



Published in final edited form as:

Appl Sci (Basel). 2022 October ; 12(19): . doi:10.3390/app12199824.

The Impact of Image Acquisition Parameters and ComBat Harmonization on the Predictive Performance of Radiomics: A Renal Cell Carcinoma Model

Abdalla Ibrahim^{1,*†}, Lin Lu^{1,†}, Hao Yang¹, Oguz Akin², Lawrence H. Schwartz¹, Binsheng Zhao¹

¹Department of Radiology, Columbia University Irving Medical Center, New York, NY 10032, USA

²Department of Radiology, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA

Abstract

Radiomics, one of the potential methods for developing clinical biomarker, is one of the exponentially growing research fields. In addition to its potential, several limitations have been identified in this field, and most importantly the effects of variations in imaging parameters on radiomic features (RFs). In this study, we investigate the potential of RFs to predict overall survival in patients with clear cell renal cell carcinoma, as well as the impact of ComBat harmonization on the performance of RF models. We assessed the robustness of the results by performing the analyses a thousand times. Publicly available CT scans of 179 patients were retrospectively collected and analyzed. The scans were acquired using different imaging vendors and parameters in different medical centers. The performance was calculated by averaging the metrics over all runs. On average, the clinical model significantly outperformed the radiomic models. The use of ComBat harmonization, on average, did not significantly improve the performance of radiomic models. Hence, the variability in image acquisition and reconstruction parameters significantly affect the performance of radiomic models. The development of radiomic specific harmonization techniques remain a necessity for the advancement of the field.

Keywords

handcrafted radiomics; harmonization; ccRCC

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

*Correspondence: ai2481@cumc.columbia.edu.

†These authors contributed equally to this work.

Author Contributions: Conceptualization, A.I., L.L., and B.Z.; methodology, A.I.; software, H.Y.; formal analysis, A.I.; data curation, A.I., L.L., and O.A.; writing—original draft preparation, A.I.; writing—review and editing, L.L., H.Y., O.A., L.H.S., and B.Z.; visualization, A.I., and L.L.; supervision, L.H.S. and B.Z. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/app12199824/s1>, Table S1: Most selected radiomic features in the different approaches.

Informed Consent Statement: Not applicable.

1. Introduction

With radiomics, quantitative medical image analysis can be performed by extracting quantitative features hypothesized to decode biologic information not even detectable by the human eye [1,2]. The field has rapidly expanded, with large numbers of published studies regarding the potential of radiomics for clinical applications. Determining lesions' malignant potential [3–5], assessing response to therapy [6–8], and predicting overall survival [9,10] are some of the clinical endpoints investigated.

Despite the many possibilities with radiomics, there are a number of limitations that have hindered the translation of radiomic signatures (models) into clinical practice [11,12]. As is the case of almost all other biomarkers [13], radiomic features (RFs) must be repeatable and reproducible to be used in clinical settings [14,15]. One of the limitations of real-world data is that CT scanners from different vendors, different models of CT scanners from the same vendor, and the imaging center implementation of these CT scanners result in many variations of image acquisition parameters. Previous studies have reported that the reproducibility of RFs can be significantly affected by the variations in image acquisition parameters including scan acquisition and reconstruction settings [15–19]. Therefore, it is ideal to evaluate and potentially “correct” these settings in order to provide the most accurate radiomics analyses. Two potential solutions for these problems include (1) the selection and exclusive use of robust RFs by studying the results of reproducibility analyses and (2) the reduction of the variations in RF values not attributed to biologic differences [14] by either pre-processing, for example image resampling, and/or post-processing, for example with ComBat harmonization techniques [20].

A number of studies have investigated the effects of different preprocessing methods, such as image resampling [16,21,22], kernel normalization [23–25], and image discretization [26–28]. These studies reported inconsistent performance depending on the data analyzed. ComBat harmonization, a method originally developed for removing batch effects from gene expression arrays [29], has been investigated as a potential harmonization method in radiomics analyses. ComBat is a function that is based on empirical Bayes and attempts to remove the effects attributed to machinery differences while preserving biologic information. This is done by pooling the gene expressions and providing ComBat formula with the biologic covariates that are known to affect the expression of genes being harmonized [29]. However, there are several limitations to fulfilling the assumptions of ComBat in radiomics analyses. First, radiomic features are differently affected by the variations in imaging parameters as previously reported. Therefore, pooling all RFs in a single ComBat harmonization run might affect the performance of ComBat. Second, the aim of radiomic analyses is to investigate the correlations with biologic information. As such, it is not possible to provide the correct biologic covariates to the ComBat formula, risking the loss of biologic information in the process. Lastly, ComBat is a data-driven harmonization method, i.e., the estimates of batch effects can vary significantly with the variations in the data being harmonized. In this manner, every time new data point are added, the harmonization and modeling process have to be repeated [30], which makes the application of ComBat not suitable for clinical practice without specific adjustments of ComBat for RFs. Hence, we are performing this experiment to evaluate these hypotheses.

Clear cell renal cell carcinoma (ccRCC) is the most common subtype of renal adenocarcinoma, the most common type of renal malignancies which account for around 90% of kidney cancers [31]. ccRCC has variable outcomes for individual patients, and consequently, the predictability of overall survival in ccRCC patients remains challenging clinically [32]. RFs could be of potential utility for predicting overall survival in these patients. Previous studies reported promising results [30,33–35]. However, the translation to clinical applications remains extremely challenging due to the above-mentioned limitations. For example, a previous study utilized the same dataset to assess to which degree do the variations in imaging parameters affect the performance of radiomic signatures [30]. The study concluded that the generalizability of radiomic signatures can be negatively affected by multiple factors, such as the differences in slice thickness and tumor size, however, more insight is needed.

In this study, we aim to assess the potential of RFs to predict overall survival in ccRCC patients, using real-world, heterogeneously acquired CT scans and in addition, we also assess the impact of different approaches of ComBat harmonization on the performance of RFs to predict overall survival.

2. Materials and Methods

2.1. Imaging Data

The analyzed data in this study are publicly available on The Cancer Imaging Archive (TCIA) [36]: (<https://wiki.cancerimagingarchive.net/display/Public/TCGA-KIRC>, accessed on 3 April 2022) [37]. We utilize 267 ccRCC patients collected from multiple medical centers nationwide. Of these, 179 CT scans were included in the analysis based on the availability of both the imaging and survival data. The imaging vendors and parameters for this data set are summarized in Table 1.

Additionally, the data contained scans that were acquired in different contrast-enhancement imaging phases (70 arterial scans and 109 portal venous scans). Therefore, the data was analyzed in three ways and three datasets were created accordingly: (i) All of the data—Dataset 1; (ii) Arterial scans only—Dataset 2; and (iii) Portal venous scans only—Dataset 3.

2.2. Tumor Segmentation and Feature Extraction

An abdominal radiologist who was blinded to the patients' clinical information segmented the renal lesions on each scan. A MATLAB-based (MathWorks, Natick, Massachusetts) application was used to visualize and segment the tumors on CT scan. The application is based on a semi-automated algorithm, which combines the region-based active contours and a level set approach, in a slice-by-slice manner. The images were then isotropically resampled to $0.5 \times 0.5 \times 0.5 \text{ mm}^3$. A total of 1160 RFs were extracted from each segmented tumor via the Columbia Image Feature Extractor (CIFE) [30].

2.3. ComBat Harmonization

ComBat method is an empirical Bayes-based method that is used to estimate the effects of different batches on the quantitative readings. For RFs, ComBat formula assumes that a feature value can be approximated by the equation:

$$Y_{ij} = \alpha + \beta X_{ij} + \gamma_i + \delta_i \epsilon_{ij} \quad (1)$$

where α is the average value for HRF Y_{ij} for ROI j on scanner i ; X is a design matrix of the biologic covariates that are known to affect the value of HRFs; β is the vector of regression coefficients corresponding to each biologic covariate; γ_i is the additive effect of scanner i on HRFs, δ_i is the multiplicative scanner effect, and ϵ_{ij} is an error term, presupposed to be normally distributed with zero mean. Following the estimation of effects, ComBat performs feature harmonization as follows:

$$Y_{ij}^{\text{ComBat}} = ((Y_{ij} - \alpha^{\wedge} - \beta^{\wedge} X_{ij} - \gamma_i^*) / (\delta_i^*)) + \alpha + \beta^{\wedge} X_{ij} \quad (2)$$

where α^{\wedge} and β^{\wedge} are estimators of the parameters α and β , respectively; and γ_i^* and δ_i^* are the empirical Bayes estimates for the parameters γ_i and δ_i , respectively [38]. ComBat harmonization was applied in this study using the slice thickness and pixel spacing values as the batch. For the slice thickness, the scans were divided into thin slices (< 3 mm) and thick slices (> 3 mm). For the pixel spacing, the scans were divided into three groups: between 0.5 and 0.7 mm, between 0.71 and 0.80 mm, and between 0.81 and 0.98 mm. The scans were acquired using similar convolution kernels in the majority of the scans, making the number of observations for the minority kernels not sufficient for ComBat application. The same applies when the medical center where the scans were acquired is considered as the batch.

2.4. Analysis Strategy and Pipeline

The workflow consisted of four major steps: the collection and curation of imaging dataset; the division of the data into three datasets as described in Section 2.1; the extraction of RFs; and the statistical analysis. Figure 1 describes the workflow of the study. Fifteen different experiments were performed, and the performance of the different approaches is compared accordingly.

Each dataset was analyzed using five different approaches. In the first approach, RFs extracted from the CT scans were directly used to assess overall survival (Ori). In the second approach, the clinical variable “percentage of tumor tissue necrosis” was used to assess overall survival (Necrosis) [39]. In the third approach, the percentage of tumor tissue necrosis and RFs were combined to assess overall survival (Ori_Necrosis). In the fourth approach, RFs were harmonized using ComBat harmonization, with the slice thickness being the batch, in order to assess overall survival. In the fifth approach, pixel spacing was used as the batch to harmonize RFs with ComBat, and the harmonized RFs were used to predict overall survival. To avoid reporting significant results by chance, the analysis was repeated 1000 times in each of the approaches. In each repeat, the data are split randomly into 70% training and 30% validation sets. Feature selection and cox regression are performed in each repeat. The concordance index (C-index) was used to evaluate the

performance of developed cox regression models [40]. The C-index of each of the final developed models on the validation set is kept in each repeat, and an average is calculated on the indices saved per approach. By performing the analysis 1000 times, we investigate whether the selected radiomic features and calculated performances are robust.

The set of features to be included in the modeling was contingent upon which of the five approaches (explained above in strategy) was utilized. In each of the repeats, following the division of data into training and validation sets, highly correlated features (identified using Spearman's correlation [41] with a cut-off of $R > 0.90$) were removed. If two RFs were found to be highly correlated, the RF that was removed was the one with higher average correlation with the other RFs. The remaining RFs were then used to build a cox regression model [42]. Backward feature selection [43] was applied on the generated cox model to select the important features. The selected features were then used to build the final cox regression model. Due to the small dataset size, if the number of the selected features was higher than 5, the 5 features with the most contributions to the model were used to build the final model. The model performance was then validated on the validation set using the C-index. Student *t*-test [44] was used to assess the difference on average between the different approaches used. All statistical were performed using R [45] language on RStudio (Version 2022.2.1.461, RStudio Inc., Boston, MA, USA) [46].

3. Results

3.1. Patient Characteristics

In this study, 179 out of 267 patients were included in the final analysis based on the availability of pre-surgical CT scans and survival data. The patients included had a median age of 59 years, 119 (66.5%) were male, and 109 (60.9%) were diagnosed at Stages I and II, with an average (\pm standard deviation (SD)) overall survival of 41 ± 24.7 months. The patients included received different treatment regimens, including immunotherapy, targeted molecular therapy, and chemotherapy.

3.2. Performance of Original RFs and Percentage of Tumor Tissue Necrosis

The results of the different analysis approaches varied slightly according to the data being analyzed and followed similar patterns (Figure 2). For the first approach, when Dataset 1 was analyzed, the performance of RFs (C-index) across the 1000 repeats ranged between 0.33 and 0.83, with an average (\pm SD) of 0.55 ± 0.08 . The RF "Intensity minimum" was selected in 91.8% of the times, followed by intensity skewness (77.3% of the times), intensity peak position (70.6%), and intensity kurtosis (64.8%). The remaining features were selected in less than 50% of the times.

In Dataset 2, the c index values across the 1000 repeats ranged between 0.07 and 0.98, with an average (\pm SD) of 0.50 ± 0.18 . None of the features was selected more than 50% of the times. Intensity minimum and intensity kurtosis were the most selected features (49.5 % and 46%, respectively).

In Dataset 3, the C-index values ranged between 0.31 and 0.95, with an average (\pm SD) of 0.62 ± 0.09 . None of the features was selected more than 50% of the times. Intensity

minimum and intensity kurtosis were the most selected features (49.5 % and 46%, respectively). Intensity minimum was the most selected RF (79.2%) followed by intensity skewness (61.4%) and intensity peak position (52.2%).

For the second approach, we assessed the ability of the clinical variables “percentage of tumor tissue necrosis” to stratify the patients in long and short survivors. The C-index values ranged (average \pm SD) between 0.20 and 0.82 (0.64 ± 0.07), 0.12 and 1.00 (0.58 ± 0.12), and 0.11 and 0.91 (0.67 ± 0.09) in Dataset 1, Dataset 2, and Dataset 3, respectively.

For the third approach, combining the percentage of tumor tissue necrosis with the RFs resulted in C-indices with an average (\pm SD) of 0.58 ± 0.07 , 0.52 ± 0.16 , 0.61 ± 0.09 in Dataset 1, Dataset 2 and Dataset 3 respectively.

T-tests showed that on average, the percentage of tumor tissue necrosis significantly (p -value < 0.05) outperformed the RFs-based models as well as the combination of RFs and the percentage of tumor tissue necrosis across all the three datasets.

3.3. Impact of ComBat Harmonization

For the fourth approach, the slice thickness was used as the batch (Figure 3), the C-index values ranged (average \pm SD) between 0.30 and 0.79 (0.55 ± 0.08) in Dataset 1, with intensity maximum being the most selected feature (94.8% of the times), followed by intensity peak position (84%) intensity minimum (79.2%), intensity skewness (59%), and intensity uniformity (51.4%).

In Dataset 2, the c index values range (average \pm SD) was 0.05–0.95 (0.43 ± 0.16), with none of the features being selected in more than half of the times. Intensity minimum was the most selected feature (43.7%) followed by intensity kurtosis (35.6%). In Dataset 3, the c index values ranged between 0.29 and 0.90 with an average (\pm SD) of 0.61 ± 0.09 . Intensity minimum was the most selected feature in this scenario (76%) followed by intensity peak position (68.2%). The remaining RFs were selected in less than 50% of the times.

Using the pixel spacing the batch (the fifth approach) resulted in C-index values ranges (mean \pm standard deviation) of 0.25–0.75 (0.51 ± 0.08), 0.05–0.97 (0.47 ± 0.17), and 0.23–0.87(0.59 ± 0.09) in Dataset 1, Dataset 2, and Dataset 3, respectively.

On average, the models based on the original RFs values before ComBat harmonization significantly (p -value < 0.05) outperformed the models based on harmonized RFs across all datasets, except for the models based on RFs harmonized based on the slice thickness in Dataset 1.

4. Discussion

In this study, we aimed to investigate (i) the potential of RFs to predict the overall survival in ccRCC, without adjusting for the differences in image acquisition and reconstruction parameters, using thorough statistical analyses; and (ii) the impact of ComBat harmonization on the performance of RFs. We found that on average, radiomic signatures had an above chance level (C-index > 0.5) for the prediction of overall survival in ccRCC patients.

Since in machine learning different data splits could result in significant variations in the performance of the model developed, we ran the analyses with thousand different splits and assessed the performance of radiomic signatures. It was observed that the chances of obtaining a C-index of 0.70 or higher across the data and its subsets was less than 5%, which falls within chance level. Moreover, different RFs were selected with different splits of the data. However, three intensity features were selected in at least 70% of the times (Supplementary Table S1). These RFs could be considered correlated with overall survival in ccRCC patients. Nonetheless, definitive conclusions can only be reached following the collection of more datapoints, as well as the availability of reproducibility analysis to assess the robustness of these RFs to the variations in imaging parameters observed across the data being analyzed [14].

Furthermore, we compared the performance of radiomic signatures to that of the percentage of tumor tissue necrosis, since the latter is a known clinical predictor [39,47]. Our results showed that on average, predictions based on the percentage of tumor tissue necrosis alone significantly outperformed those based on the different radiomic signatures. Nonetheless, the most frequently selected RFs (e.g., “intensity minimum”, “intensity skewness”, as shown in the supplement Table S1) were CT attenuation-related features, which more or less reflected some information about tumor tissue necrosis. We noticed that none of the predefined RFs was highly correlated to the percentage of tumor tissue necrosis. The findings that the clinical model only—on average—performed better, suggests that the RFs likely resulted in added noise to the survival model. Nevertheless, the comparison is not representative of the full potential of RFs, since in some instances the performance of the combination was higher, and the limitations of reproducibility have to be addressed before a final conclusion could be reached.

The data analyzed in this study included scans that were acquired with different hardware and different imaging parameters, such as convolution kernel, slice thickness, and pixel spacing. Variations in these parameters have been previously reported to affect the reproducibility of RFs. A number of studies investigated the effects of variations in convolution kernels on the reproducibility of RFs. All of these studies concluded that the variations in convolution kernel significantly affects the reproducibility of RFs [17,23–25]. Similarly, the effects of variations in slice thickness [23,48–50], pixel spacing [16,50], and effective mAs [51] were reported to have significant effects on the reproducibility of RFs. Since the data analyzed in this study included data acquired with different imaging protocols, it is expected that the reproducibility of the majority of RFs is affected significantly, which could explain the wide range of C-indices obtained across the thousand repeats in each approach. It could also explain the selection of intensity based RFs across all the investigated approaches, since these RFs are expected to be the least affected by the variations in our data. In addition, more sensitive RFs that decode texture were not frequently selected, which is in line with the expectations based on the literature.

We assessed the impact of ComBat harmonization on the performance of RFs using two approaches; one using the slice thickness, and the second using the pixel spacing to group the scans. We have not provided any biologic covariates to the calculations of ComBat estimates, since the aim of radiomics studies is to investigate clinical correlations. We

observed on average a decrease in the performance of RFs following ComBat harmonization in both of the approaches. The probability of obtaining a C-index of 0.70 or higher was also within chance level (<5%). When slice thickness was the batch used for ComBat harmonization, three intensity features were selected in 70% or more of the times. Two of these RFs were selected before and after harmonization. On the other hand, when pixel spacing was used as the batch, four intensity features were selected more than 70% of the times, one of which was selected before and after harmonization. In addition, *t*-test analyses showed that the performance of RFs before harmonization was significantly higher. This could be justified by a number of theories. First, the assumption of similar batch effects on the RFs is not met, which could affect the estimation of batch effects. Another possibility is that ComBat harmonization removed the biologic signals in the harmonization process. Regardless of the reason, there was no added value of performing ComBat harmonization in the reported manner in this dataset. This signifies the need for radiomics-specific harmonization methods.

While we designed statistically thorough experiments, a number of limitations are to be acknowledged. First, there was a lack of data to perform reproducibility analyses and preselect reproducible RFs, as well as the lack of a more homogenous dataset to compare the performances. Second, the relatively low number of datapoints analyzed, as well as the different treatment regimens used could adversely affect the performance of the signatures developed, especially on the arterial subset of the data. More datapoints are needed to reach more concrete conclusions. Third, the data were collected from different centers, which could introduce more variations due to the patients' characteristics and center protocols. The small numbers of patients per center did not allow for adjusting for these potential variations. In addition, given that the scans were resampled before feature extraction, the impact of ComBat harmonization could be impacted. Lastly, other factors that are known to affect the concordance of RFs, such as intra- and inter-observer variabilities need to be assessed in future studies.

5. Conclusions

In conclusion, the variations in the imaging parameters in the data under analysis significantly affects the performance of developed radiomic signatures. The application of ComBat harmonization in our data did not improve the performance of radiomic signatures, however reproducibility analyses may enhance the performance of ComBat by identifying specific RFs to utilize. Radiomics-specific harmonization techniques currently remain one of the major needs and challenges in the field of radiomics.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Funding:

Authors acknowledge financial support from the National Institutes of Health (U01 CA225431). The content is solely the responsibility of the authors and does not necessarily represent the views of the funding source.

Data Availability Statement:

The data is publicly available on [TCIA.org](https://wiki.cancerimagingarchive.net/display/Public/TCGA-KIRC): (<https://wiki.cancerimagingarchive.net/display/Public/TCGA-KIRC>, accessed on 3 April 2022).

References

1. Lambin P; Rios-Velazquez E; Leijenaar R; Carvalho S; van Stiphout RGPM; Granton P; Zegers CML; Gillies R; Boellard R; Dekker A; et al. Radiomics: Extracting More Information from Medical Images Using Advanced Feature Analysis. *Eur. J. Cancer* 2012, 48, 441–446. [PubMed: 22257792]
2. Gillies RJ; Kinahan PE; Hricak H Radiomics: Images Are More than Pictures, They Are Data. *Radiology* 2016, 278, 563–577. [PubMed: 26579733]
3. Linning E; Lu L; Li L; Yang H; Schwartz LH; Zhao B Radiomics for Classification of Lung Cancer Histological Subtypes Based on Nonenhanced Computed Tomography. *Acad. Radiol.* 2019, 26, 1245–1252. [PubMed: 30502076]
4. Yao Z; Dong Y; Wu G; Zhang Q; Yang D; Yu J-H; Wang W-P Preoperative Diagnosis and Prediction of Hepatocellular Carcinoma: Radiomics Analysis Based on Multi-Modal Ultrasound Images. *BMC Cancer* 2018, 18, 1089. [PubMed: 30419849]
5. Xu Y; Shu Z; Song G; Liu Y; Pang P; Wen X; Gong X The Role of Preoperative Computed Tomography Radiomics in Distinguishing Benign and Malignant Tumors of the Parotid Gland. *Front. Oncol.* 2021, 11, 634452. [PubMed: 33777789]
6. Granzier RWY; Ibrahim A; Primakov SP; Samiei S; van Nijnatten TJA; de Boer M; Heuts EM; Hulsmans F-J; Chatterjee A; Lambin P; et al. MRI-Based Radiomics Analysis for the Pretreatment Prediction of Pathologic Complete Tumor Response to Neoadjuvant Systemic Therapy in Breast Cancer Patients: A Multicenter Study. *Cancers* 2021, 13, 2447. [PubMed: 34070016]
7. Tharmalingam H; Tsang YM; Alonzi R; Beasley W; Taylor NJ; McWilliam A; Padhani A; Choudhury A; Hoskin PJ Changes in Magnetic Resonance Imaging Radiomic Features in Response to Androgen Deprivation Therapy in Patients with Intermediate- and High-Risk Prostate Cancer. *Clin. Oncol.* 2022, 34, e246–e253.
8. Dercle L; Lu L; Schwartz LH; Qian M; Tejpar S; Eggleton P; Zhao B; Piessevaux H Radiomics Response Signature for Identification of Metastatic Colorectal Cancer Sensitive to Therapies Targeting EGFR Pathway. *J. Natl. Cancer Inst.* 2020, 112, 902–912. [PubMed: 32016387]
9. Li Q; Bai H; Chen Y; Sun Q; Liu L; Zhou S; Wang G; Liang C; Li Z-C A Fully-Automatic Multiparametric Radiomics Model: Towards Reproducible and Prognostic Imaging Signature for Prediction of Overall Survival in Glioblastoma Multiforme. *Sci. Rep.* 2017, 7, 14331. [PubMed: 29085044]
10. Larue RTHM; Klaassen R; Jochems A; Leijenaar RTH; Hulshof MCCM; van Berge Henegouwen MI; Schreurs WMJ; Sosef MN; van Elmpt W; van Laarhoven HWM; et al. Pre-Treatment CT Radiomics to Predict 3-Year Overall Survival Following Chemoradiotherapy of Esophageal Cancer. *Acta Oncol.* 2018, 57, 1475–1481. [PubMed: 30067421]
11. Yip SSF; Aerts HJWL Applications and Limitations of Radiomics. *Phys. Med. Biol.* 2016, 61, R150–R166. [PubMed: 27269645]
12. Ibrahim A; Primakov S; Beuque M; Woodruff HC; Halilaj I; Wu G; Refaee T; Granzier R; Widaatalla Y; Hustinx R; et al. Radiomics for Precision Medicine: Current Challenges, future Prospects, and the Proposal of a New Framework. *Methods* 2020, 188, 20–29. [PubMed: 32504782]
13. Strimbu K; Tavel JA What Are Biomarkers? *Curr. Opin. HIV AIDS* 2010, 5, 463. [PubMed: 20978388]
14. Ibrahim A; Refaee T; Leijenaar RTH; Primakov S; Hustinx R; Mottaghy FM; Woodruff HC; Maidment ADA; Lambin P The Application of a Workflow Integrating the Variable Reproducibility and Harmonizability of Radiomic Features on a Phantom Dataset. *PLoS ONE* 2021, 16, e0251147. [PubMed: 33961646]

15. Jha AK; Mithun S; Jaiswar V; Sherkhane UB; Purandare NC; Prabhash K; Rangarajan V; Dekker A; Wee L; Traverso A Repeatability and Reproducibility Study of Radiomic Features on a Phantom and Human Cohort. *Sci. Rep.* 2021, 11, 2055. [PubMed: 33479392]
16. Ibrahim A; Refaee T; Primakov S; Barufaldi B; Acciavatti RJ; Granzier RWY; Hustinx R; Mottaghy FM; Woodruff HC; Wildberger JE; et al. The Effects of In-Plane Spatial Resolution on CT-Based Radiomic Features' Stability with and without ComBat Harmonization. *Cancers* 2021, 13, 1848. [PubMed: 33924382]
17. Meyer M; Ronald J; Vernuccio F; Nelson RC; Ramirez-Giraldo JC; Solomon J; Patel BN; Samei E; Marin D Reproducibility of CT Radiomic Features within the Same Patient: Influence of Radiation Dose and CT Reconstruction Settings. *Radiology* 2019, 293, 583–591. [PubMed: 31573400]
18. Zhao B Understanding Sources of Variation to Improve the Reproducibility of Radiomics. *Front. Oncol.* 2021, 11, 633176. [PubMed: 33854969]
19. Yamashita R; Perrin T; Chakraborty J; Chou JF; Horvat N; Koszalka MA; Midya A; Gonen M; Allen P; Jarnagin WR; et al. Radiomic Feature Reproducibility in Contrast-Enhanced CT of the Pancreas Is Affected by Variabilities in Scan Parameters and Manual Segmentation. *Eur. Radiol.* 2020, 30, 195–205. [PubMed: 31392481]
20. Mali SA; Ibrahim A; Woodruff HC; Andrearczyk V; Müller H; Primakov S; Salahuddin Z; Chatterjee A; Lambin P Making Radiomics More Reproducible across Scanner and Imaging Protocol Variations: A Review of Harmonization Methods. *J. Pers. Med.* 2021, 11, 842. [PubMed: 34575619]
21. Li Y; Ammari S; Balleyguier C; Lassau N; Chouzenoux E Impact of Preprocessing and Harmonization Methods on the Removal of Scanner Effects in Brain MRI Radiomic Features. *Cancers* 2021, 13, 3000. [PubMed: 34203896]
22. Traverso A; Kazmierski M; Shi Z; Kalendralis P; Welch M; Nissen HD; Jaffray D; Dekker A; Wee L Stability of Radiomic Features of Apparent Diffusion Coefficient (ADC) Maps for Locally Advanced Rectal Cancer in Response to Image Pre-Processing. *Phys. Med.* 2019, 61, 44–51. [PubMed: 31151578]
23. He L; Huang Y; Ma Z; Liang C; Liang C; Liu Z Effects of Contrast-Enhancement, Reconstruction Slice Thickness and Convolution Kernel on the Diagnostic Performance of Radiomics Signature in Solitary Pulmonary Nodule. *Sci. Rep.* 2016, 6, 34921. [PubMed: 27721474]
24. Denzler S; Vuong D; Bogowicz M; Pavic M; Frauenfelder T; Thierstein S; Eboulet EI; Maurer B; Schniering J; Gabry HS; et al. Impact of CT Convolution Kernel on Robustness of Radiomic Features for Different Lung Diseases and Tissue Types. *Br. J. Radiol.* 2021, 94, 20200947. [PubMed: 33544646]
25. Refaee T; Salahuddin Z; Widaatalla Y; Primakov S; Woodruff HC; Hustinx R; Mottaghy FM; Ibrahim A; Lambin P CT Reconstruction Kernels and the Effect of Pre- and Post-Processing on the Reproducibility of Handcrafted Radiomic Features. *J. Pers. Med.* 2022, 12, 553. [PubMed: 35455668]
26. Altazi BA; Zhang GG; Fernandez DC; Montejo ME; Hunt D; Werner J; Biagioli MC; Moros EG Reproducibility of F18-FDG PET Radiomic Features for Different Cervical Tumor Segmentation Methods, Gray-Level Discretization, and Reconstruction Algorithms. *J. Appl. Clin. Med. Phys.* 2017, 18, 32–48.
27. Larue RTHM; van Timmeren JE; de Jong EEC; Feliciani G; Leijenaar RTH; Schreurs WMJ; Sosef MN; Raat FHPJ; van der Zande FHR; Das M; et al. Influence of Gray Level Discretization on Radiomic Feature Stability for Different CT Scanners, Tube Currents and Slice Thicknesses: A Comprehensive Phantom Study. *Acta Oncol.* 2017, 56, 1544–1553. [PubMed: 28885084]
28. Schwier M; van Griethuysen J; Vangel MG; Pieper S; Peled S; Tempany C; Aerts HJWL; Kikinis R; Fennessy FM; Fedorov A Repeatability of Multiparametric Prostate MRI Radiomics Features. *Sci. Rep.* 2019, 9, 9441. [PubMed: 31263116]
29. Johnson WE; Li C; Rabinovic A Adjusting Batch Effects in Microarray Expression Data Using Empirical Bayes Methods. *Biostatistics* 2007, 8, 118–127. [PubMed: 16632515]
30. Lu L; Ahmed FS; Akin O; Luk L; Guo X; Yang H; Yoon J; Hakimi AA; Schwartz LH; Zhao B Uncontrolled Confounders May Lead to False or Overvalued Radiomics Signature: A Proof of Concept Using Survival Analysis in a Multicenter Cohort of Kidney Cancer. *Front. Oncol.* 2021, 11, 638185. [PubMed: 34123789]

31. Cancer Genome Atlas Research Network Comprehensive Molecular Characterization of Clear Cell Renal Cell Carcinoma. *Nature* 2013, 499, 43–49. [PubMed: 23792563]
32. Zhang G; Wu Y; Zhang J; Fang Z; Liu Z; Xu Z; Fan Y Nomograms for Predicting Long-Term Overall Survival and Disease-Specific Survival of Patients with Clear Cell Renal Cell Carcinoma. *Onco. Targets. Ther.* 2018, 11, 5535–5544. [PubMed: 30233214]
33. Khodabakhshi Z; Amini M; Mostafaei S; Haddadi Avval A; Nazari M; Oveisi M; Shiri I; Zaidi H Overall Survival Prediction in Renal Cell Carcinoma Patients Using Computed Tomography Radiomic and Clinical Information. *J. Digit. Imaging* 2021, 34, 1086–1098. [PubMed: 34382117]
34. Nazari M; Shiri I; Zaidi H Radiomics-Based Machine Learning Model to Predict Risk of Death within 5-Years in Clear Cell Renal Cell Carcinoma Patients. *Comput. Biol. Med.* 2021, 129, 104135. [PubMed: 33254045]
35. Gao R; Qin H; Lin P; Ma C; Li C; Wen R; Huang J; Wan D; Wen D; Liang Y; et al. Development and Validation of a Radiomic Nomogram for Predicting the Prognosis of Kidney Renal Clear Cell Carcinoma. *Front. Oncol.* 2021, 11, 613668. [PubMed: 34295804]
36. 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. [PubMed: 23884657]
37. Kirby J TCGA-KIRC. Available online: <https://wiki.cancerimagingarchive.net/display/Public/TCGA-KIRC> (accessed on 10 June 2022).
38. Fortin J-P; Parker D; Tunç B; Watanabe T; Elliott MA; Ruparel K; Roalf DR; Satterthwaite TD; Gur RC; Gur RE; et al. Harmonization of Multi-Site Diffusion Tensor Imaging Data. *Neuroimage* 2017, 161, 149–170. [PubMed: 28826946]
39. Ahmed FS; Akin O; Shaish H; Luk L; Guo X; Yang H; Zabor E; Ostrovnya I; Hakimi AA; Zhao B; et al. Nonenhancing Component of Clear Cell Renal Cell Carcinoma on Computed Tomography Correlates with Tumor Necrosis and Stage and Serves as a Size-Independent Prognostic Biomarker. *J. Comput. Assist. Tomogr.* 2019, 43, 628–633. [PubMed: 31162237]
40. Lawless JF; Singhal K Efficient Screening of Nonnormal Regression Models. *Biometrics* 1978, 34, 318–327.
41. Zar JH Spearman Rank Correlation. In *Encyclopedia of Biostatistics*; Wiley Online Library: Hoboken, NJ, USA, 2005.
42. Smith L Cox Regression Model; Louisiana State University: Baton Rouge, LA, USA, 2004.
43. Schemper M; Stare J Explained Variation in Survival Analysis. *Stat. Med.* 1996, 15, 1999–2012. [PubMed: 8896135]
44. Zimmerman DW; Zumbo BD Rank Transformations and the Power of the Student T Test and Welch T test for Non-Normal Populations with Unequal Variances. *Can. J. Exp. Psychol.* 1993, 47, 523–539.
45. Team, R.C. R Language Definition; R foundation for statistical computing: Vienna, Austria, 2000.
46. Gandrud C Reproducible Research with R and R Studio; CRC Press: Boca Raton, FL, USA, 2013; ISBN 9781466572843.
47. Zigeuner R; Hutterer G; Chromecki T; Kampel-Kettner K; Imamovic A; Langner C; Rehak P; Pummer K The Mayo Clinic Stage, Size, Grade and Necrosis (ssign) Score for Clear Cell Renal Cell Carcinoma: External Validation by a Single European Centre Applying. *Eur. Urol. Suppl.* 2008, 3, 280.
48. Zhao B; Tan Y; Tsai W-Y; Qi J; Xie C; Lu L; Schwartz LH Reproducibility of Radiomics for Deciphering Tumor Phenotype with Imaging. *Sci. Rep.* 2016, 6, 23428. [PubMed: 27009765]
49. Shafiq-ul-Hassan M; Zhang GG; Latifi K Intrinsic Dependencies of CT Radiomic Features on Voxel Size and Number of Gray Levels. *Medical* 2017, 44, 1050–1062.
50. Shafiq-Ul-Hassan M; Latifi K; Zhang G; Ullah G; Gillies R; Moros E Voxel Size and Gray Level Normalization of CT Radiomic Features in Lung Cancer. *Sci. Rep.* 2018, 8, 10545. [PubMed: 30002441]
51. Zhovannik I; Bussink J; Traverso A; Shi Z; Kalendralis P; Wee L; Dekker A; Fijten R; Monshouwer R Learning from Scanners: Bias Reduction and Feature Correction in Radiomics. *Clin. Transl. Radiat. Oncol.* 2019, 19, 33–38. [PubMed: 31417963]

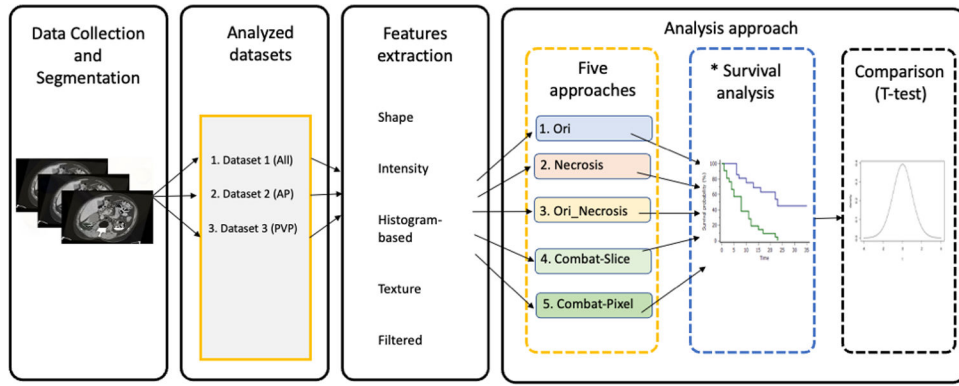


Figure 1. Study workflow. * The survival analysis is performed 1000 analyses per approach.

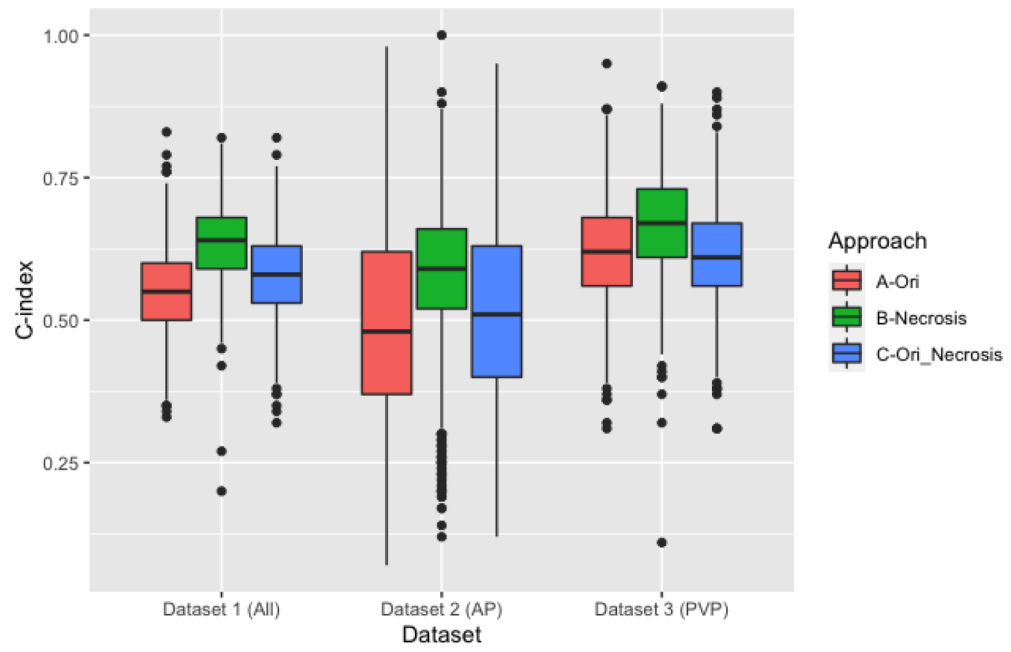


Figure 2. Boxplot comparison of the C-indices of the clinical and original RF models.

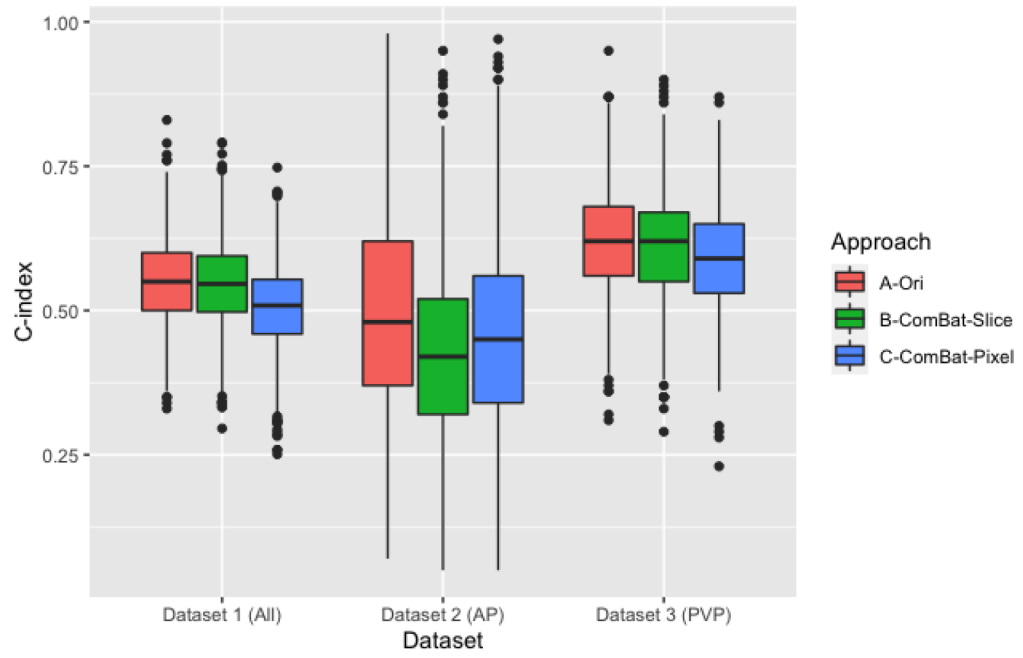


Figure 3. Boxplot comparison of the C-indices before and after ComBat harmonization.

Table 1.

Description of the data included.

Vendor	Number of Scans	Convolution Kernels	Slice Thickness (mm)	Pixel Spacing (mm ²)
GE	154	Standard, Soft	1.25–7.5	$0.7 \times 0.7 - 0.78 \times 0.78$
Philips	3	A, B	5	$0.74 \times 0.74 - 0.86 \times 0.86$
Siemens	22	B30f, B30s, B31f, B31s	3, 5	$0.54 \times 0.54 - 0.98 \times 0.98$