

Comparison of radiomic pre-processing steps in the reproducible prediction of disease free survival across multi-scanners/centers

Marta Ferreira (✉ martaferreira22cloud@gmail.com)

University of Liege: Universite de Liege <https://orcid.org/0000-0003-4532-6744>

Pierre Lovinfosse

CHU de Liège: Centre hospitalier universitaire de Liege

Johanne Hermesse

CHU de Liège: Centre hospitalier universitaire de Liege

Marjolein Decuyper

CHU de Liège: Centre hospitalier universitaire de Liege

Caroline Rousseau

Université de Nantes: Universite de Nantes

François Lucia

Centre Hospitalier Universitaire de Brest: CHRU de Brest

Ulrike Schick

Centre Hospitalier Universitaire de Brest: CHRU de Brest

Caroline Reinhold

McGill University Health Centre

Philippe Robin

Centre Hospitalier Universitaire de Brest: CHRU de Brest

Mathieu Hatt

Université de Brest: Université de Bretagne Occidentale

Dimitris Visvikis

Université de Brest: Université de Bretagne Occidentale

Claire Bernard

Central University Hospital of Liege: Centre hospitalier universitaire de Liege

Ralph T.H. Leijenaar

Maastricht University Hospital: Maastricht Universitair Medisch Centrum+

Frédéric Kridelka

Central University Hospital of Liege: Centre hospitalier universitaire de Liege

Philippe Lambin

Maastricht University Hospital: Maastricht Universitair Medisch Centrum+

Roland Hustinx

University of Liege: Universite de Liege

Patrick E. Meyer

Liege University: Universite de Liege

Methodology

Keywords: Radiomics, Intensity discretization, Pre-processing radiomics steps

Posted Date: September 8th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-875843/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background

Features reproducibility and the generalizability of the models are currently among the most important limitations when integrating radiomics into the clinics. Radiomic features are sensitive to imaging acquisition protocols, reconstruction algorithms and parameters, as well as by the different steps of the usual radiomics workflow. We propose a framework for comparing the reproducibility of different pre-processing steps in PET/CT radiomic analysis in the prediction of disease free survival (DFS) across multi-scanners/centers.

Results

We evaluated and compared the prediction performance of several models that differ in i) the type of intensity discretization, ii) feature selection method, iii) features type i.e. original or tumour to liver ratio radiomic features (OR or TLR). We trained our models using data from one scanner/center and tested on two external scanner/centers. Our results show that there is a low reproducibility in predictions across scanners and discretization methods. Despite of this, TLR based models were generally more robust than OR. Maximum relevance minimum redundancy (MRMR) forward feature selection with Pearson correlation was the feature selection method that had the best mean area under the precision recall curve when using it combining the features from all discretization's bin's number (D_All_FBN) with TLR features for two of the four classifiers.

Conclusion

We evaluated and compared the prediction performance of several models in a data set containing hundred fifty-eight patients with locally advanced cervical cancer (LACC) from three distinct scanners. In our cohort of LAAC patients pre-processing of radiomic features in [¹⁸F]FDG PET affects DFS predictions performances across scanners and combining the D_All_FBN TLR approach with the MRMR forward Pearson feature selection method might help increasing robustness of radiomic studies.

Background

Radiomics consists of characterizing tumour phenotypes via the extraction of high-dimensional quantitative features from medical images, with the aim of supporting clinical decision-making [1–3]. Radiomic features have been used to characterize cancer subtypes and aggressiveness or to predict the response to treatment [4–8] and have increasingly been combined with machine learning (ML) techniques in order to predict a specific clinical outcome [5–6][9–13].

The challenges facing the integration of radiomics into the clinics are many. Features reproducibility and the generalizability of the models are currently among the most important limitations. Even though standardization guidelines have helped to mitigate some of these reproducibility challenges [14], radiomic features are sensitive to imaging acquisition protocols and reconstruction algorithms and parameters, but can also be affected by the different steps of the usual radiomics workflow [15–19]. 2-[¹⁸F]fluoro-2-deoxy-D-glucose ([¹⁸F]FDG) positron emission tomography combined with computed tomography (PET/CT) imaging is especially prone to reproducibility issues due to frequent variations in pre-acquisition settings and scanner properties [20], despite standardization efforts of acquisition and reconstruction protocols within the context of multicenter trials [21].

The intensity discretization scheme is one of the steps in radiomics workflow known to affect models reproducibility. Most of radiomic studies assume a specific discretization method based on the results of previous studies [5][20]. Additionally, some studies used phantoms or internal data sets and evaluated the impact of some of the pre-processing steps using test-retest analysis [18][22].

In this study, we compared the effect of 6 [¹⁸F]FDG PET intensity discretization methods on the prediction of disease free survival (DFS) in locally advanced cervical cancer (LACC) patients from 3 different scanners/centers. Moreover, we evaluated the effect of combining these different discretization approaches with 7 distinct feature selection (FS) methods as well as with the effect of feature transformation using tumour to liver ratio (TLR) features as done in our previous work [23]. In contrast to what was done in our previous work, we trained 84 distinct models using the data of one of the scanners and evaluated these models on the remaining two. We further evaluated these different steps using 4 different classifiers in order to better enhance the study robustness.

Materials And Methods

Data

One hundred and fifty-eight patients with LACC imaged between 2010 and 2016, in three different scanners were included in this retrospective study. PET/CT studies were performed in the CHU of Liège, where 89 studies were acquired using a Philips Gemini TF or BB (scanner A), in the CHU of Brest and ICO St Herblain, where 17 and 34, respectively, were acquired using a Siemens Biograph mCT (scanner B) and at the McGill University Health Center, where 18 studies were performed with a General Electric Discovery ST (scanner C). The patient's clinical characteristics, treatment, acquisition and reconstruction protocols are described in our previous study [23].

Experimental design

DFS was dichotomized into a binary endpoint, i.e. recurrence or no recurrence, independently of the time-to-event. Next we compared the prediction performance of 84 different models, which differed in i) the type of intensity discretization used before feature calculation, ii) the feature selection method iii)

the features type, i.e., original radiomics (OR) or TLR radiomics.

Models were trained on the data from scanner A and then evaluated on data from two remaining external scanners independently. The metric used to evaluate model performance was the area under the curve of the precision recall curve (AUCpr) (Fig. 1).

Statistical and ML analyses were performed using R software, version 4.0.1.

Segmentation and interpolation

PET images from scanners B and C were interpolated using the research toolbox (Oncoradiomics SA, Liège, Belgium), up-sampling or down-sampling the images using a linear method, so that all datasets had isotropic voxels of $4 \times 4 \times 4 \text{ mm}^3$ (i.e., the voxel size in images of scanner A). The 3D primary tumour volumes were segmented in the [^{18}F]FDG PET images using the 2 classes semi-automatic Fuzzy Local Adaptive Bayesian algorithm [24]. Volumes of 20 cm^3 in the liver were manually drawn in order to investigate the predictive value of TLR radiomic features, as explained below. All segmentations were reviewed and edited if needed by one nuclear medicine physician with 9 years of experience in clinical PET/CT.

Images Radiomic features

We extracted two hundred and fifteen features from the segmented volumes, which included first order grey level statistics, geometry, fractals, texture matrix based features and others. Features were extracted using the Oncoradiomics research toolbox and their detailed description can be found in supplementary data of our previous study [23]. All features were calculated according to the Imaging biomarkers standardization initiative (IBSI) [25]. We also studied the ratio of the features values calculated in the tumour and in the liver (TLR versions of features), except for the shape features as done in our previous study [23]. We hypothesized that TLR features may reduce the variability of radioactive dose uptake within the different patients and across centers by normalizing the radiomic features using the liver which is an organ with an homogenous and reproducible uptake. There were no missing data for any patient.

Radiomic features intensity discretization

Image intensities were discretized using the two schemes currently standardized by the IBSI: fixed bin number (FBN, with 32 and 64 bins) and fixed bin width (FBW, with 4 different widths of 0.05, 0.1, 0.2 and 0.5 Standardized Uptake Values) [25]. These two sets of features were considered either alone or by: 1) joining the features from all discretization's widths/bin's number (D_All_FBW, D_All_FBN), 2) combining the four discretization's widths from the FBW discretization method or combining the two number of bins from FBN through the calculation of their median value (D_Med_FBN, D_Med_FBW).

Features selection, classifiers and model selection

We applied 7 different FS methods to identify the 5 most relevant features: 1-Accuracy decrease obtained from the embedded FS of the random forest (RF) classifier; 2- Gini impurity decrease obtained from the embedded FS of the RF classifier; 3- forward FS using maximum relevance minimum redundancy (MRMR) method with Pearson correlation; 4- backward FS using MRMR with Pearson correlation; 5- forward FS using MRMR with Spearman correlation; 6- backward FS using MRMR with Spearman correlation; 7- forward MRMR based on the mutual information (MI). We also considered 4 ML classifiers: RF, support vector machine (SVM) with radial kernel, Naïve Bayes (NB) and a logistic regression (LR) [26–28]. We used for each classifier the default hyperparameters values in their respective R packages. We used 5-fold cross-validation in the training data to internally validate and select the models with better predictions for each classifier independently. Additionally, models were trained using all the training data then tested in the two external data sets.

A paired Wilcoxon Rank Sum test was used to test whether the predictions for each discretization scheme were statistically significantly different from each other in the two external validation schemes. Wilcoxon Rank Sum tests were considered significant if $p < 0.05$. Holm-Bonferroni correction method was used to correct for multiple hypothesis testing.

Results

Table 1 depicts the mean AUCpr between the three validation schemes, i.e. i) Internal validation using 5-fold cross validation using scanner A ii) external validation using scanner B iii) external validation using scanner C. The table shows the mean AUCpr of the models using RF, SVM, LR and NB classifier using the different FS methods applied to the OR and TLR features. Additionally, the results shown for the standard FBW and FBN discretization schemes correspond to the model with discretization width/bin number that had a better AUCpr in the internal validation scheme. The AUCpr of the three validation schemes individually are shown in the supplementary material.

Our results showed a low reproducibility between scanners. The discretization scheme that showed the higher AUCpr in the validation scheme was D_Med_FBN combined with TLR features. This was not the case for the two external scanners. (Supplementary material).

Table 1
MEAN AUCPR BETWEEN VALIDATION SCHEMES

	FBW OR	FBN OR	D_Med_FBW OR	D_Med_FBN OR	D_Ail_FBW OR	D_Ail_FBN OR	FBW TLR	FBN TLR	D_Med_FBW TLR	D_Med_FBN TLR	I T
RF + MRMR_Forward_pearson	0.51	0.46	0.44	0.48	0.52	0.46	0.44	0.55	0.44	0.52	C
RF + MRMR_Backward_pearson	0.45	0.38	0.45	0.42	0.43	0.39	0.49	0.48	0.44	0.44	C
RF + MRMR_Forward_spearman	0.46	0.50	0.42	0.57	0.43	0.47	0.50	0.47	0.45	0.57	C
RF + MRMR_Backward_spearman	0.39	0.47	0.39	0.42	0.46	0.43	0.51	0.55	0.49	0.51	C
RF + MRMR_MI	0.46	0.44	0.47	0.45	0.46	0.40	0.46	0.51	0.45	0.52	C
RF + RF_Accuracy	0.52	0.47	0.54	0.55	0.53	0.49	0.53	0.56	0.54	0.58	C
RF + RF_Gini	0.47	0.49	0.42	0.50	0.51	0.48	0.52	0.56	0.53	0.52	C
SVM + MRMR_Forward_pearson	0.48	0.49	0.44	0.47	0.52	0.49	0.60	0.51	0.44	0.57	C
SVM + MRMR_Backward_pearson	0.37	0.41	0.37	0.45	0.41	0.43	0.58	0.48	0.42	0.42	C
SVM + MRMR_Forward_spearman	0.44	0.40	0.42	0.52	0.52	0.40	0.48	0.43	0.42	0.43	C
SVM + MRMR_Backward_spearman	0.38	0.46	0.35	0.46	0.40	0.38	0.42	0.47	0.41	0.36	C
SVM + MRMR_MI	0.38	0.39	0.47	0.33	0.43	0.37	0.46	0.48	0.45	0.52	C
SVM + RF_Accuracy	0.52	0.47	0.39	0.44	0.49	0.42	0.51	0.52	0.50	0.51	C
SVM + RF_Gini	0.56	0.51	0.35	0.45	0.56	0.45	0.51	0.50	0.55	0.44	C
LR + MRMR_Forward_pearson	0.49	0.51	0.48	0.48	0.48	0.51	0.55	0.57	0.51	0.51	C
LR + MRMR_Backward_pearson	0.47	0.46	0.47	0.47	0.50	0.47	0.50	0.50	0.48	0.41	C
LR + MRMR_Forward_spearman	0.34	0.48	0.34	0.52	0.38	0.43	0.43	0.51	0.38	0.50	C
LR + MRMR_Backward_spearman	0.43	0.48	0.39	0.48	0.39	0.40	0.48	0.47	0.45	0.56	C
LR + MRMR_MI	0.39	0.44	0.45	0.51	0.41	0.46	0.42	0.48	0.43	0.53	C
LR + RF_Accuracy	0.53	0.48	0.53	0.50	0.50	0.46	0.51	0.57	0.50	0.51	C
LR + RF_Gini	0.50	0.50	0.51	0.49	0.52	0.46	0.48	0.55	0.50	0.52	C
NB + MRMR_Forward_pearson	0.45	0.50	0.44	0.57	0.47	0.50	0.52	0.50	0.47	0.50	C
NB + MRMR_Backward_pearson	0.44	0.49	0.43	0.43	0.50	0.46	0.48	0.49	0.42	0.38	C
NB + MRMR_Forward_spearman	0.40	0.49	0.35	0.52	0.42	0.43	0.44	0.49	0.46	0.53	C
NB + MRMR_Backward_spearman	0.43	0.51	0.42	0.46	0.37	0.46	0.45	0.49	0.41	0.52	C
NB + MRMR_MI	0.45	0.51	0.41	0.52	0.41	0.53	0.43	0.46	0.42	0.48	C

Mean AUCPr between the three validation schemes using the four classifiers (RF, SVM, LR and NB) and the seven FS methods represented in the columns and discretization schemes in the rows. The features discretization were applied to the OR and TLR features.

	FBW OR	FBN OR	D_Med_FBW OR	D_Med_FBN OR	D_All_FBW OR	D_All_FBN OR	FBW TLR	FBN TLR	D_Med_FBW TLR	D_Med_FBN TLR	D_All_FBW TLR	D_All_FBN TLR
NB + RF_Accuracy	0.52	0.56	0.49	0.55	0.49	0.51	0.48	0.56	0.52	0.47	0.47	0.47
NB + RF_Gini	0.54	0.51	0.52	0.54	0.55	0.46	0.47	0.46	0.42	0.46	0.46	0.46

Mean AUCpr between the three validation schemes using the four classifiers (RF, SVM, LR and NB) and the seven FS methods represented in the columns and discretization schemes in the rows. The features discretization were applied to the OR and TLR features.

Despite the models low reproducibility, D_All_FBN with TLR features was the model with the better mean AUCpr for the LR and NB classifier (0.57 and 0.58 respectively). It was also the second model with overall higher AUCpr in the two independent scanners with AUCpr of 0.45 and 0.7 in scanner B and C respectively (Fig. 2). For the RF classifier the model with higher AUCpr was D_Med_FBN TLR whereas for SVM it was FBW TLR. Regarding the FS method, when combined with the D_All_FBN TLR, MRMR Forward with Pearson correlation was the optimal FS method for at least one of the four classifiers in all validation schemes. MRMR Forward with Pearson correlation is also the FS method that showed better mean AUCpr for 4 out of the 6 discretization schemes when using TLR features. The discretization schemes that showed higher mean AUCpr across classifiers were D_Med_FBN, D_All_FBW and D_All_FBN. All of them were used with TLR features (Fig. 3). Despite of the good performance of D_All_FBN with TLR, the only discretization methods that showed to be statistically significant from D_All_FBN TLR were FBW with OR features, and FBW, FBN, D_Med_FBW, D_Med_FBN, D_All_FBW with TLR features. FBW combined with OR features was the only discretization scheme statistically different from all the others (Table 2). TLR based models had higher mean AUCpr values than OR based models for all the classifiers and discretization schemes, except when using D_med_FBW with LR or FBW and D_Med_FBN with NB (Fig. 3). TLR based models predictions in the two external validation schemes were statistically significant from all the OR based models (p-value < 0.05).

Table 2

Wilcoxon rank sum test corrected p-values showing the statistical significance between discretization schemes

	FBW_OR	FBN_OR	D_Med_FBW_OR	D_Med_FBN_OR	D_All_FBW_OR	D_All_FBN_OR	FBW_TLR	FBN_TLR	D_Med_FBW_TLR	D_Med_FBN_TLR	D_All_FBW_TLR	D_All_FBN_TLR
FBW_OR												
FBN_OR	2.11E-11											
D_Med_FBW_OR	5.22E-07	1.48E+01										
D_Med_FBN_OR	5.43E-13	4.30E+00	2.95E-03									
D_All_FBW_OR	5.63E-09	4.59E+01	7.51E+00	3.30E+01								
D_All_FBN_OR	2.07E-04	2.87E+02	1.86E+01	6.37E-05	1.26E-01							
FBW_TLR	8.77E-50	3.24E-14	2.01E-37	1.32E-11	1.93E-18	1.60E-25						
FBN_TLR	4.41E-25	7.47E-03	3.71E-15	9.63E-03	2.87E-08	2.26E-15	9.22E-07					
D_Med_FBW_TLR	3.33E-44	5.81E-10	1.61E-28	1.62E-06	1.58E-15	2.15E-21	5.18E-02	2.05E-01				
D_Med_FBN_TLR	2.70E-40	7.49E-24	8.82E-29	4.52E-18	1.37E-16	4.89E-36	6.43E+01	1.41E-09	6.97E+00			
D_All_FBW_TLR	1.83E-25	2.94E-13	4.28E-16	1.18E-12	1.16E-12	1.25E-13	4.97E+00	5.07E-03	5.33E+01	1.71E+00		
D_All_FBN_TLR	1.51E-04	6.42E+01	7.73E+00	3.18E+01	3.18E+01	2.18E+01	2.60E-23	2.26E-09	8.29E-16	4.13E-24	4.20E-10	

Discussion

Radiomics aims at converting data from medical images into quantitative features providing valuable information regarding the clinical management of patients. These features can be combined with statistical/ML methods in order to derive predictive models of clinically relevant endpoints. The radiomics workflow consists however of multiple pre-processing steps that can affect the radiomic features values and therefore their clinical relevance. In this study we

compared the prediction performance of numerous models that differed in their discretization and FS methods. This comparison was carried out within the context of predicting DFS from [¹⁸F]FDG PET images in a multi-scanner/center [¹⁸F]FDG PET cohort of LACC patients. Additionally, we investigated the effect of features transformation using features ratios with an organ of reference and we combined it with the previous pre-processing steps. We also compared the models performances using four classifiers. Multi-classifier radiomics predictive models, ensemble classifiers or the combination of different classifiers performances to measure feature importance, consistently tend to outperform traditional single classifier approaches [29–31]. Moreover, the choice of classification method is one the most dominant sources of performance variation in radiomics studies [12]. Due to this, we believe that comparing the results of multiple classifiers is needed when evaluating the robustness of the workflow. The discretization scheme is one of the factors that affect radiomic features reproducibility. FBW discretization in PET has been recommended [18][25], although some studies have also reported more favourable properties using FBN [32]. This is related to the fact that FBN and FBW have different drawbacks and advantages. FBW preserves the relationship between PET units and the corresponding physical substrate, contrary to arbitrary units (such as in some non-quantitative magnetic resonance imaging sequences). FBN on the other hand does not preserve such relationship but introduces a normalization effect that can be favourable when contrast is considered important or when the actual original image intensity value does not have a ‘meaning’. In our study D_All_FBN and FBW both combined with TLR features were the discretization scheme that showed the best AUCpr in the two external scanners. Combining features discretized with different widths/bin numbers can as shown in our study introduce complementary information and be a more reproducible and simple strategy as it also avoids the uncertain assumption or the extensive search of the optimal feature discretization width/bin number. Combining feature discretization schemes has also been done in previous studies [33–34]. Furthermore, our results show that when using D_All_FBN with TLR the FS scheme with higher mean AUCpr in the three validation schemes is MRMR Forward with Pearson correlation for 2 of the 4 classifiers. FS is an effective strategy to improve radiomics-based predictive studies. Different FS strategies are used in radiomics studies, each with his pros and cons, and some known to work better with certain type of features or classifiers [35].

Finally, we also evaluated the feature type, i.e. OR and TLR radiomics. We have shown in our previous study that using the ratio of the tumour features with a reference organ (TLR radiomics) improves the predictive performance of radiomics model in LACC. In contrast to our previous study, we trained our models using only data from one clinical center/scanner and evaluated our models in two external scanners. We emphasize the conclusions of our previous study, by observing that all of the most robust models used TLR features instead of OR. This can be caused by a normalizing effect of the SUVs on each patient. The importance of data normalization/transformation has also been accessed for other radiomic studies and shown to improve models performances [36–38]. Moreover, feature transformation using an organ of reference has also been investigated by other authors, leading to normalized images and increased reproducibility of radiomic features [39–40].

The results of our study are encouraging and can potentially be used as a first recommendation approach to improve reproducibility of radiomics studies across multi-scanners/centers in LACC. However, it corresponds to a preliminary study and we still observed performance differences between the 3 scanners: for some scanners some models work quite well while for other scanners other models work better. This could be caused by the variation in scanner properties, the different number of patients in each scanner, or the variation in tumour recurrence rate for each population.

In future work, the tumour segmentations should be done fully automatically using a recently validated approach [41], instead of being done by a single observer in a semi-automated way, since segmentation variations within patients can affect radiomics reproducibility [20]. Within this work, we also did not try combining all the features from the two discretization schemes, mainly due to the resulting need for longer computation times. Combining discretization methods, exploring more classifiers, combining the information from PET with other image modalities and explore image or matrix fusion strategies as done successfully before by other authors [27] will be investigated in future work. Our findings should also be validated using a larger data set and within the context of other pathologies.

Conclusion

In the present paper, we proposed a framework for comparing different pre-processing PET strategies using a multi-center series with the intent of predicting DFS in LACC patients. Our results show that there is a low reproducibility in predictions across scanners and discretization methods. Combining features calculated with different numbers of bins, relying on the normalizing effect of tumour to liver ratio, and using the maximum relevance minimum redundancy feature selection method was found to increase the robustness of the developed models and could be recommended for future radiomic studies in a similar context. These recommendations should now be evaluated in larger cohorts and in different cancer types.

Abbreviations

AUCpr: area under the curve of the precision recall curve

CT: computed tomography

DFS: disease free survival

D_All_FBW: discretization using all widths of the fixed bin width discretization

D_All_FBN: discretization using all bin's number of the fixed bin number discretization

D_Med_FBW: discretization using the median value of features with the different widths used in the fixed bin width discretization

D_Med_FBN: discretization using the median value of features with the different bins used in the fixed bin number discretization

[¹⁸F]FDG: 2-[¹⁸F]fluoro-2-deoxy-D-glucose

FBW: fixed bin width

FBN: fixed bin number

FS: feature selection

IBSI: Imaging biomarkers standardization initiative

LACC: locally advanced cervical cancer

LR: logistic regression

MI: mutual information

ML: machine learning

NB: Naïve Bayes

MRMR: maximum relevance minimum redundancy

OR: original radiomics

PET: positron emission tomography

RF: random forest

SVM: support vector machine

TLR: tumour to liver ratio

Declarations

Funding

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 766276. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Competing interests

Dr Philippe Lambin reports, within and outside the submitted work, grants/sponsored research agreements from Varian medical, Oncoradiomics, ptTheragnostic/DNAmito, Health Innovation Ventures. He received an advisor/presenter fee and/or reimbursement of travel costs/external grant writing fee and/or in kind manpower contribution from Oncoradiomics, BHV, Merck, Varian, Elekta, ptTheragnostic and Convert pharmaceuticals. Dr Lambin has shares in the company Oncoradiomics, Convert pharmaceuticals, MedC2 and LivingMed Biotech, he is co-inventor of two issued patents with royalties on radiomics (PCT/NL2014/050248, PCT/NL2014/050728) licensed to Oncoradiomics and one issue patent on mtDNA (PCT/EP2014/059089) licensed to ptTheragnostic/DNAmito, three non-patented invention (softwares) licensed to ptTheragnostic/DNAmito, Oncoradiomics and Health Innovation Ventures and three non-issues, non licensed patents on Deep Learning-Radiomics and LSRT (N2024482, N2024889, N2024889). He confirms that none of the above entities or funding was involved in the preparation of this paper.

Ethical approval and consent to participate

All procedures were performed in accordance with the principles of the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards. The study design and exemption from informed consent were approved by the Institutional Review Board of Liege University Hospital.

Consent for publication

For this type of retrospective study formal consent is not required.

Availability of data and materials

The datasets generated and/or analysed during the current study are available in <https://github.com/msilvaferreira/Phd/tree/master/FDG%20PET%20radiomics%20to%20predict%20disease%20free%20survival%20in%20Cervical%20Canc>

Authors' contributions

Study concepts/study design: Marta Ferreira, Patrick E.Meyer and Roland Hustinx

Literature research: Marta Ferreira and Patrick E.Meyer

Clinical studies: Johanne Hermesse, Marjolein Decuypere, Philippe Robin and Frédéric Kridelka

Data analysis: Marta Ferreira, Patrick E.Meyer, Pierre Lovinfosse and Roland Hustinx

Statistical analysis: Marta Ferreira and Patrick E.Meyer

Manuscript drafting, editing and revision: All authors read and approved the final manuscript

Acknowledgements

Not applicable

References

1. Lambin P, et al. Radiomics: Extracting more information from medical images using advanced feature analysis. *Eur J Cancer*. 2012;48:441–6.
2. Aerts HJWL, et al., “Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach,” *Nat. Commun.*, vol. 5, 2014.
3. Lambin P, et al. Radiomics: The bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol*. 2017;14:749–62.
4. Tsujikawa T, et al., “18F-FDG PET radiomics approaches: comparing and clustering features in cervical cancer,” *Ann. Nucl. Med.*, vol. 31, 2017.
5. Altazi BA, et al. Investigating multi-radiomic models for enhancing prediction power of cervical cancer treatment outcomes. *Phys Medica*. 2018;46:180–8.
6. Hao H, et al., “Shell feature: A new radiomics descriptor for predicting distant failure after radiotherapy in non-small cell lung cancer and cervix cancer,” *Phys. Med. Biol.*, vol. 63, 2018.
7. Lucia F, et al. Prediction of outcome using pretreatment 18 F-FDG PET/CT and MRI radiomics in locally advanced cervical cancer treated with chemoradiotherapy. *Eur J Nucl Med Mol Imaging*. 2018;45:768–86.
8. Bowen SR, et al. Tumor radiomic heterogeneity: Multiparametric functional imaging to characterize variability and predict response following cervical cancer radiation therapy. *J Magn Reson Imaging*. 2017;47:1388–96.
9. Shen WC, et al. Prediction of local relapse and distant metastasis in patients with definitive chemoradiotherapy-treated cervical cancer by deep learning from [18F]-fluorodeoxyglucose positron emission tomography/computed tomography. *Eur Radiol*. 2019;29:6741–9.
10. Sun W, Jiang M, Dang J, Chang P, Yin FF. “Effect of machine learning methods on predicting NSCLC overall survival time based on Radiomics analysis,” *Radiat. Oncol.*, 2018.
11. Leger S, Zwanenburg A, Pilz K, Lohaus F. “A comparative study of machine learning methods for time-to-event survival data for radiomics risk modelling”, pp. 1–11, 2017.
12. Parmar C, Grossmann P, Bussink J, Lambin P, Aerts HJWL. “Machine Learning methods for Quantitative Radiomic Biomarkers,” *Sci. Rep.*, vol. 5, 2015.
13. Deist TM, et al., “Machine learning algorithms for outcome prediction in (chemo)radiotherapy: An empirical comparison of classifiers,” *Med. Phys.*, 2018.
14. Lei M, et al., “Benchmarking features from different radiomics toolkits / toolboxes using Image Biomarkers Standardization Initiative,” *arXiv*, 2020.
15. Lovinfosse P, Visvikis D, Hustinx R, Hatt M. “FDG PET radiomics: a review of the methodological aspects,” pp. 379–391, 2018.
16. Shiri I, Rahmim A, Ghaffarian P, Geramifar P, Abdollahi H, Bitarafan-rajabi A. The impact of image reconstruction settings on 18F-FDG PET radiomic features: multi-scanner phantom and patient studies Lesions to Background Ratio Time of Flight Full Width at Half Maximum Response Evaluation Criteria in Solid Tumours. *Eur Radiol*. 2017;27:4498–509.
17. Shafiq-ul-hassan M, et al. Intrinsic dependencies of CT radiomic features on voxel size and number of gray levels. ” *Med Phys*. 2017;44:1050–62.
18. Leijenaar RTH, et al., “The effect of SUV discretization in quantitative FDG-PET Radiomics: the need for standardized methodology in tumor texture analysis,” *Nat. Publ. Gr.*, pp. 1–10, 2015.
19. Altazi BA, et al. 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.
20. Liberini V, et al., “Impact of segmentation and discretization on radiomic features in 68Ga-DOTA-TOC PET/CT images of neuroendocrine tumor,” *EJNMMI Phys.*, vol. 8, 2021.
21. Aide N, Lasnon C, Veit-Haibach P, Sera T, Sattler B, Boellaard R. EANM/EARL harmonization strategies in PET quantification: from daily practice to multicentre oncological studies. *Eur J Nucl Med Mol Imaging*. 2017;44:17–31.
22. Schwier M, et al., “Repeatability of multiparametric prostate MRI radiomics features,” *Sci. Rep.*, 2019.
23. Ferreira M, et al., “[18F]FDG PET radiomics to predict disease-free survival in cervical cancer: a multi-scanner/center study with external validation,” *Eur. J. Nucl. Med. Mol. Imaging*, 2021.
24. Hatt M, Cheze C, Turzo A, Roux C. A fuzzy locally adaptive Bayesian segmentation approach for volume determination in PET. ” *IEEE Trans Med Imaging*. 2009;28:881–93.
25. Zwanenburg A, et al. The image biomarker standardization initiative: Standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology*. 2020;295:328–38.
26. Hanchuan Peng F. Long, “Feature Selection Based on Mutual Information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy. *IEEE Trans Pattern Anal Mach Intell*. 2005;27:1226–38.
27. Breiman L, Friedman J, Olshen R, Classification and Regression Trees. 1984.
28. Géron A. Hands-On Machine Learning with Scikit-Learn. 1st ed.: O’Reilly; 2017.
29. Zhou Z, et al., “Constructing multi-modality and multi-classifier radiomics predictive models through reliable classifier fusion,” *IEEE Comput. Soc.*, 2017.

30. Zhou Z, et al. "Multifactorial cancer treatment outcome prediction through multifaceted radiomics. ", arXiv: Medical Physics; 2018.
31. Osman AFI. A Multi-parametric MRI-Based Radiomics Signature and a Practical ML Model for Stratifying Glioblastoma Patients Based on Survival Toward Precision Oncology. *Front Comput Neurosci.* 2019;13:1–15.
32. Presotto L, et al. *Physica Medica* Original paper PET textural features stability and pattern discrimination power for radiomics analysis: An 'ad-hoc' phantoms study. *Phys Medica.* 2018;50:66–74.
33. Vallières M, et al. Radiomics strategies for risk assessment of tumour failure in head-and-neck cancer. *Sci Rep.* 2017;7:1–14.
34. Lv W, Ashrafinia S, Ma J, Lu L, Rahmim A. Multi-level multi-modality fusion radiomics: Application to PET and CT imaging for prognostication of head and neck cancer. *IEEE J Biomed Heal Informatics.* 2020;24:2268–77.
35. N. S.-M. & A. A.-B. Verónica Bolón-Canedo, *Feature Selection for High-Dimensional Data.* Springer, 2016.
36. Scalco E, et al. T2w-MRI signal normalization affects radiomics features reproducibility. *Med Phys.* 2020;47:1680–91.
37. Haga A, et al. Standardization of imaging features for radiomics analysis. *J Med Investig.* 2019;66:35–7.
38. Li XT, Huang RY. Standardization of imaging methods for machine learning in neuro-oncology. *Neuro-Oncology Adv.* 2020;2:iv49–55.
39. Isaksson LJ, et al. Effects of MRI image normalization techniques in prostate cancer radiomics. *Phys Medica.* 2020;71:7–13.
40. Traverso A, et al. Sensitivity of radiomic features to inter-observer variability and image pre-processing in Apparent Diffusion Coefficient (ADC) maps of cervix cancer patients. *Radiother Oncol.* 2020;143:88–94.
41. Iantsen A, et al., "Convolutional neural networks for PET functional volume fully automatic segmentation: development and validation in a multi-center setting," *Eur. J. Nucl. Med. Mol. Imaging,* 2021.

Figures

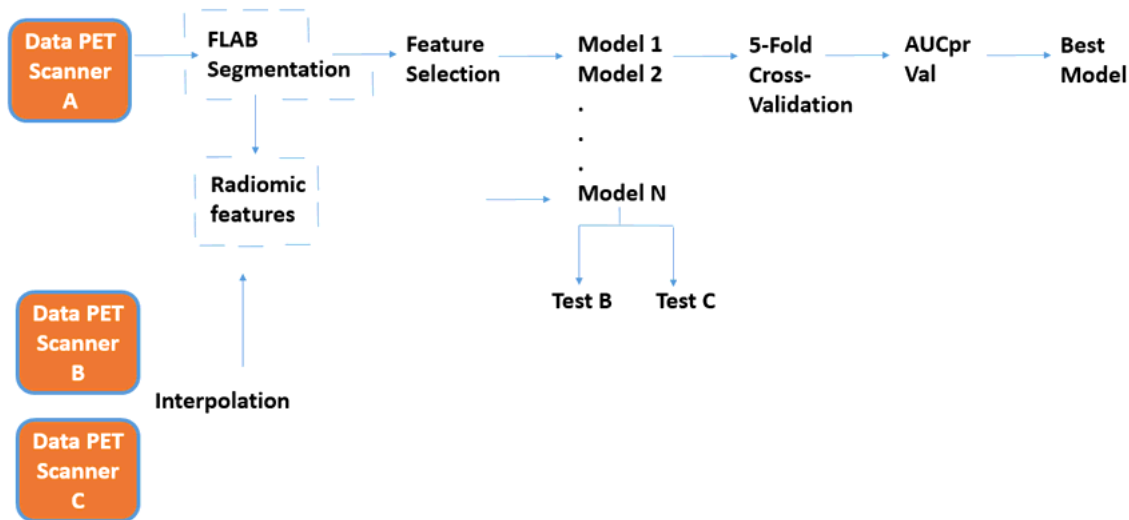


Figure 1

Radiomics workflow.

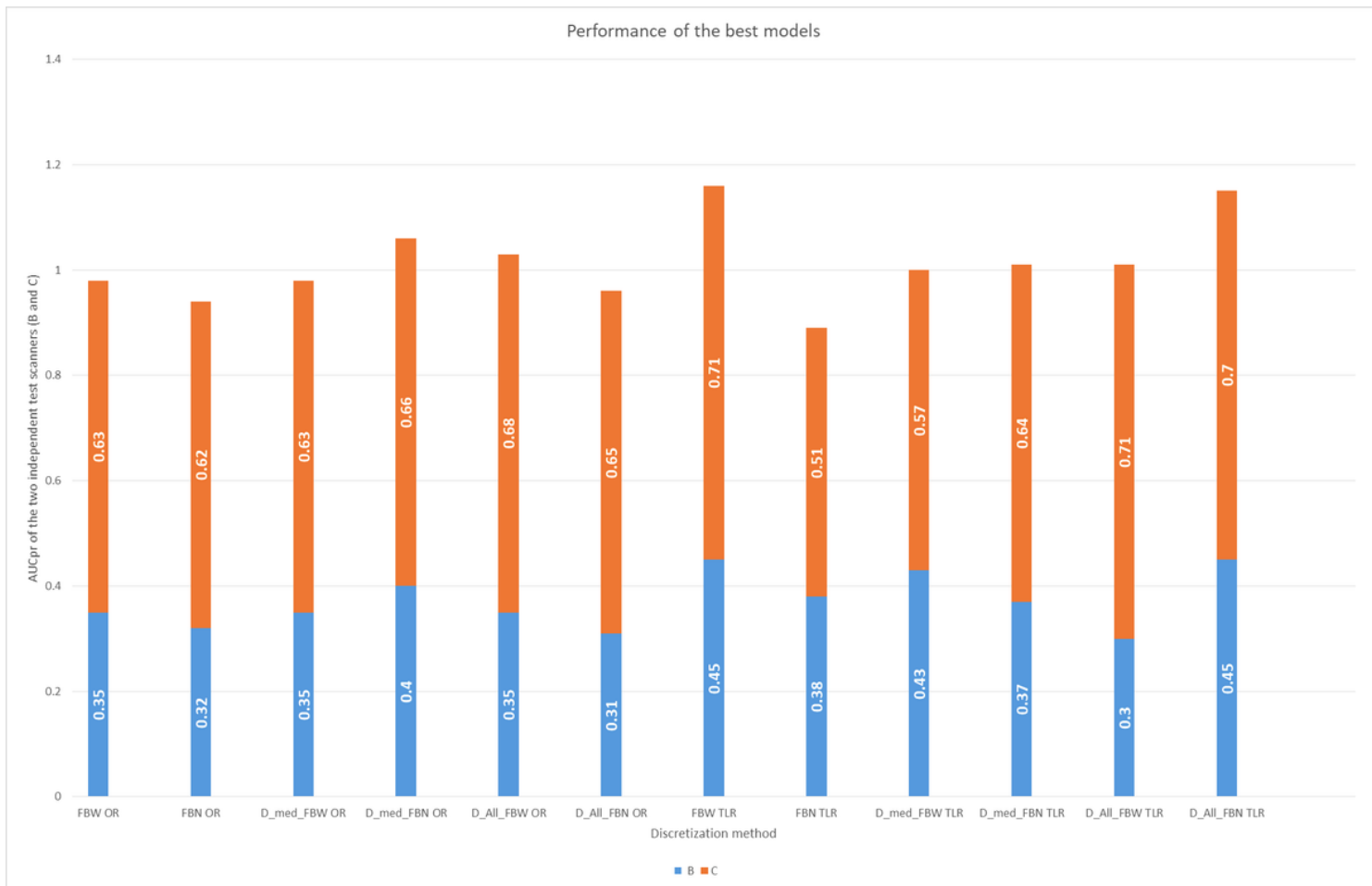


Figure 2

Bar plot showing the AUCpr in the two external scanners (B and C) for all discretization strategies. The models depicted in the plot use the FS method and classifier which showed better performance in the internal validation scheme (scanner A).

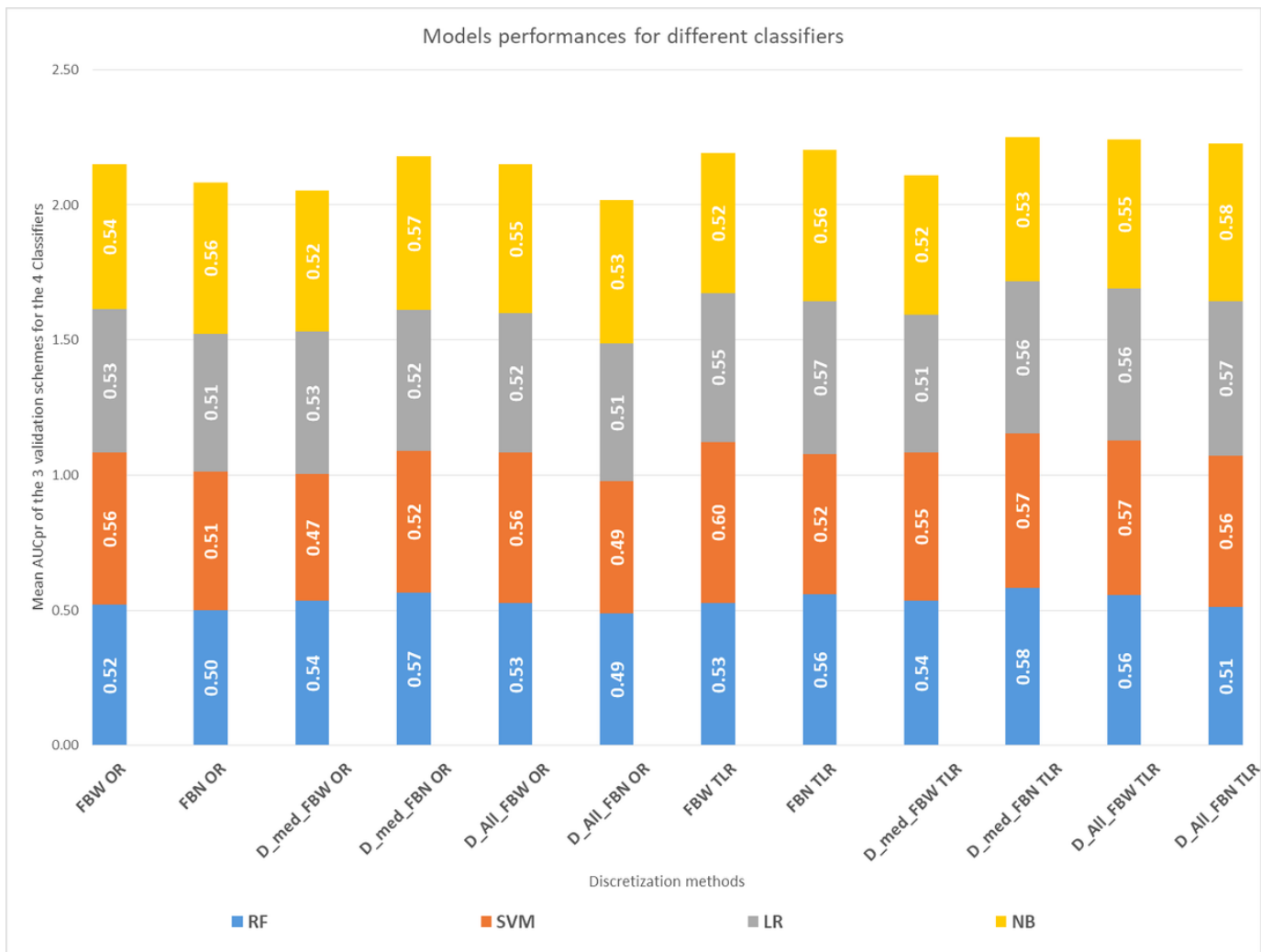


Figure 3
 Bar plot showing the models with FS that showed the better performances for all intensity discretization schemes and for the four classifiers. The AUCpr correspond to the mean value of the 3 validation schemes.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementarymaterial.pdf](#)