert, O.; Echegaray, S.; Ayers, K.; Zhou, M.; Shafiq, M.; Zheng, H.; Benson, J.A.; Zhang, W.; Leung, A.; et al. A radiogenomic dataset of non-small cell lung cancer. *Sci. Data* **2018**, *5*, 180202, doi:10.1038/sdata.2018.202.
23. Gevaert, O.; Xu, J.; Hoang, C.D.; Leung, A.N.; Xu, Y.; Quon, A.; Rubin, D.L.; Napel, S.; Plevritis, S.K. Non-small cell lung cancer: Identifying prognostic imaging biomarkers by leveraging public gene expression microarray data—methods and preliminary results. *Radiology* **2012**, *264*, 387–396.
24. Yushkevich, P.; Piven, J.; Hazlett, H.C.; Smith, R.G.; Ho, S.; Gee, J.C.; Gerig, G. User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *Neuroimage* **2006**, *31*, 1116–1128.
25. Dice, L.R. Measures of the amount of ecologic association between species. *Ecology* **1945**, *26*, 297–302.
26. Meng, Y.; Sun, J.; Qu, N.; Zhang, G.; Yu, T.; Piao, H. Application of radiomics for personalized treatment of cancer patients. *Cancer Manag. Res.* **2019**, *11*, 10851–10858.
27. Dalal, V.; Carmicheal, J.; Dhaliwal, A.; Jain, M.; Kaur, S.; Batra, S.K. Radiomics in stratification of pancreatic cystic lesions: Machine learning in action. *Cancer Lett.* **2020**, *469*, 228–237.
28. Waninger, J.J.; Green, M.D.; Cheze Le Rest, C.; Rosen, B.; El Naqa, I. Integrating radiomics into clinical trial design. *Q. J. Nucl. Med. Mol. Imaging* **2019**, *63*, 339–346.
29. Chaddad, A.; Kucharczyk, M.J.; Daniel, P.; Sabri, S.; Jean-Claude, B.J.; Niazi, T.; Abdulkarim, B. Radiomics in glioblastoma: Current status and challenges facing clinical implementation. *Front. Oncol.* **2019**, *9*, 374.
30. Liu, Z.; Wang, S.; Dong, D.; Wei, J.; Fang, C.; Zhou, X.; Sun, K.; Li, L.; Li, B.; Wang, M. The applications of radiomics in precision diagnosis and treatment of oncology: Opportunities and challenges. *Theranostics* **2019**, *9*, 1303–1322.
31. Rizzo, S.; Botta, F.; Raimondi, S.; Origgi, D.; Fanciullo, C.; Morganti, A.G.; Bellomi, M. Radiomics: The facts and the challenges of image analysis. *Eur. Radiol. Exp.* **2018**, *2*, 36.
32. Lambin, P.; Rios-Velazquez, E.; Leijenaar, R.; Carvalho, S.; van Stiphout, R.G.; Granton, P.; Zegers, C.M.; Gillies, R.; Boellard, R.; Dekker, A. Radiomics: Extracting more information from medical images using advanced feature analysis. *Eur. J. Cancer* **2012**, *48*, 441–446.
33. Lambin, P.; Leijenaar, R.T.; Deist, T.M.; Peerlings, J.; de Jong, E.E.; van Timmeren, J.; Sanduleanu, S.; Larue, R.T.; Even, A.J.; Jochems, A. Radiomics: The bridge between medical imaging and personalized medicine. *Nat. Rev. Clin. Oncol.* **2017**, *14*, 749–762.
34. Chen, B.; Zhang, R.; Gan, Y.; Yang, L.; Li, W. Development and clinical application of radiomics in lung cancer. *Radiat. Oncol.* **2017**, *12*, 154.
35. Wilson, R.; Devaraj, A. Radiomics of pulmonary nodules and lung cancer. *Transl. Lung Cancer Res.* **2017**, *6*, 86–91.
36. Gillies, R.J.; Kinahan, P.E.; Hricak, H. Radiomics: Images are more than pictures, they are data. *Radiology* **2016**, *278*, 563–577.
37. Kumar, V.; Gu, Y.; Basu, S.; Berglund, A.; Eschrich, S.A.; Schabath, M.B.; Forster, K.; Aerts, H.J.W.L.; Dekker, A.; Fenstermacher, D.; et al. Radiomics: The process and the challenges. *Magn. Reson. Imaging* **2012**, *30*, 1234–1248.
38. Henriksson, E.; Kjellen, E.; Wahlberg, P.; Ohlsson, T.; Wennerberg, J.; Brun, E. 2-Deoxy-2-[18F] fluoro-D-glucose uptake and correlation to intratumoral heterogeneity. *Anticancer Res.* **2007**, *27*, 2155–2159.
39. Yang, X.; Knopp, M.V. Quantifying tumor vascular heterogeneity with dynamic contrast-enhanced magnetic resonance imaging: A review. *J. Biomed. Biotechnol.* **2011**, *2011*, 732848.
40. Basu, S.; Kwee, T.C.; Gatenby, R.; Saboury, B.; Torigian, D.A.; Alavi, A. Evolving role of molecular imaging with PET in detecting and characterizing heterogeneity of cancer tissue at the primary and metastatic sites, a plausible explanation for failed attempts to cure malignant disorders. *Eur. J. Nucl. Med. Mol. Imaging* **2011**, *38*, 987–991.
41. Ganeshan, B.; Abaleke, S.; Young, R.C.; Chatwin, C.R.; Miles, K.A. Texture analysis of non-small cell lung cancer on unenhanced computed tomography: Initial evidence for a relationship with tumour glucose metabolism and stage. *Cancer Imaging* **2010**, *10*, 137–143.
42. Ganeshan, B.; Panayiotou, E.; Burnand, K.; Dizdarevic, S.; Miles, K. Tumour heterogeneity in non-small cell lung carcinoma assessed by CT texture analysis: A potential marker of survival. *Eur. Radiol.* **2012**, *22*, 796–802.

43. Ganeshan, B.; Goh, V.; Mandeville, H.C.; Ng, Q.S.; Hoskin, P.J.; Miles, K.A. Non-small cell lung cancer: Histopathologic correlates for texture parameters at CT. *Radiology* **2013**, *266*, 326–336.
44. El Naqa, I.; Grigsby, P.; Apte, A.; Kidd, E.; Donnelly, E.; Khullar, D.; Chaudhari, S.; Yang, D.; Schmitt, M.; Laforest, R. Exploring feature-based approaches in PET images for predicting cancer treatment outcomes. *Pattern Recognit.* **2009**, *42*, 1162–1171.
45. Vaidya, M.; Creach, K.M.; Frye, J.; Dehdashti, F.; Bradley, J.D.; El Naqa, I. Combined PET/CT image characteristics for radiotherapy tumor response in lung cancer. *Radiother. Oncol.* **2012**, *102*, 239–245.
46. Yousefi, B.; LaRiviere, M.J.; Cohen, E.A.; Buckingham, T.H.; Yee, S.S.; Black, T.A.; Chien, A.L.; Noel, P.; Hwang, W.; Katz, S.I.; et al. Combining radiomic phenotypes of non-small cell lung cancer with liquid biopsy data may improve prediction of response to EGFR inhibitors. *Sci. Rep.* **2021**, *11*, 9984, doi:10.1038/s41598-021-88239-y.
47. Emaminejad, N.; Qian, W.; Guan, Y.; Tan, M.; Qiu, Y.; Liu, H.; Zheng, B. Fusion of quantitative image and genomic biomarkers to improve prognosis assessment of early stage lung cancer patients. *IEEE Trans. Biomed. Eng.* **2016**, *63*, 1034–1043.
48. Weltens, C.; Menten, J.; Feron, M.; Bellon, E.; Demaerel, P.; Maes, F.; Van de Bogaert, W.; van der Schueren, E. Interobserver variations in gross tumor volume delineation of brain tumors on computed tomography and impact of magnetic resonance imaging. *Radiother. Oncol.* **2001**, *60*, 49–59.
49. Leunens, G.; Menten, J.; Weltens, C.; Verstraete, J.; van der Schueren, E. Quality assessment of medical decision making in radiation oncology: Variability in target volume delineation for brain tumours. *Radiother. Oncol.* **1993**, *29*, 169–175.
50. Cazzaniga, L.F.; Marinoni, M.A.; Bossi, A.; Bianchi, E.; Cagna, E.; Cosentino, D.; Scandolaro, L.; Valli, M.; Frigero, M. Interphysician variability in defining the planning target volume in the irradiation of prostate and seminal vesicles. *Radiother. Oncol.* **1998**, *47*, 293–296.
51. Hamilton, C.S.; Denham, J.W.; Joseph, D.J.; Lamb, D.S.; Spry, N.A.; Gray, A.J.; Atkinson, C.H.; Wynne, C.J.; Abdelaal, A.; Bydder, P.V. Treatment and planning decisions in non-small cell carcinoma of the lung: An Australasian patterns of practice study. *Clin. Oncol. (R. Coll. Radiol.)* **1992**, *4*, 141–147.
52. Tai, P.; Van Dyk, J.; Yu, E.; Battista, J.; Stitt, L.; Coad, T. Variability of target volume delineation in cervical esophageal cancer. *Int. J. Radiat. Oncol. Biol. Phys.* **1998**, *42*, 277–288.
53. Valley, J.F.; Mirimanoff, R.O. Comparison of treatment techniques for lung cancer. *Radiother. Oncol.* **1993**, *28*, 168–173.
54. Graham, M.V.; Purdy, J.A.; Emami, B.; Matthews, J.W.; Harms, W.B. Preliminary results of a prospective trial using three dimensional radiotherapy for lung cancer. *Int. J. Radiat. Oncol. Biol. Phys.* **1995**, *33*, 993–1000.
55. Senan, S.; van Sörnsen de Koste, J.; Samson, M.; Tankink, H.; Jansen, P.; Nowak, P.J.; Krol, A.D.; Schmitz, P.; Lagerwaard, F.J. Evaluation of a target contouring protocol for 3D conformal radiotherapy in non-small cell lung cancer. *Radiother. Oncol.* **1999**, *53*, 247–255.
56. Harris, K.M.; Adams, H.; Lloyd, D.C.; Harvey, D.J. The effect on apparent size of simulated pulmonary nodules of using three standard CT window settings. *Clin. Radiol.* **1993**, *47*, 241–244.
57. Graham, M.V.; Matthews, J.W.; Harms, W.B.; Emami, B.; Glazer, H.S.; Purdy, J.A. Three-dimensional radiation treatment planning study for patients with carcinoma of the lung. *Int. J. Radiat. Oncol. Biol. Phys.* **1994**, *29*, 1105–1117.
58. Leijenaar, R.T.; Carvalho, S.; Velazquez, E.R.; van Elmp, W.J.; Parmar, C.; Hoekstra, O.S.; Hoekstra, C.J.; Boellaard, R.; Dekker, A.L.; Gillies, R.J.; et al. Stability of FDG-PET Radiomics features: An integrated analysis of test-retest and inter-observer variability. *Acta Oncol.* **2013**, *52*, 1391–1397.
59. Forgacs, A.; Pall Jonsson, H.; Dahlbom, M.; Daver, F.; Difranco, M.; Opposits, G.; Krizsan, A.; Garai, I.; Czernin, J.; Varga, J.; et al. A study on the basic criteria for selecting heterogeneity parameters of F18-FDG PET images. *PLoS ONE* **2016**, *11*, e0164113.
60. Buch, K.; Li, B.; Qureshi, M.M.; Kuno, H.; Anderson, S.W.; Sakai, O. Quantitative assessment of variation in CT parameters on texture features: Pilot study using a nonanatomic phantom. *Am. J. Neuroradiol.* **2017**, *38*, 981–985.
61. Zhao, B.; Tan, Y.; Tsai, W.Y.; Schwartz, L.H.; Lu, L. Exploring variability in CT characterization of tumors: A preliminary phantom study. *Transl. Oncol.* **2014**, *7*, 88–93.
62. Logue, J.P.; Sharrock, C.L.; Cowan, R.A.; Read, G.; Marrs, J.; Mott, D. Clinical variability of target volume description in conformal radiotherapy planning. *Int. J. Radiat. Oncol. Biol. Phys.* **1998**, *41*, 929–931.
63. Van de Steene, J.; Linthout, N.; de Mey, J.; Vinh-Hung, V.; Claassens, C.; Noppen, M.; Bel, A.; Storme, G. Definition of gross tumor volume in lung cancer: Inter-observer variability. *Radiother. Oncol.* **2002**, *62*, 37–49.
64. Giraud, P.; Elles, S.; Helfre, S.; De Rycke, Y.; Servois, V.; Carette, M.F.; Alzieu, C.; Bondiau, P.Y.; Dubray, B.; Touboul, E.; et al. Conformal radiotherapy for lung cancer: Different delineation of the gross tumor volume (GTV) by radiologists and radiation oncologists. *Radiother. Oncol.* **2002**, *62*, 27–36, doi:10.1016/s0167-8140(01)00444-3.
65. Haga, A.; Takahashi, W.; Aoki, S.; Nawa, K.; Yamashita, H.; Abe, O.; Nakagawa, K. Classification of early stage non-small cell lung cancers on computed tomographic images into histological types using radiomic features: Interobserver delineation variability analysis. *Radiol. Phys. Technol.* **2018**, *11*, 27–35.
66. Pearson, K. LIII. On lines and planes of closest fit to systems of points in space. *Lond. Edinb. Dublin Philos. Mag. J. Sci.* **1901**, *2*, 559–572.
67. Hotelling, H. Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.* **1933**, *24*, 417–441.
68. Jolliffe, I.T. *Principle Component Analysis*, 2nd ed.; Springer: New York, NY, USA, 2002; 487p.
69. Pearson, K., VII. Note on regression and inheritance in the case of two parents. *Proc. R. Soc. Lond.* **1895**, *58*, doi:10.1098/rspl.1895.0041.

-
70. Sørensen, T. A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons. *K. Dan. Vidensk. Selsk. Biol. Skr.* **1948**, V, 3–34.