A European Cancer Image Platform Linked to Biological and Health Data for Next-Generation Artificial Intelligence and Precision Medicine in Oncology

# Deliverable D4.5: Standardisation of cancer image data and features

| | |
|---|---|
| Reference | D4.5_ EuCanImage_UM_v1 |
| Lead Beneficiary | UM |
| Author(s) | Philippe Lambin, Henry Woodruff, Shruti A. Mali |
| Dissemination level | Public |
| Type | Report |
| Official Delivery Date | 31st March 2023 |
| Date of validation of the WP leader | 30th March 2023 |
| Date of validation by the Project Coordinator | 30th March 2023 |
| Project Coordinator Signature |  |

EuCanImage is funded by the European Union's H2020 Framework Under Grant Agreement No 952103

## Version log

| Issue Date | Version | Involved | Comments |
|---|---|---|---|
| 23/03/2023 | V0.1 | Philippe Lambin, Henry Woodruff, Shruti A. Mali | Initial draft |
| 30/03/2023 | V0.2 | Fred Prior | Feedback |
| 30/03/2023 | V1 | Anais Emelie & Karim Lekadir | Revised and corrected final version. |
| | | | |
| | | | |

## Executive Summary

Standardising imaging data is a critical step in the field of medical image analysis. Before applying machine learning or deep learning techniques on these images/features, the data need to be 'cleaned' and standardised to attain maximum reproducibility. In this context, we introduce the precision-medicine-toolbox that allows researchers to perform data curation, image pre-processing and handcrafted radiomics feature exploration tasks. This toolbox will also benefit the researchers without a strong programming background to implement these critical steps and increase the reproducibility of quantitative medical imaging research. Furthermore, we deliver a guideline document that informs researchers in standardizing imaging data at both image level and feature level. With the help of MLToolbox (built on top of WORC), precision-medicine-toolbox and guideline document, consensus will be reached with the AI developers on defining the pre-processing pre-sets per use case. The goal is to eventually define these pre-sets and apply them in XNAT so that pre-processed images are readily available to download.

## Table of Contents



## Acronyms

| Name | Abbreviation |
|---|---|
| Computed tomography | CT |
| Magnetic resonance imaging | MRI |
| Positron emission tomography | PET |
| Digital imaging and communications in medicine | DICOM |
| Deep learning | DL |
| The image biomarker standardisation initiative | IBSI |

# 1 Introduction

The basis for this task is divided into 3 tiers. First, the ML Toolbox that is built on top of the existing WORC framework (Starmans et al. 2021) and integrated in the VRE. This toolbox consists of various pre-processing steps such as clipping with defined intensity boundaries, z-score/min-med normalization, resampling the voxels spacings and bias field correction. The MLtoolbox also consists of ComBat tool for feature harmonization. Second, we deliver precision-medicine-toolbox that is an open-source toolkit that would aid the AI developers in their pre-processing pipelines if they wish to tailor the process for example, locally running pre-processing functions on imaging data according to their needs. This toolbox further provides insights on handcrafted radiomic features by performing exploration tasks by visualizing and calculating basic statistics. Third, we deliver the review paper 'Making radiomics more reproducible across scanner and imaging protocol variation: A review of harmonization methods' which would serve as a basis for the guideline document. The next step after this deliverable is to come to a consensus with the AI developers and decide on the pre-processing pre-sets per use case. The plan is to implement these pre-sets in the XNAT itself to be able to download data that is already pre-processed. Refer Figure 1 for the overview of this task in the EuCanImage workflow.

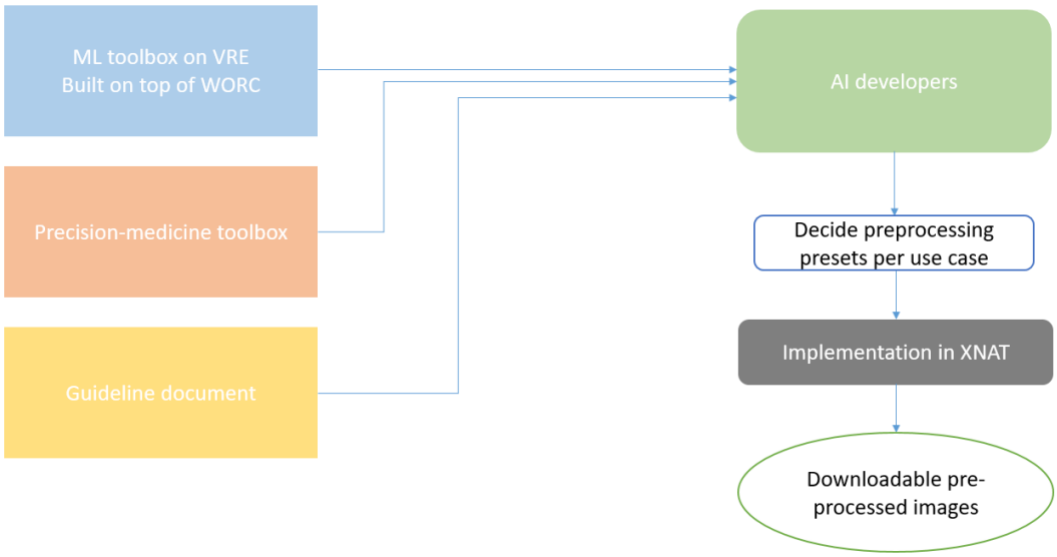


*Figure 1 Overview of Task 4.5*

# 2 Image standardization toolbox

Medical image analysis plays a key role in precision medicine as it allows the clinicians to identify anatomical abnormalities and it is routinely used in clinical assessment. Data curation and pre-processing of medical images are critical steps in the quantitative medical image analysis that can have a significant impact on the resulting model performance. In this task, we introduce a precision-medicine-toolbox [1] that allows researchers to perform data curation, image pre-processing and handcrafted radiomics extraction (via Pyradiomics) and feature exploration tasks with Python. With this open-source solution, we aim to address the data preparation and exploration problem, bridge the gap between the currently existing packages, and improve the reproducibility of quantitative medical imaging research. The

precision-medicine-toolbox that was developed in line with this task and would allow AI developers to perform certain image pre-processing tasks if they want to do it locally or in case they want to customize their pre-processing pipeline.

## 2.1 The precision-medicine-toolbox

The precision-medicine-toolbox allows researchers to perform data curation, image pre-processing and handcrafted radiomics feature exploration tasks. This toolbox will also benefit the researchers without a strong programming background to implement these critical steps and increase the reproducibility of quantitative medical imaging research. The functionalities of the toolbox are shown in Figure 2.

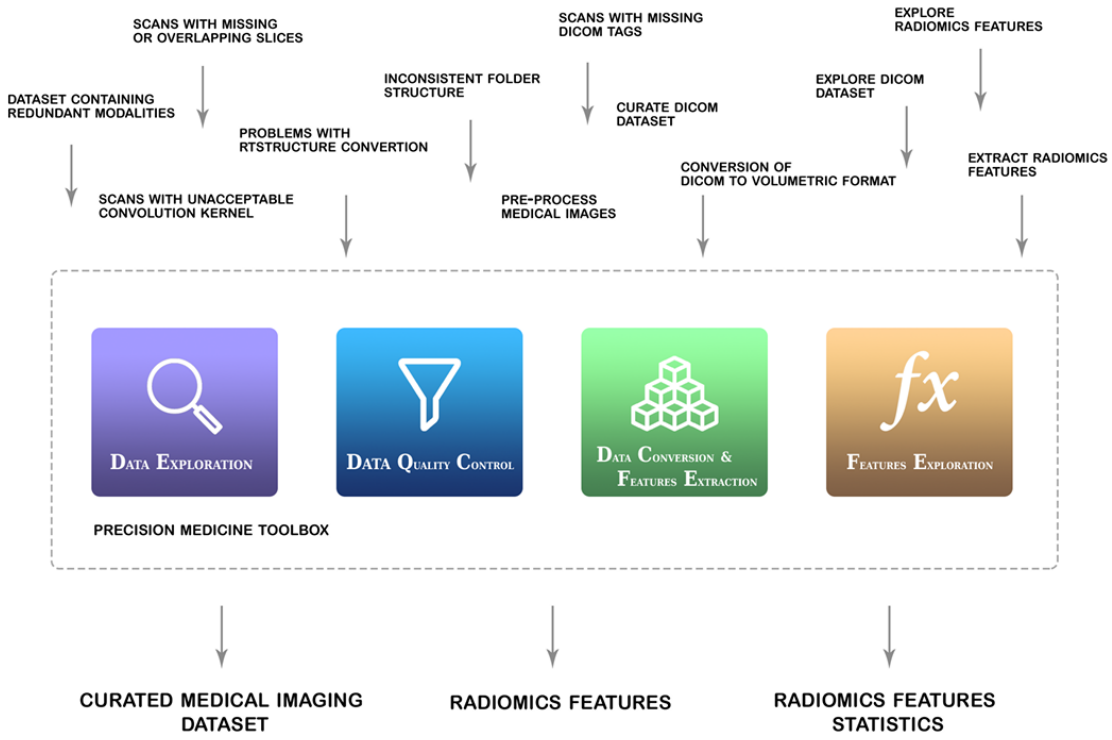


*Figure 2 Overview of the precision-medicine-toolbox*

The imaging module allows for pre-processing and exploration of the imaging datasets. Currently, the following functions are implemented:

- dataset parameters exploration by parsing of the imaging metadata,
- dataset basic quality examination, including check of imaging modality, slice thickness, number of slices, in-plane resolution and pixel spacing, and reconstruction kernel,
- conversion of DICOM dataset into volumetric Nearly Raw Rusted Data (NRRD) dataset,
- image basic pre-processing, including bias field correction (for MRI) , intensity rescaling and normalization, histogram matching, intensities resampling, histogram equalization, image reshaping,

• unrolling NRRD images and ROI masks into Joint Photographic experts Group (JPEG) slices for a quick check of the converted images or any existing NRRD or MetaImage Medical Format (MHA)dataset,
• radiomics features extraction from NRRD/MHA data using PyRadiomics package.

Furthermore, this model also includes a features module that allows for the exploration of the feature datasets. It allows for the basic analysis of the features. Currently, the following functions are implemented:

- visualization of feature values distributions in classes,
- visualization of features mutual Spearman correlation matrix,
- calculation of corrected p-values for Mann-Whitney test for features mean values in groups,
- visualization of univariate receiver operating characteristic (ROC) curves for each feature and calculation of the area under the curve (AUC),
- volumetric analysis, including visualization of volume-based precision-recall curve and calculation of Spearman correlation coefficient between every feature and volume,
- calculation of basic statistics (number of missing values, mean, std, min, max, Mann-Whitney test p-values for binary classes, univariate ROC AUC for binary classes, Spearman's correlation with volume if volumetric feature name is sent to function) for every feature.

The precision-medicine-toolbox is implemented in Python (Python Software Foundation, Wilmington, DA, U.S.) and requires version 3.6 or higher. The source code is hosted on GitHub (https://github.com/primakov/precision-medicine-toolbox) and Zenodo platform (DOI 10.5281/zenodo.6126913). The precision-medicine-toolbox package has been released under the BSD-3-Clause License and is available from the Python Package Index (PyPI) repository (https://pypi.org/project/precision-medicine-toolbox/). An easy installation of the latest version is possible with "pip install precision-medicine-toolbox" command. The project has the following structure:

- README.MD:file with the project overview,
- Requirements.txt: file with the list of the packages to be installed,
- LICENSE: statement of the license applicable to the project's software and manuscripts, • .gitignore: specification of the files, intentionally untracked by Git,
- .readthedocs.yaml: Read the Docs configuration file,
- Mkdocs.yml: Mkdocs configuration file,
- Setup.cfg and setup.py: configuration files for PyPi package,
- Data: folder with the raw data for the examples as well as generated files,
- Docs: folder with the documentation files,
- Examples: folder with the examples:
    - Imaging_module.ipynb: tutorial illustrating functionality for the imaging datasets,
    - Features_module.ipynb: tutorial illustrating functionality for the features datasets,
- Pmtool: folder with the toolbox scripts:
    - __init__.py: initialization file,

- data_set.py: script defining the base class for imaging datasets,
- tool_box.py: script defining the inheriting class for imaging datasets methods,
- features_set.py: script defining the base class for features datasets,
- analysis_box.py: script defining the inheriting class for features datasets methods.

# 3 Guidelines for standardizing imaging data

This section contains general guidelines that would eventually aid in standardizing imaging data at both image level and feature level. This section will briefly cover three topics under the deliverable: (1) guidelines to classify image acquisition protocols, (2) standardised definitions of features and (3) identification of reproducible radiomic features. For detailed information on guidelines related to harmonization strategies, to make radiomic features more reproducible, we deliver a narrative review '**Making Radiomics More Reproducible across Scanner and Imaging Protocol Variations: A Review of Harmonization Methods**' [2]. Different harmonization solutions are discussed and divided into two main categories: image domain and feature domain. The image domain category comprises methods such as the standardization of image acquisition, post-processing of raw sensor-level image data, data augmentation techniques, and style transfer. The feature domain category consists of methods such as the identification of reproducible features and normalization techniques such as statistical normalization, intensity harmonization, ComBat and its derivatives, and normalization using deep learning. We also reflect upon the importance of deep learning solutions for addressing variability across multi-centric radiomic studies. The overall objective of this review is to address the advantages, disadvantages, and challenges posed by these harmonization methods. Figure 3 shows an overview of different harmonization methods that are applicable at different stages of medical imaging.

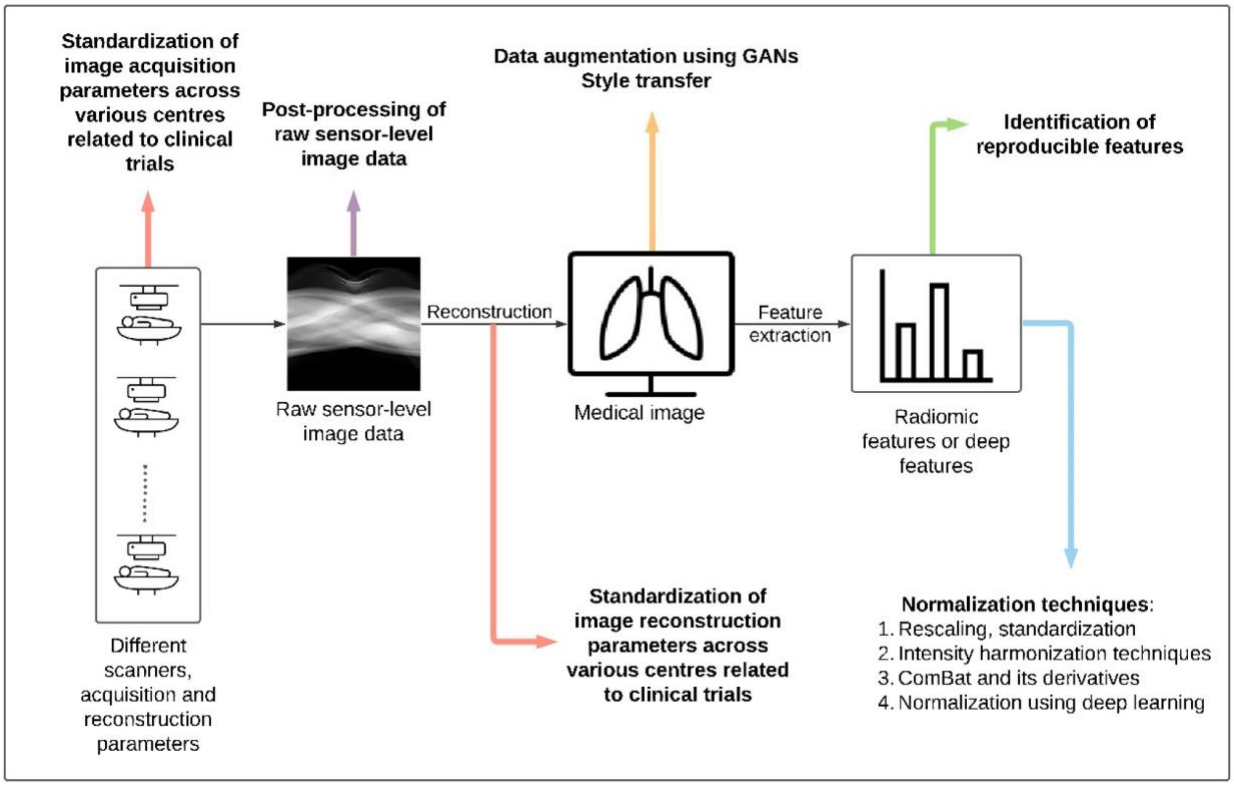


*Figure 3 Overview of harmonization methods at different stages of medical imaging*

A powerful radiomics approach involves the extraction of various quantitative features from medical images, storing this data in a federated form of database [3] where several individual databases function as an entity, and successive mining of data to acquire relevant clinical outcomes [4]. Large quantities of data are required to develop robust predictive models and this amount of data is usually obtained from multiple hospitals and/or institutions. Furthermore, due to continuous improvement in scanner and protocol settings, this type of data is a moving target. To compensate for the effects scanner/protocol variability might have on the predictive models, large quantities of data are needed to make systems generalize. In these cases, federated (or distributed) learning could be adapted so as to allow sharing of data between hospitals/institutes to develop robust predictive models [4]. Major management problems still exist even though there are databases that are collecting and cross-referencing massive amounts of radiomics information in addition to other related patient data from millions of case studies [5–8].

Radiomic feature extraction can be categorized into two main approaches: hand-crafted (derived from traditional statistical and computer vision methods) and deep learning (DL). Hand-crafted radiomics characteristics (such as texture, shape, intensity) provide information on the particular area of the medical imaging scan, often referred to as the region or volume of interest (ROI or VOI), which could be a tumour, a tissue or an organ as a whole [9]. DL algorithms learn complex visual features and perform ROI segmentation using cascading layers with non-linearities by using 'sliding' kernels in convolutional neural networks (CNN), while hand-crafted features represent the spatial appearances (texture and shape) by

mathematically extracting spatial distribution on inter-pixel relationships, signal intensities, gray-scale patterns and spectral properties [10]. DL has the benefit of not necessarily requiring prior segmentation masks of the medical imaging scan. However, DL is a 'black box' approach, i.e. the lack of interpretability of the models and the deep features generated are seen as a key limitation in clinical applications [11]. DL also requires a larger amount of data and/or pre-trained models often trained on diverse domains (e.g. photographic images), in order to perform efficiently and effectively. The vast majority of published radiomic models lack consistent evaluation of performance, sufficient large-scale annotated datasets for radiomic studies, reproducibility, clinical efficacy, and large-scale validation on sufficiently large cohorts, despite these being prerequisites for clinical translation [12,13]. Furthermore, there is a lack of reproducibility of radiomic features while translating results into clinical practice [14]. Ideally, the features extracted using radiomics represent imaging biomarkers and should be independent of image acquisition parameters or protocols [15]. For example, if a patient is scanned in different hospitals, the quantitative features extracted from all these scans should either have similar values or the correct transformation should be known. Scanner protocols and hardware are constantly changing over time and differ across hospitals. The same scanner can also be configured differently. Frequent software updates might have an influence on images produced. A major consequence of these scanner and protocol variations is a domain shift [16], i.e. a shift in data distribution across various centres/time/machines/software.

## 3.1 Guidelines to classify image acquisition protocols

For multicentric prospective studies, the ideal way to standardise radiomic features is to define and follow imaging protocols that define scanner types in conjunction with acquisition and reconstruction parameters. For example, the European Society for Therapeutic Radiology and Oncology (ESTRO) panel provides guidelines for procedures and methods for image-guided radiation therapy (IGRT) in prostate cancer [17,18]. This panel consulted a large base of the radiation oncology community from the European Union and developed guidelines for delineating localized prostate cancer in (computed tomography) CT and magnetic resonance images (MRI). ESTRO also has a working group focusing on cervical carcinoma for developing and validating methods and imaging parameters from various institutions [19]. For standardization of PET imaging, the European Association of Nuclear Medicine (EANM) [20] launched the EARL (EANM Research Ltd) program covering areas such as scan acquisition, processing of images, and image interpretation. Pfaehler et al. [21] conducted a study to investigate the effects of harmonizing image reconstructions on feature reproducibility and concluded that EARL compliant image reconstruction harmonized a wide selection of radiomic features. A similar initiative by the American Society for Radiation Oncology (ASTRO) [22] was created to develop a 'practice parameter', for IGRT and to provide quality assurance standards, personnel qualifications, indications and guided documentation [23] for imaging. In MRI, however, such guidelines do not exist [24] and most of the MRI modalities are not even quantitative [25]. Efforts have been taken in the past, concerning MRI imaging, for example by UCHealth [26] to reduce the number of MRI protocols from 168 to 66 across scanners and centres by selecting an appropriate clinics-driven protocol and standardization process. Another set of guidelines is provided by the FDA (Food and Drug Administration)

[27] to focus on image acquisition in clinical trials conducted to support the authorization of drugs and biological products. Ever since this draft by FDA was released in 2015, it has become a reference standard for most promoters and industries of clinical trials. Another such way to ensure that the protocols are well documented could be by following Transparent Reporting of Medical Image Acquisition (TRIAC [28,29]) guidelines for future proof radiomics or if a public protocol is used. TRIAC guidelines describe five different levels of evidence for reporting imaging protocols.

- Level 0 indicates that the protocol has not been formally approved with a reference number
- Level 1 indicates that the protocol has been approved with a reference number in the archive of the department
- Level 2 indicates that the protocol has been approved with formal quality assurance (recommended minimum level for prospective trials)
- Level 3 indicates that the protocol is established internationally and has been published in guideline documents and peer-reviewed papers
- Level 4 indicates that the protocol is Future proof i.e., the protocol follows TRIAC Level 3, FAIR principles and retains raw data.

It would be worthwhile to work with raw sensor-level data, right before reconstructing the image and apply harmonization methods on it to remove scanner and protocol variability. Image reconstruction, necessary for human viewing and interpretation, combined with the manual contouring variability, could lead to loss of latent raw sensor-level image data and lower precision in measurements. Most machine learning (ML) and DL algorithms have been used on reconstructed images in the existing medical imaging workflow. Instead, the abilities of ML and DL could be leveraged to process the underlying raw sensor-level data to access its hidden nuances [30–32]. Radiomics signature analysis can also be performed directly on the raw image data without the need for reconstruction which adds bias and variability [33,34]. Furthermore, the reconstruction process itself can also be considered as a prediction problem utilising raw CT data (sinograms) or k-space values of MRI inputs [35]. These studies widen the scope to apply harmonization methods on raw image data and take advantage of the hidden information in the raw image data rather than applying it in the reconstructed image-space [2].

Reporting the scanner hardware settings, image acquisition, and reconstruction methods are also critical in view of standardizing imaging protocols. Most of the radiomics studies include retrospective datasets that have already been imaged in the past with pre-set scanner/s and standardizing imaging protocols at this point might not be feasible. To tackle this either pre-processing of images could be done before image analysis or a phantom study could be performed to detect inter-scanner differences and vendor-dependent features to assess the feature robustness [36–41]. This is intended to increase feature reproducibility. Few pre-processing methods have been used in previous work including isotropic voxel resampling, bias field correction [42], normalizing intensity scales using histogram equalization [43,44], gray-level discretization [45], and processing of raw sensor-level image data [33,46]. Pre-

processing step is crucial for standardizing heterogeneous datasets to increase the reproducibility of features [29]. Acquiring images from individuals at multiple time points also allows analysing feature robustness across temporal variabilities (e.g., organ movement). Once the model has been decided, defining model constraints is crucial for its development. E.g., defining inclusion and exclusion criteria for model inputs; detecting and eliminating biases (e.g., sex, ethnicity, socio-economic factors, data imbalance) occurring due to diversity and distribution across diverse patient groups within the dataset/s.

## 3.2 Standardised definitions of features following the recommendations of Image Biomarker Standardization Initiative (IBSI)

The IBSI aimed to standardise the extraction of radiomic features which were assessed in three phases [29]. One of the reasons for lack of reproducibility of radiomic features is the scarcity of consensus-based guidelines and standardised definitions for the process of high-throughput extraction of quantitative features, eventually meant to aid clinical decision support systems. A total of 169 radiomic features were standardised which further verified and calibrated different radiomics softwares. A reference manual [47] is available that presents efforts to chart a course through part of this frontier by presenting consensus-based recommendations, guidelines, definitions and reference values for image biomarkers and defining a general radiomics image processing scheme. This manual will help improve the reproducibility of radiomic studies. This reference manual is divided into several chapters that describe processing of acquired and reconstructed (medical) imaging for high-throughput computation of image biomarkers (Chapter 2: Image processing) [48]; that define a diverse set of image biomarkers (Chapter 3: Image features) [49]; that describe guidelines for reporting on radiomic studies and provide nomenclature for image biomarkers (Chapter 4: Radiomics reporting guidelines and nomenclature) [50]; and that describe the data sets and image processing configurations used to find reference values for image biomarkers (Chapter 5: Reference data sets) [51].

## 3.3 Identification of reproducible features

Various studies have tested the reproducibility, variability and repeatability of features extracted from various phantom and patient studies over different reconstruction and acquisition parameters in the case of multi-centric datasets and examine the reproducibility of radiomic features. High reproducibility can be assessed by improving the intra-individual reproducibility, multi-centre/multi-machine reproducibility, multi reader reproducibility, image reconstruction and pre/post-processing methods. Clustering based techniques such as k-means clusters, principal component analysis and t-distributed stochastic neighbour embedding can be utilised to visualize and measure inter-class/intra-class correlation. Conducting test-retest studies on phantoms could contribute towards reproducibility. Using a phantom allows a controlled analysis that isolates the variation due to scanner variation from other variations related with patient acquisition. Phantoms can also be scanned by specific scanners with special clinical settings to specifically improve the normalization of the features

for clinical use. The normalization could therefore be updated to follow the latest imaging advances and standards [2]. Meaningful features could be selected by using intra-class correlation coefficient (ICC) and concordance correlation coefficient (CCC) cut-off values so as to get rid of redundant features. In case of multiple segmentations, stable features could be selected across different annotations so as to reduce the dimensionality of the radiomic features.

In the context of PET images, a study by Shiri et al. [38] investigated the impact of various image reconstruction settings on several PET/CT radiomic features obtained from a phantom dataset (developed in-house National Electrical Manufacturers Association [NEMA]) and a patients dataset from two different scanners. Radiomic features were grouped into intensity-based, geometry-based and texture-based features and their reproducibility and variability were evaluated using the coefficient of variation (COV). The results from both phantom and patient studies showed that 47% of all radiomic features were reproducible. Almost half of intensity-based and texture-based and all of the geometry-based features were found to be reproducible respectively. The intensity and geometry-based features were also found to be reproducible in another study by Vuong et al. [52], where the authors investigate if the PET/CT radiomics models can be transferred to PET/MRI models by checking the reproducibility of radiomic features against different test-retest and attenuation correction variability. However, Shiri et al. [38] used a phantom body filled with homogeneous activity rather than heterogeneous activity, which does not properly imitate the human tissue. A similar study by Bailly et al. [53] analysed the reproducibility of texture features in PET scans across different acquisition and reconstruction parameters in the context of multi-centre trials. They found out that only a few features were strongly reproducible and acceptable for multi-centre trials. Nevertheless, this study checked the reproducibility of texture features evaluated against reconstruction parameters coming from just one manufacturer. Many such studies have been carried out to check the reproducibility of radiomic features in PET scans [54–68] where they only assess the impact of variability in scanner and imaging parameters and do not provide concrete image and/or feature harmonization methods to obtain reproducible features.

In the case of CT scans, Prayer et al. [39] conducted a trial to investigate the inter-and intra-scanner repeatability and reproducibility of computed tomography (CT) radiomic features (radiomic feature) of fibrosing interstitial lung disease (fILD). The dataset was obtained from IRB-approved test-retest study with sixty fILD patients. Results showed that intra and inter-scanner reproducibility were highly affected by the variation in slice thicknesses than the variation in reconstruction kernels under study and were reconstruction parameter-specific respectively. The CT radiomic features showed excellent reconstruction parameter-specific repeatability for the test-retest study. However, the sample size of the data used was small, and to check the variability of features only two scanners were used. Careful selection of radiomic features is critical to ensure plausible outcomes in heterogeneous CT datasets. Similar studies have been conducted in the past where the reproducibility of CT radiomic features was investigated using phantom data [69–72] as well as patient data [14,73,74]. The phantom studies were carried forward to reduce the exposure to patients however, they are not real substitutes of heterogeneous human tissues.

Considering MRI, a recent study using radiomics to investigate the reproducibility of features across several MRI scanners and scanning protocol parameters was carried out using both

phantom data and patient (volunteer) data by Lee et al. [40]. This study also investigated repeatability by measuring the variability of radiomic features using a test-retest strategy. The variability of radiomic features across different MRI scanners and protocols was evaluated using the ICC and the repeatability was evaluated using the coefficient of variation (COV). The COV measurements showed that there was very little difference in the variability between filtering and normalizing effects which were used for pre-processing. The ICC measurements showed higher repeatability for the phantom data than for the patient data. However, this study was not able to prevent the effects of the volunteer's movements on the radiomic values despite simulating movements while scanning. A similar study, conducted by Peerlings et al. [41], extracted stable parametric MRI radiomic features with a minimum concordance correlation coefficient of 0.85 between data derived from 61 patients' test and retest apparent diffusion coefficient (ADC) maps across various MRI-systems, tissues and vendors. A review by Traverso et al. [75] mentions that there are not many phantom studies conducted to investigate the reproducibility of MRI radiomic features. Most of them cover various sites such as the brain [76,77], the gastro-intestinal tract [78–80] and the prostate [81,82], although this limitation was addressed by Rai et al. [83] by developing a novel 3D MRI radiomic phantom to assess the robustness and reproducibility of MRI radiomic features across multiple centres.

Furthermore, to correct for image protocol differences, ComBat [84] method can be used. ComBat harmonization is a statistical method that was developed originally to harmonize gene expression arrays [85]. ComBat was designed to provide estimates of the effects of assigned batches which have a single technical difference between each other, while taking into account the effect of biological covariates on the variables or features being harmonized. The estimations are calculated using Bayesian models, and a location/scale shift is performed accordingly to adjust the values of different features. The application of ComBat on radiomics features was first introduced by Fortin et al. [84]. The authors used ComBat to harmonize cortical thickness measurements calculated on diffusion imaging tensor data to remove variations in feature values attributed to differences in acquisition and reconstruction parameters. ComBat removes interscanner variability for these measurements and can also preserve biological correlations. The ComBat functionality is present in The MLToolbox that is built on top of the existing WORC framework [86] and integrated in the VRE.

# 4 Conclusion

This task introduces a toolbox, to curate, pre-process imaging data and explore handcrafted radiomic features. The precision-medicine-toolbox will benefit the researchers without a strong programming background to implement these critical steps and increase the reproducibility of quantitative medical imaging research. For this task we also deliver a guideline document, an article [2], that informs researchers in standardising imaging data at both image level and feature level. The MLToolbox is available on the VRE platform built on top of WORC which supports commonly used pre-processing functions. Moving forward, consensus will be attained with the AI developers to decide the pre-processing pre-sets for each use case, so that it can be implemented in XNAT. Once this consensus is reached with the AI developers, the functionalities can be implemented in XNAT so that the pre-

processed images are available for further AI development.

# 5 References


1. Primakov S, Lavrova E, Salahuddin Z, Woodruff HC, Lambin P. Precision-medicine-toolbox: An open-source python package for facilitation of quantitative medical imaging and radiomics analysis. arXiv [eess.IV]. 2022. Available: http://arxiv.org/abs/2202.13965

2. Mali SA, Ibrahim A, Woodruff HC, Andrearczyk V, Müller H, Primakov S, et al. Making Radiomics More Reproducible across Scanner and Imaging Protocol Variations: A Review of Harmonization Methods. J Pers Med. 2021;11. doi:10.3390/jpm11090842

3. Zerka F, Barakat S, Walsh S, Bogowicz M, Leijenaar RTH, Jochems A, et al. Systematic Review of Privacy-Preserving Distributed Machine Learning From Federated Databases in Health Care. JCO Clin Cancer Inform. 2020;4: 184–200.

4. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. Radiology. 2016;278: 563–577.

5. Roelofs E, Persoon L, Nijsten S, Wiessler W, Dekker A, Lambin P. Benefits of a clinical data warehouse with data mining tools to collect data for a radiotherapy trial. Radiotherapy and Oncology. 2013. pp. 174–179. doi:10.1016/j.radonc.2012.09.019

6. Roelofs E, Dekker A, Meldolesi E, van Stiphout RGPM, Valentini V, Lambin P. International data-sharing for radiotherapy research: An open-source based infrastructure for multicentric clinical data mining. Radiotherapy and Oncology. 2014. pp. 370–374. doi:10.1016/j.radonc.2013.11.001

7. Miotto R, Li L, Kidd BA, Dudley JT. Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records. Scientific Reports. 2016. doi:10.1038/srep26094

8. Nead KT, Gaskin G, Chester C, Swisher-McClure S, Dudley JT, Leeper NJ, et al. Androgen Deprivation Therapy and Future Alzheimer's Disease Risk. Journal of Clinical Oncology. 2016. pp. 566–571. doi:10.1200/jco.2015.63.6266

9. Gatenby RA, Grove O, Gillies RJ. Quantitative imaging in cancer evolution and ecology. Radiology. 2013;269: 8–15.

10. van Timmeren JE, Cester D, Tanadini-Lang S, Alkadhi H, Baessler B. Radiomics in medical imaging—"how-to" guide and critical reflection. Insights Imaging. 2020;11: 91.

11. Alzubaidi L, Zhang J, Humaidi AJ, Al-Dujaili A, Duan Y, Al-Shamma O, et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. J Big Data. 2021;8: 53.

12. Limkin EJ, Sun R, Dercle L, Zacharaki EI, Robert C, Reuzé S, et al. Promises and challenges for the implementation of computational medical imaging (radiomics) in oncology. Annals of Oncology. 2017. pp. 1191–1206. doi:10.1093/annonc/mdx034

13. Vickers AJ. Prediction Models: Revolutionary in Principle, But Do They Do More Good Than Harm? Journal of Clinical Oncology. 2011. pp. 2951–2952. doi:10.1200/jco.2011.36.1329

14. Foy JJ, Al-Hallaq HA, Grekoski V, Tran T, Guruvadoo K, Armato SG Iii, et al. Harmonization of radiomic feature variability resulting from differences in CT image acquisition and reconstruction:

assessment in a cadaveric liver. Phys Med Biol. 2020;65: 205008.

15. Andrearczyk V, Depeursinge A, Müller H. Neural network training for cross-protocol radiomic feature standardization in computed tomography. J Med Imaging (Bellingham). 2019;6: 024008.

16. Perone CS, Ballester P, Barros RC, Cohen-Adad J. Unsupervised domain adaptation for medical imaging segmentation with self-ensembling. Neuroimage. 2019;194: 1–11.

17. Mottet N, Bellmunt J, Bolla M, Briers E, Cumberbatch MG, De Santis M, et al. EAU-ESTRO-SIOG Guidelines on Prostate Cancer. Part 1: Screening, Diagnosis, and Local Treatment with Curative Intent. Eur Urol. 2017;71: 618–629.

18. Cornford P, Bellmunt J, Bolla M, Briers E, De Santis M, Gross T, et al. EAU-ESTRO-SIOG Guidelines on Prostate Cancer. Part II: Treatment of Relapsing, Metastatic, and Castration-Resistant Prostate Cancer. Eur Urol. 2017;71: 630–642.

19. Pötter R, Haie-Meder C, Van Limbergen E, Barillot I, De Brabandere M, Dimopoulos J, et al. Recommendations from gynaecological (GYN) GEC ESTRO working group (II): concepts and terms in 3D image-based treatment planning in cervix cancer brachytherapy-3D dose volume parameters and aspects of 3D image-based anatomy, radiation physics, radiobiology. Radiother Oncol. 2006;78: 67–77.

20. Boellaard R, Delgado-Bolton R, Oyen WJG, Giammarile F, Tatsch K, Eschner W, et al. FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. Eur J Nucl Med Mol Imaging. 2015;42: 328–354.

21. Pfaehler E, van Sluis J, Merema BBJ, van Ooijen P, Berendsen RCM, van Velden FHP, et al. Experimental Multicenter and Multivendor Evaluation of the Performance of PET Radiomic Features Using 3-Dimensionally Printed Phantom Inserts. J Nucl Med. 2020;61: 469–476.

22. Yu JB, Beck TF, Anscher MS, Baschnagel AM, Brock KK, Carlson DJ, et al. Analysis of the 2017 American Society for Radiation Oncology (ASTRO) Research Portfolio. Int J Radiat Oncol Biol Phys. 2019;103: 297–304.

23. Luh JY, Albuquerque KV, Cheng C, Ermoian RP, Nabavizadeh N, Parsai H, et al. ACR-ASTRO Practice Parameter for Image-guided Radiation Therapy (IGRT). Am J Clin Oncol. 2020;43: 459–468.

24. Carré A, Klausner G, Edjlali M, Lerousseau M, Briend-Diop J, Sun R, et al. Standardization of brain MR images across machines and protocols: bridging the gap for MRI-based radiomics. Sci Rep. 2020;10: 12340.

25. Schick U, Lucia F, Dissaux G, Visvikis D, Badic B, Masson I, et al. MRI-derived radiomics: methodology and clinical applications in the field of pelvic oncology. Br J Radiol. 2019;92: 20190105.

26. Sachs PB, Hunt K, Mansoubi F, Borgstede J. CT and MR Protocol Standardization Across a Large Health System: Providing a Consistent Radiologist, Patient, and Referring Provider Experience. J Digit Imaging. 2017;30: 11–16.

27. Center for Drug Evaluation, Research. Clinical trial imaging endpoint process standards. 5 Dec 2020 [cited 12 Apr 2021]. Available: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/clinical-trial-imaging-endpoint-process-standards-guidance-industry

28. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. Ann Intern Med. 2015;162: 55–63.

29. Zwanenburg A, Vallières M, Abdalah MA, Aerts HJWL, Andrearczyk V, Apte A, et al. The Image Biomarker Standardization Initiative: Standardised Quantitative Radiomics for High-Throughput

Image-based Phenotyping. Radiology. 2020;295: 328–338.

30. Chung C, Kalpathy-Cramer J, Knopp MV, Jaffray DA. In the Era of Deep Learning, Why Reconstruct an Image at All? J Am Coll Radiol. 2021;18: 170–173.

31. De Man Q, Haneda E, Claus B, Fitzgerald P, De Man B, Qian G, et al. A two-dimensional feasibility study of deep learning-based feature detection and characterization directly from CT sinograms. Med Phys. 2019;46: e790–e800.

32. Gao Y, Tan J, Liang Z, Li L, Huo Y. Improved computer-aided detection of pulmonary nodules via deep learning in the sinogram domain. Vis Comput Ind Biomed Art. 2019;2: 15.

33. Lee H, Huang C, Yune S, Tajmir SH, Kim M, Do S. Machine Friendly Machine Learning: Interpretation of Computed Tomography Without Image Reconstruction. Sci Rep. 2019;9: 15540.

34. Lambin P, Woodruff H. Method of performing radiomics analysis on image data May 21, 2021, Ref: P129643NL01, Application number n°: N2028271. Patent. P129643NL01.

35. Ravishankar S, Ye JC, Fessler JA. Image Reconstruction: From Sparsity to Data-adaptive Methods and Machine Learning. Proc IEEE Inst Electr Electron Eng. 2020;108: 86–109.

36. Ibrahim A, Refaee T, Leijenaar RTH, Primakov S, Hustinx R, Mottaghy FM, et al. The application of a workflow integrating the variable reproducibility and harmonizability of radiomic features on a phantom dataset. PLoS One. 2021;16: e0251147.

37. Ibrahim A, Refaee T, Primakov S, Barufaldi B, Acciavatti RJ, Granzier RWY, et al. The Effects of In-Plane Spatial Resolution on CT-Based Radiomic Features' Stability with and without ComBat Harmonization. Cancers . 2021;13. doi:10.3390/cancers13081848

38. 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. Eur Radiol. 2017;27: 4498–4509.

39. Prayer F, Hofmanninger J, Weber M, Kifjak D, Willenpart A, Pan J, et al. Variability of computed tomography radiomics features of fibrosing interstitial lung disease: A test-retest study. Methods. 2021;188: 98–104.

40. Lee J, Steinmann A, Ding Y, Lee H, Owens C, Wang J, et al. Radiomics feature robustness as measured using an MRI phantom. Sci Rep. 2021;11: 3973.

41. Peerlings J, Woodruff HC, Winfield JM, Ibrahim A, Van Beers BE, Heerschap A, et al. Stability of radiomics features in apparent diffusion coefficient maps from a multi-centre test-retest trial. Sci Rep. 2019;9: 4800.

42. Van Leemput K, Maes F, Vandermeulen D, Suetens P. Automated model-based bias field correction of MR images of the brain. IEEE Trans Med Imaging. 1999;18: 885–896.

43. Masson I, Da-ano R, Lucia F, Doré M. Statistical harmonization can improve the development of a multicenter CT-based radiomic model predictive of nonresponse to induction chemotherapy in laryngeal …. Medical. 2021. Available: https://aapm.onlinelibrary.wiley.com/doi/abs/10.1002/mp.14948?casa_token=4OwnRHSh-CoAAAAA:RgduV-2HMf2Qkof-MOFgRQvjRonz12AtyaGd1rY2i_7XKjD_d_WQVJFKEOtxbyYrRXgH8eH5ExwP0GV3

44. Lucia F, Visvikis D, Vallières M, Desseroit M-C, Miranda O, Robin P, et al. External validation of a combined PET and MRI radiomics model for prediction of recurrence in cervical cancer patients treated with chemoradiotherapy. Eur J Nucl Med Mol Imaging. 2019;46: 864–877.

45. Duron L, Balvay D, Vande Perre S, Bouchouicha A, Savatovsky J, Sadik J-C, et al. Gray-level discretization impacts reproducible MRI radiomics texture features. PLoS One. 2019;14:

e0213459.

46. Gallardo-Estrella L, Lynch DA, Prokop M, Stinson D, Zach J, Judy PF, et al. Normalizing computed tomography data reconstructed with different filter kernels: effect on emphysema quantification. Eur Radiol. 2016;26: 478–486.

47. Introduction — IBSI 0.0.1dev documentation. [cited 23 Mar 2023]. Available: https://ibsi.readthedocs.io/en/latest/01_Introduction.html

48. Image processing — IBSI 0.0.1dev documentation. [cited 23 Mar 2023]. Available: https://ibsi.readthedocs.io/en/latest/02_Image_processing.html

49. Image features — IBSI 0.0.1dev documentation. [cited 23 Mar 2023]. Available: https://ibsi.readthedocs.io/en/latest/03_Image_features.html

50. Radiomics reporting guidelines and nomenclature — IBSI 0.0.1dev documentation. [cited 23 Mar 2023]. Available: https://ibsi.readthedocs.io/en/latest/04_Radiomics_reporting_guidelines_and_nomenclature.html

51. Reference data sets — IBSI 0.0.1dev documentation. [cited 23 Mar 2023]. Available: https://ibsi.readthedocs.io/en/latest/05_Reference_data_sets.html

52. Vuong D, Tanadini-Lang S, Huellner MW, Veit-Haibach P, Unkelbach J, Andratschke N, et al. Interchangeability of radiomic features between [18F]-FDG PET/CT and [18F]-FDG PET/MR. Med Phys. 2019;46: 1677–1685.

53. Bailly C, Bodet-Milin C, Couespel S, Necib H, Kraeber-Bodéré F, Ansquer C, et al. Revisiting the Robustness of PET-Based Textural Features in the Context of Multi-Centric Trials. PLoS One. 2016;11: e0159984.

54. Pfaehler E, Beukinga RJ, de Jong JR, Slart RHJA, Slump CH, Dierckx RAJO, et al. Repeatability of 18 F-FDG PET radiomic features: A phantom study to explore sensitivity to image reconstruction settings, noise, and delineation method. Med Phys. 2019;46: 665–678.

55. van Timmeren JE, Carvalho S, Leijenaar RTH, Troost EGC, van Elmpt W, de Ruysscher D, et al. Challenges and caveats of a multi-center retrospective radiomics study: an example of early treatment response assessment for NSCLC patients using FDG-PET/CT radiomics. PLoS One. 2019;14: e0217536.

56. Orlhac F, Soussan M, Maisonobe J-A, Garcia CA, Vanderlinden B, Buvat I. Tumor texture analysis in 18F-FDG PET: relationships between texture parameters, histogram indices, standardised uptake values, metabolic volumes, and total lesion glycolysis. J Nucl Med. 2014;55: 414–422.

57. Cortes-Rodicio J, Sanchez-Merino G, Garcia-Fidalgo MA, Tobalina-Larrea I. Identification of low variability textural features for heterogeneity quantification of 18F-FDG PET/CT imaging. Rev Esp Med Nucl Imagen Mol. 2016;35: 379–384.

58. Nyflot MJ, Yang F, Byrd D, Bowen SR, Sandison GA, Kinahan PE. Quantitative radiomics: impact of stochastic effects on textural feature analysis implies the need for standards. J Med Imaging (Bellingham). 2015;2: 041002.

59. Galavis PE, Hollensen C, Jallow N, Paliwal B, Jeraj R. Variability of textural features in FDG PET images due to different acquisition modes and reconstruction parameters. Acta Oncol. 2010;49: 1012–1016.

60. van Velden FHP, Kramer GM, Frings V, Nissen IA, Mulder ER, de Langen AJ, et al. Repeatability of Radiomic Features in Non-Small-Cell Lung Cancer [18F]FDG-PET/CT Studies: Impact of Reconstruction and Delineation. Molecular Imaging and Biology. 2016. pp. 788–795. doi:10.1007/s11307-016-0940-2

61. Leijenaar RTH, Nalbantov G, Carvalho S, van Elmpt WJC, Troost EGC, Boellaard R, et al. The effect of SUV discretization in quantitative FDG-PET Radiomics: the need for standardised methodology in tumor texture analysis. Sci Rep. 2015;5: 11075.

62. Shiri I, Rahmim A, Abdollahi H, Ghafarian P, Bitarafan A, Ay M, et al. Radiomics texture features variability and reproducibility in advance image reconstruction setting of oncological PET/CT. EUROPEAN JOURNAL OF NUCLEAR MEDICINE AND MOLECULAR IMAGING. SPRINGER 233 SPRING ST, NEW YORK, NY 10013 USA; 2016. pp. S150–S150.

63. Yan J, Chu-Shern JL, Loi HY, Khor LK, Sinha AK, Quek ST, et al. Impact of Image Reconstruction Settings on Texture Features in 18F-FDG PET. J Nucl Med. 2015;56: 1667–1673.

64. Belli ML, Mori M, Broggi S, Cattaneo GM, Bettinardi V, Dell'Oca I, et al. Quantifying the robustness of [18F]FDG-PET/CT radiomic features with respect to tumor delineation in head and neck and pancreatic cancer patients. Phys Med. 2018;49: 105–111.

65. Desseroit MC, Tixier F, Weber WA. Reliability of PET/CT shape and heterogeneity features in functional and morphologic components of non–small cell lung cancer tumors: a repeatability analysis in a …. Journal of Nuclear. 2017. Available: https://jnm.snmjournals.org/content/58/3/406.short

66. Orlhac F, Soussan M, Chouahnia K, Martinod E, Buvat I. 18F-FDG PET-Derived Textural Indices Reflect Tissue-Specific Uptake Pattern in Non-Small Cell Lung Cancer. PLoS One. 2015;10: e0145063.

67. Ger RB, Meier JG, Pahlka RB, Gay S, Mumme R, Fuller CD, et al. Effects of alterations in positron emission tomography imaging parameters on radiomics features. PLoS One. 2019;14: e0221877.

68. Leijenaar RTH, Carvalho S, Velazquez ER, van Elmpt WJC, Parmar C, Hoekstra OS, et al. Stability of FDG-PET Radiomics features: an integrated analysis of test-retest and inter-observer variability. Acta Oncol. 2013;52: 1391–1397.

69. Mackin D, Fave X, Zhang L, Fried D, Yang J, Taylor B, et al. Measuring Computed Tomography Scanner Variability of Radiomics Features. Invest Radiol. 2015;50: 757–765.

70. Caramella C, Allorant A, Orlhac F, Bidault F, Asselain B, Ammari S, et al. Can we trust the calculation of texture indices of CT images? A phantom study. Medical Physics. 2018. pp. 1529–1536. doi:10.1002/mp.12809

71. Berenguer R, Pastor-Juan MDR, Canales-Vázquez J, Castro-García M, Villas MV, Mansilla Legorburo F, et al. Radiomics of CT Features May Be Nonreproducible and Redundant: Influence of CT Acquisition Parameters. Radiology. 2018;288: 407–415.

72. Zhovannik I, Bussink J, Traverso A, Shi Z, Kalendralis P, Wee L, et al. Learning from scanners: Bias reduction and feature correction in radiomics. Clin Transl Radiat Oncol. 2019;19: 33–38.

73. Balagurunathan Y, Gu Y, Wang H, Kumar V, Grove O, Hawkins S, et al. Reproducibility and Prognosis of Quantitative Features Extracted from CT Images. Transl Oncol. 2014;7: 72–87.

74. Al-Kadi OS. Assessment of texture measures susceptibility to noise in conventional and contrast enhanced computed tomography lung tumour images. Computerized Medical Imaging and Graphics. 2010. pp. 494–503. doi:10.1016/j.compmedimag.2009.12.011

75. 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.

76. Su C, Jiang J, Zhang S, Shi J, Xu K, Shen N, et al. Radiomics based on multicontrast MRI can precisely differentiate among glioma subtypes and predict tumour-proliferative behaviour. Eur Radiol. 2019;29: 1986–1996.

77. Han Y, Xie Z, Zang Y, Zhang S, Gu D, Zhou M, et al. Non-invasive genotype prediction of chromosome 1p/19q co-deletion by development and validation of an MRI-based radiomics signature in lower-grade gliomas. Journal of Neuro-Oncology. 2018. pp. 297–306. doi:10.1007/s11060-018-2953-y

78. Liu Z, Zhang X-Y, Shi Y-J, Wang L, Zhu H-T, Tang Z, et al. Radiomics Analysis for Evaluation of Pathological Complete Response to Neoadjuvant Chemoradiotherapy in Locally Advanced Rectal Cancer. Clin Cancer Res. 2017;23: 7253–7262.

79. Horvat N, Veeraraghavan H, Khan M, Blazic I, Zheng J, Capanu M, et al. MR Imaging of Rectal Cancer: Radiomics Analysis to Assess Treatment Response after Neoadjuvant Therapy. Radiology. 2018;287: 833–843.

80. Nie K, Shi L, Chen Q, Hu X, Jabbour SK, Yue N, et al. Rectal Cancer: Assessment of Neoadjuvant Chemoradiation Outcome based on Radiomics of Multiparametric MRI. Clin Cancer Res. 2016;22: 5256–5264.

81. Min X, Li M, Dong D, Feng Z, Zhang P, Ke Z, et al. Multi-parametric MRI-based radiomics signature for discriminating between clinically significant and insignificant prostate cancer: Cross-validation of a machine learning method. European Journal of Radiology. 2019. pp. 16–21. doi:10.1016/j.ejrad.2019.03.010

82. Viswanath SE, Chirra PV, Yim MC, Rofsky NM, Purysko AS, Rosen MA, et al. Comparing radiomic classifiers and classifier ensembles for detection of peripheral zone prostate tumors on T2-weighted MRI: a multi-site study. BMC Med Imaging. 2019;19: 22.

83. Rai R, Holloway LC, Brink C, Field M, Christiansen RL, Sun Y, et al. Multicenter evaluation of MRI-based radiomic features: A phantom study. Med Phys. 2020;47: 3054–3063.

84. Fortin J-P, Parker D, Tunç B, Watanabe T, Elliott MA, Ruparel K, et al. Harmonization of multi-site diffusion tensor imaging data. NeuroImage. 2017. pp. 149–170. doi:10.1016/j.neuroimage.2017.08.047

85. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics. 2007;8: 118–127.

86. Starmans MPA, van der Voort SR, Phil T, Timbergen MJM, Vos M, Padmos GA, et al. Reproducible radiomics through automated machine learning validated on twelve clinical applications. arXiv [eess.IV]. 2021. Available: http://arxiv.org/abs/2108.08618