# The Image Biomarker Standardization Initiative: Standardized convolutional filters for quantitative radiomics Authors and affiliations

Philip Whybra, Alex Zwanenburg, Vincent Andrearczyk, Roger Schaer, Aditya P Apte, Alexandre Ayotte, Bhakti Baheti, Spyridon Bakas, Andrea Bettinelli, Ronald Boellaard, et al.

Instructions: https://pubs.rsna.org/page/radiology/author-instructions#original_research

# Title Page

## Title

The Image Biomarker Standardization Initiative: Standardized convolutional filters for quantitative radiomics

## Authors and affiliations

Philip Whybra*, Alex Zwanenburg*, Vincent Andrearczyk, Roger Schaer, Aditya P Apte, Alexandre Ayotte, Bhakti Baheti, Spyridon Bakas, Andrea Bettinelli, Ronald Boellaard, Luca Boldrini, Irène Buvat, Gary J R Cook, Florian Dietsche, Nicola Dinapoli, Hubert S Gabryś, Vicky Goh, Matthias Guckenberger, Mathieu Hatt, Mahdi Hosseinzadeh, Aditi Iyer, Jacopo Lenkowicz, Mahdi A L Loutfi, Steffen Löck, Francesca Marturano, Olivier Morin, Christophe Nioche, Fanny Orlhac, Sarthak Pati, Arman Rahmim, Seyed Masoud Rezaeijo, Christopher G Rookyard, Mohammad R Salmanpour, Andreas Schindele, Isaac Shiri, Emiliano Spezi, Stephanie Tanadini-Lang, Florent Tixier, Taman Upadhaya, Vincenzo Valentini, Joost J M van Griethuysen, Fereshteh Yousefirizi, Habib Zaidi, Henning Müller, Martin Vallières, Adrien Depeursinge

* P.W. and A.Z. contributed equally to this work.

From the School of Engineering, Cardiff University, Cardiff, United Kingdom (P.W., E.S.); OncoRay – National Center for Radiation Research in Oncology, Faculty of Medicine and University Hospital Carl Gustav Carus, Technische Universität Dresden, Helmholtz-Zentrum Dresden - Rossendorf, Dresden, Germany (A.Z., S.L.); National Center for Tumor Diseases (NCT), Partner Site Dresden, Germany: German Cancer Research Center (DKFZ), Heidelberg, Germany, Faculty of Medicine and University Hospital Carl Gustav Carus, Technische Universität Dresden, Dresden, Germany, and Helmholtz Association / Helmholtz-Zentrum Dresden - Rossendorf (HZDR), Dresden, Germany (A.Z.); Institute of Informatics, University of Applied Sciences and Arts Western Switzerland (HES-SO), Sierre, Switzerland (V.A., R.S., H.M., A.D.); Department of Medical Physics, Memorial Sloan Kettering Cancer Center, New York, NY, USA (A.P.A., A.I.); Department of Computer Science, Université de Sherbrooke, Sherbrooke, QC, Canada (A.A., M.A.L.L., M.V.); Center for Artificial Intelligence and Data Science for Integrated Diagnostics (AI2D) and Center for Biomedical Image Computing and Analytics (CBICA), University of Pennsylvania, Philadelphia, PA, USA (B.B., S.B., S.P.); Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA (B.B., S.B., S.P.); Department of Radiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA (B.B., S.B., S.P.); Medical Physics Department, Veneto Institute of Oncology IOV - IRCCS, Padua, Italy (A.B., F.M.); Radiology and Nuclear Medicine, Amsterdam UMC, Amsterdam, the Netherlands (R.B.); Fondazione Policlinico Universitario “A. Gemelli” IRCCS, Rome, Italy (L.B., N.D., J.L.); Institut Curie, Université PSL, Inserm

U1288, Laboratoire d'Imagerie Translationnelle en Oncologie, Orsay, France (I.B., C.N., F.O.); Cancer Imaging, School of Biomedical Engineering and Imaging Sciences, King's College London, London, United Kingdom (G.J.R.C., V.G., C.G.R.); Department of Radiation Oncology, University Hospital Zurich, University of Zurich, Zurich, Switzerland (F.D., H.S.G., M.G., S.T.); Department of Radiology, Guy's & St Thomas' NHS Foundation Trust, London, United Kingdom (V.G.); LaTIM, INSERM, UMR 1101, Université de Bretagne-Occidentale, Brest, France (M.Ha., F.T.); Technological Virtual Collaboration (TECVICO Corp.), Vancouver, BC, Canada (M.Ho., M.R.S.); Department of Electrical and Computer Engineering, Tarbiat Modares University, Tehran, Iran (M.Ho.); Department of Radiation Oncology, University of California San Francisco, San Francisco, CA, USA (O.M., T.U.); Departments of Radiology and Physics, University of British Columbia, Vancouver, BC, Canada (A.R.); Department of Medical Physics, Faculty of Medicine, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran. (S.M.R.); Repository Unit, Cancer Research UK National Cancer Imaging Translational Accelerator, United Kingdom (C.G.R.); Department of Integrative Oncology, BC Cancer Research Institute, Vancouver, BC, Canada (M.R.S., F.Y.); Department of Nuclear Medicine, Universitätsklinikum Augsburg, Augsburg, Germany (A.S.); Division of Nuclear Medicine and Molecular Imaging, Geneva University Hospital, Geneva, Switzerland (I.S., H.Z.); Dipartimento Radiodiagnostica, Radioterapia ed Ematologia, Fondazione Policlinico Universitario "A. Gemelli" IRCCS, Rome, Italy (V.V.); Professore ordinario di Radioterapia, Università Cattolica del Sacro Cuore - Milano, Milan, Italy (V.V.); Department of Radiology, The Netherlands Cancer Institute, Amsterdam, the Netherlands (J.J.M.v.G.); Department of Radiology, UMC Utrecht, Utrecht, the Netherlands (J.J.M.v.G.); Centre de recherche du Centre hospitalier universitaire de Sherbrooke (CHUS), Sherbrooke, QC, Canada (M.V.); and Department of Nuclear Medicine and Molecular Imaging, Lausanne University Hospital (CHUV), Lausanne, Switzerland (A.D.).

## Corresponding author

Alex Zwanenburg, alexander.zwanenburg@nct-dresden.de, +49 351 458 7442, Fetscherstraße 74 / PF 41, 01307 Dresden, Germany

## Funding information

The authors were supported by the National Cancer Institute grants P30CA008748 (A.P.A.), U01CA242871 (B.B., S.B.) and U24CA189523 (B.B., S.B.); UK Research & Innovation London Medical Imaging and Artificial Intelligence Centre (G.J.R.C.); UK Wellcome / Engineering and Physical Sciences Research Council Centre for Medical Engineering at King's College London (WT 203148/Z/16/Z) (G.J.R.C.); Cancer Research UK National Cancer Imaging Translational Accelerator awards C1519/A28682 (G.J.R.C., C.G.R.) and C4278/A27066 (V.G.); Swiss National Science Foundation grants 310030_170159 (H.S.G.), CRSII5_183478 (S.T.), 320030_176052 (H.Z.), 205320_179069 (A.D.), and 325230_197477 (A.D.); Natural Sciences and Engineering Research Council of Canada Discovery Grant (RGPIN-2019-06467) (A.R.); UK Engineering and Physical Sciences Research Council (EP/N509449/1) (E.S.); Canada CIFAR AI Chairs Program (M.V.); Swiss Personalized Health Network IMAGINE and QA4IQI projects (A.D.); and RCSO IsNET HECKTOR project (A.D.).

## Manuscript Type

Original research

## Word Count for Text

2323

## Data sharing statement

Data generated by the authors or analyzed during the study are available at: https://github.com/theibsi/data_sets (imaging data) and https://github.com/theibsi/ibsi_2_reference_data (reference response maps and reference feature values). Code used to analyze the data and obtain the results can be found here: https://github.com/theibsi/ibsi_2_data_analysis (commit 778f813).

# Abbreviated title page

## Title

The Image Biomarker Standardization Initiative: Standardized convolutional filters for quantitative radiomics

## Article type

Original research

## Summary Statement

***Eight types of convolutional filters were standardized for radiomics analyses.***

## Key Results

- Mean filters, Laplacian-of-Gaussian filters, Laws kernels, Gabor kernels, as well as separable and non-separable wavelets (including decomposed forms) are now well-defined and standardized for radiomic analyses.
- Applying a filter to an image creates a response map. Thirty-three reference response maps and 323 reference feature values computed from response maps were identified.
- A web-based tool is provided for validating filter implementations in radiomics software packages.

## Abbreviations

IBSI: Image Biomarker Standardization Initiative
ICC: intraclass correlation coefficient
ROI: region of interest

# Abstract

Background: Filters applied to medical imaging can highlight various structures and patterns. Despite their widespread use in radiomics analyses, lack of standardization of convolutional filters negatively affects reproducibility.

Purpose: To standardize convolutional filters for radiomics analyses.

Materials and Methods: This study consisted of three phases. In the first phase, we aimed to establish 36 reference response maps for convolutional filters based on digital phantoms: mean, Laplacian-of-Gaussian, Laws kernels, Gabor kernels, separable and non-separable wavelets (including decomposed forms) and Riesz transformations of convolutional filters. In the second phase, we aimed to find reference values for 396 intensity-based features computed from response maps of 22 filter and image processing configurations, based on computed tomography (CT) imaging. Afterwards, reproducibility of standardized convolutional filters and feature values was assessed during a validation phase, using a public dataset of multi-modal imaging (CT, FDG-PET, T1w-MR) from 51 patients with soft-tissue sarcoma.

Results: In phase 1, 15 teams from 7 countries were able to find reference response maps for 33 of 36 filter configurations. In phase 2, 11 teams were able to find reference feature values for 323 of 396 features. Consensus on reference feature values for Riesz transformations was not established. During the validation phase, 458 of 486 features were found to be reproducible among 9 teams. Coefficient of variation and quartile coefficient of dispersion features were found to be poorly reproducible for band- and high-pass filters.

Conclusion: Eight types of convolutional filters for radiomics were standardized and reference responses and reference feature values for verification and calibration of radiomics software packages were obtained. A web-based tool is available for checking compliance of radiomics software with the reference.

# Main body

## Introduction

Radiomics is the high-throughput processing of medical images to support clinical decision making (1). The intermediate steps in this process, and variability therein, are known to affect the reproducibility of the resulting decision support tools. The Image Biomarker Standardization Initiative (IBSI) previously established standards for digital image processing and computation of hand-crafted, quantitative radiomic features (2). This has improved interchangeability between IBSI-compliant radiomics software packages, provided that the same image processing and feature parameters are used (3,4).

Filters are frequently used in radiomics studies as they allow characteristics to be emphasized such as edges (e.g., sharp tumor boundaries), vessels, blobs, or holes in medical imaging. However, the use of Laplacian-of-Gaussian, wavelets and other filters has

not been standardized, and their use has been found to be poorly reproducible (5). Many of these filters are convolutional filters, hence the focus of this work. With convolution, filters are systematically slid across the entire image, yielding a response map that spatially locates the image characteristics mentioned earlier. Radiomic features, such as mean intensity, can then be computed from the region of interest (ROI) within the response map.

Here, the IBSI aimed to improve reproducibility of convolutional filters for radiomics by: *(a)* establishing definitions for convolutional filters, including commonly used ones such as wavelets; *(b)* integrating the convolutional filters into the previously established general radiomics image processing scheme (2); *(c)* providing datasets, associated reference response maps and reference feature values, as well as tools for verification and calibration of radiomics software packages; and *(d)* updating the IBSI guidelines for methodological reporting to include details concerning convolutional filters.

## Materials and Methods

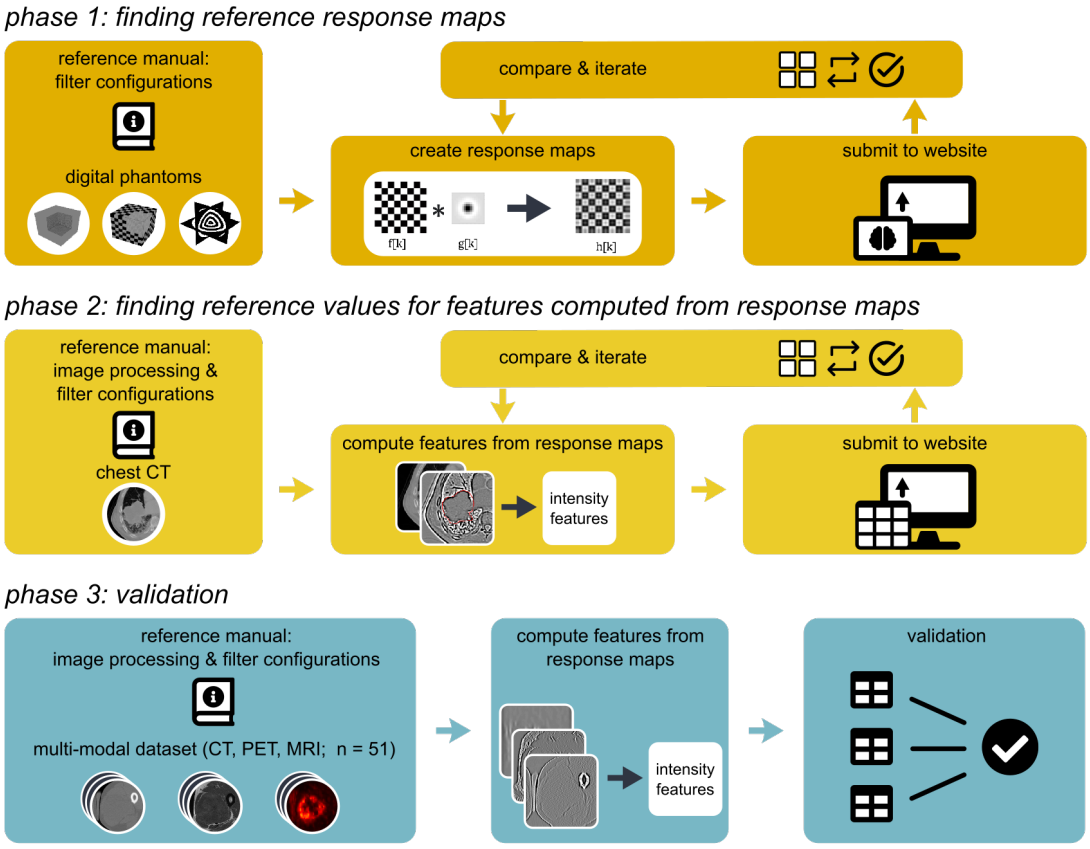


**Figure 1:** Study overview. The study is divided into three phases. In the first phase, convolutional filters were applied to digital phantoms to identify reference response maps. In the second phase, reference values were identified for intensity-based features computed from response maps of a chest CT image. In the third phase, the results of the first two phases were validated using a multi-modal dataset. Unlike the first two phases, the validation phase is not iterative.

## Study design

This standardization effort was divided into three phases (Fig. 1). During the first two phases the implementation and use of convolutional filters were standardized. Phase 1 concerned the creation of reference response maps, i.e., the expected result of applying a convolutional filter with specific parameters to an image. In phase 2, convolutional filters were integrated into a radiomics workflow for the purpose of finding reference values for radiomic features computed from response maps. In phase 3, we assessed whether standardization of convolutional filters resulted in reproducible feature values.

## Convolutional filters

Convolutional filters transform an image to a response map by convolution. These filters consist of numerical weights that are pre-defined or parameterized in the spatial domain or in the frequency (Fourier) domain. Several convolutional filters were assessed, i.e., mean filter, Laplacian-of-Gaussian filter, Laws kernels, Gabor kernels, separable and non-separable wavelets, and Riesz transformations of convolutional filters, see Fig. 2. Further details are supplied in supplementary note 1 and in the reference manual.

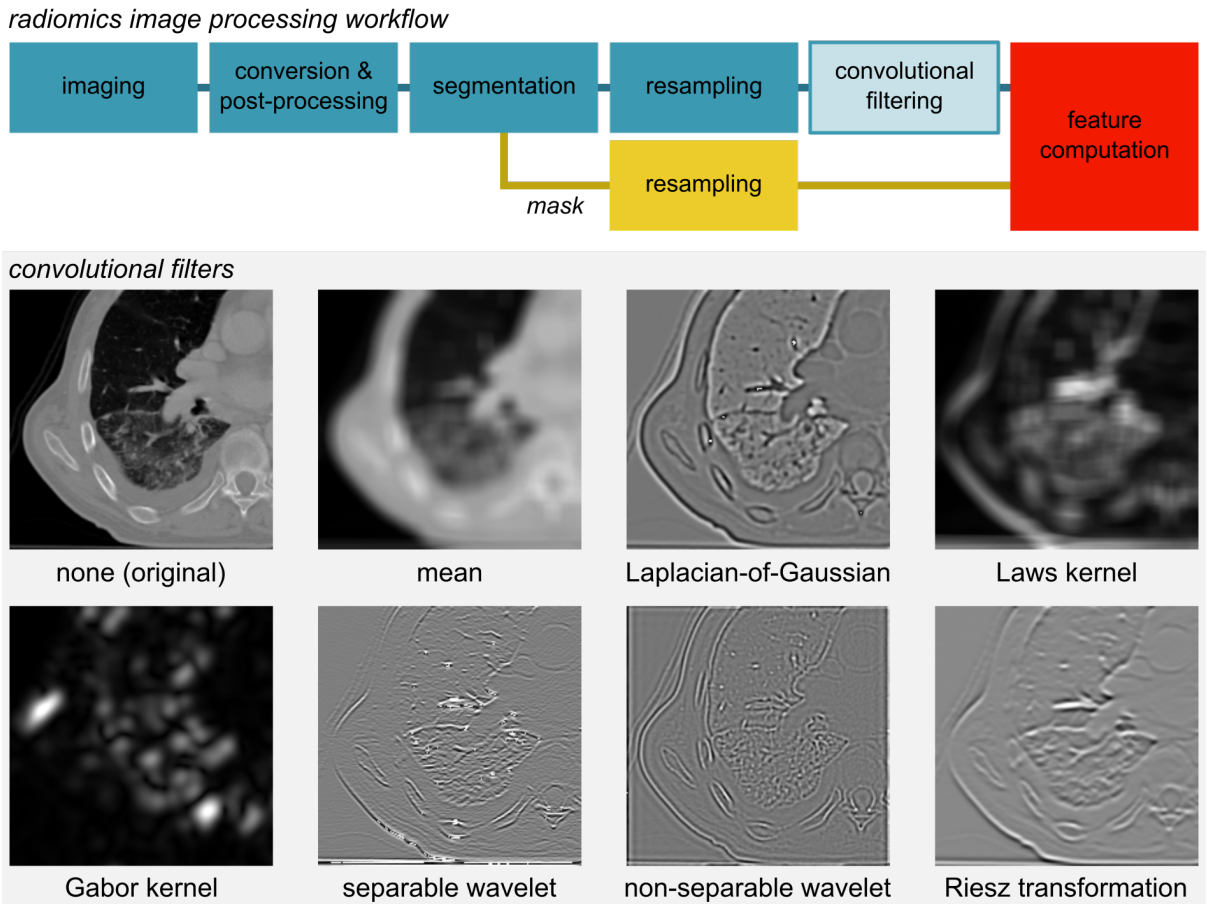


**Figure 2:** Convolutional filtering is positioned after resampling in the overall radiomics workflow (top panel). This workflow starts with an image that is obtained from a repository or archiving system in a digital format, such as DICOM. Then the image is optionally converted (e.g., from PET activity to standardized uptake values) and post-processed (e.g., MR bias-field correction). Segmentation masks are either loaded in a digital format, or automatically created. Both image and segmentation masks are then optionally resampled. Response

maps are created by optionally filtering the image. Both response map and segmentation mask are then used to compute hand-crafted radiomic features. This study attempts to standardize several types of convolution filters (bottom panel). The original CT image is shown for reference. Decomposition of separable and non-separable wavelets is not shown. The image processing scheme and convolutional filters are described in more detail in the reference manual.

## Participating teams

Teams of radiomics researchers were invited to participate in the IBSI. Participation was voluntary and open for the duration of the study. Teams were eligible to participate if they *(a)* developed their own radiomics software, and *(b)* their software is IBSI compliant. A team was not required to participate in all phases of the study.

## Phase 1: Establishing reference response maps

In phase 1, several digital phantoms were used, and thirty-six convolutional filter configurations were defined to establish reference response maps (supplementary note 2). Teams computed response maps for each filter configuration and uploaded these to a central website (https://ibsi.radiomics.hevs.ch/; supplementary note 7).

The level of consensus for each response map was assessed using the same metrics as previously (2): *(a)* by the number of teams that matched the tentative reference response map (supplementary note 4); and (*b*) the previous number divided by the number of teams that contributed a response map. Level of consensus was then: *none*, if the tentative reference response map was not produced by over 50% of contributing teams; *weak*, match between fewer than three teams; *moderate*, three to five; *strong*, six to nine; *very strong*, ten or more.

## Phase 2: Defining feature reference values

Convolutional filtering was integrated into the general radiomics image processing scheme (Fig. 2). Image processing and convolutional filter configurations were then defined for each filter. Both 2D and 3D filter configurations were created, yielding twenty-two configurations in total (supplementary note 3). Teams computed a response map for each configuration from a publicly available CT image of a patient with lung cancer (6). Eighteen intensity-based features were then computed from the gross tumor volume ROI in each response map (supplementary note 5). After computing feature values, teams uploaded their results to the website. The level of consensus for feature values was assessed using the same metrics as in phase 1 by using contributed values for each feature as input and comparing matches within a tolerance margin (supplementary note 5).

## Phase 3: Validation

After completing phases 1 and 2, teams were asked to compute intensity-based features from the gross tumor volume segmentation in response maps of a multimodality imaging cohort (co-registered CT, $^{18}$F-FDG-PET, and T1-weighted MRI). This cohort consisted of 51 patients with soft-tissue sarcoma obtained from The Cancer Imaging Archive (7–9). PET and MRI were pre-processed to ensure that conversion of PET activity concentration to

standardized uptake value and MR bias field correction and normalization could not affect validation results (supplementary note 3). Nine image processing and convolutional filter configurations were specified for each modality. Teams were blinded to the results submitted by other teams. After submitting results, obvious configuration errors were reported back to the submitting team.

Reproducibility was assessed using the two-way, random-effects, single rater absolute agreement intraclass correlation coefficient (ICC) (10). Based on Koo and Li (11), reproducibility of each feature was assigned to one of the following categories, based on the lower bound of the 95% confidence interval of the ICC (12): poor, lower bound less than 0.50; moderate, between 0.50 and 0.75; good, between 0.75 and 0.90; and excellent, greater than 0.90.

# Results

Fifteen teams from seven countries participated in the first phase, eleven teams in the second phase, and nine teams in the validation phase. Eleven teams had developed publicly available software: CaPTk, CERR, FAST, LIFEx, MIRAS, MIRP, moddicom, S-IBEX, SPAARC, and the McGill and Université de Sherbrooke teams (see supplementary note 6).

Of the thirty-six response maps that were assessed in the first phase, moderate or better consensus was found for seventeen (47%) at the initial timepoint (Fig. 3). At the final time-point, moderate or better consensus was achieved for thirty-three (92%) configurations, of which twenty-four (67%) were very strong. Full consensus was reached for configurations corresponding to mean filters, Laplacian-of-Gaussian filters, Laws kernels, Gabor kernels, as well as separable and non-separable wavelets (including decomposed forms). No or only weak consensus was achieved for three (8%) configurations, corresponding to configurations involving Riesz transformations (Fig. E1).

At the initial time point of the second phase, moderate or better consensus was achieved for 198 (50%) of 396 features, aggregated over twenty-two different filter configurations. At the final time-point 323 (82%) features had at least moderate consensus. Again, full consensus was reached for features computed from response maps of mean filters, Laplacian-of-Gaussian filters, Laws kernels, Gabor kernels, as well as separable and non-separable wavelets (including decomposed forms), except for the quantile coefficient of dispersion feature for three-dimensional non-separable wavelets. No consensus was established for features based on (steered) Riesz transformations (Fig. E2) because too few teams submitted values for these features.

In summary, eight types of convolutional filters were standardized in the first two phases. The reproducibility of features from response maps created by these filters was assessed in the third phase. Here, good to excellent reproducibility of features was found for 458 (92%) of 486 features. Overall, nineteen (4%) features were poorly reproducible, and were found for Laplacian-of-Gaussian, separable and non-separable wavelet filters. Most of these features were either coefficient of variation or quartile coefficient of dispersion features that represented eight and nine of nineteen features, respectively. A list of poorly reproducible features is provided in Table E1. No dependence on imaging modality could be observed.

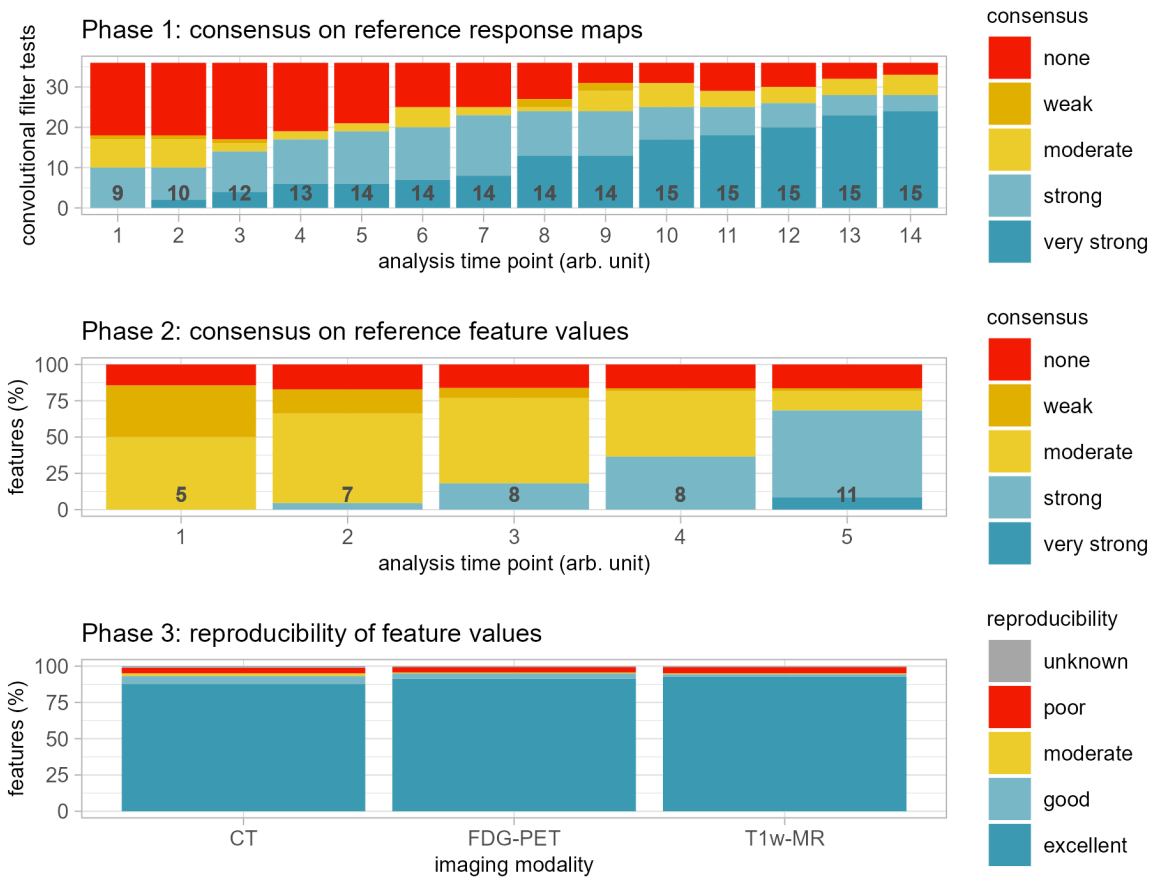


**Figure 3:** Results overview. In phase 1, participating teams computed thirty-six response maps of convolutional filters according to predefined configurations. These response maps were compared, and consensus was measured. Teams updated their implementations iteratively, which led to an improvement of consensus over time (arbitrary unit, the entire process took twenty-four months). Consensus strength was based on matching the voxel-wise difference between response maps and the tentative reference response map within a tolerance: weak, match between fewer than three teams; moderate, three to five; strong, six to nine; very strong, ten or more; none, 50% of the teams or more did not match. The number of participating teams at each timepoint is shown. In phase 2, participating teams computed 396 features from response maps of convolutional filters according to predefined filter and image processing configurations. As in phase 1, teams updated their implementations iteratively. However, unlike phase 1, improvement in consensus was mostly due to more teams enrolling over time (arbitrary unit, the entire process took eight months). Consensus strength was based on the number of teams matching the tentative reference feature value within a tolerance and was assigned according to the same categories as in phase 1. In phase 3, reproducibility of features computed from response maps was validated. Teams computed 486 features from a public dataset of fifty-one patients with soft-tissue sarcoma that were scanned using CT, FDG-PET, and T1w-MR imaging. Reproducibility was assessed using the lower bound of the 95% confidence interval of an intraclass correlation coefficient: poor, lower bound less than 0.50; moderate, between 0.50 and 0.75; good, between 0.75 and 0.90; excellent, greater than 0.90; and unknown, computed by fewer than two teams.

# Discussion

Convolutional filters are commonly used in radiomics studies. However, due to lack of proper consensus-based reference implementations, features computed from response maps provided by these filters are difficult to reproduce (5). In this study, fifteen teams from seven countries collaborated to remedy this situation by providing reference response maps, reference feature values and reference documentation. We were able to standardize and validate eight different filter types: mean, Laplacian-of-Gaussian, Laws kernels, Gabor filters, and separable and non-separable wavelet filters in both undecomposed and decomposed forms.

The presented results complement the previous results of the Image Biomarker Standardization Initiative (2). That work focused on standardizing both the image processing scheme for radiomics and a large set of radiomic features. It aimed to improve reproducibility of radiomics studies by mitigating the effect of using different radiomics software packages, and by providing a common framework for describing methodological details. This work adds to the previous by standardizing the use of convolutional filters that are frequently used in radiomics.

This work has several implications: first, we found that most types of convolutional filters are not trivial to implement reproducibly across implementations, as evidenced by the initial lack of consensus on response maps in phase 1. Thus, we must assume that current clinical or research radiomics software in advanced image analysis workstations that offer convolutional filters yield feature values that are not reproducible in external settings, thereby hampering external validation and further clinical studies.

The second implication is that “IBSI-compliant” software is now expected to reproduce the reference response maps and reference feature values found in this study, insofar as convolutional filters are available in the software, in addition to the reference feature values that were previously found (2). Compliance may be checked using web-based tools (https://ibsi.radiomics.hevs.ch/), or manually using the provided reference response maps and feature values. Compliant software is expected to produce response maps where every voxel deviates from the reference response map by at most 1% of the range of intensity values of the reference response map. A feature value should be reproduced within the tolerance margin around the reference feature value.

Third, even though we contextualized our efforts within radiological imaging, our work is relevant for quantitative image analysis in general, including digital pathology. Finally, the IBSI reporting checklist for methodological details is now replaced by an updated version that includes reporting items for convolutional filters.

We standardized mean filters, Laplacian-of-Gaussian filters, Laws kernels, Gabor filters, and separable and non-separable wavelets. We also assessed Riesz transformations of convolutional filters in phases 1 and 2. Despite their attractive characteristics from a signal processing point-of-view, implementing Riesz transformations was not trivial and too few teams did so. Therefore, we could not provide reference response maps and reference values for Riesz transformations, and we did not assess these during the validation phase.

As Riesz transformations are only very rarely used in radiomics studies, the impact should be minimal.

Most features computed from response maps were reproducible after validation. However, several features could not be computed in a reproducible manner, notably the coefficient of variation and quartile coefficient features in conjunction with high- and band-pass convolutional filters. Such filters are characterized by a response map with a mean intensity of zero. Coefficient of variation relies on a division by the mean intensity in the ROI that is close to zero. The quartile coefficient of dispersion relies on division by the sum values of the first and third quartiles of intensities in the ROI, which can be very similar except for their sign, as the distribution is again centered around zero. In the presence of high- and band-pass convolutional filters, the division operation present in both features led to otherwise negligible numeric differences between teams becoming relevant. This led to poor reproducibility. Therefore, these features should not be used in combination with high- and band-pass filters.

In this work, we standardized intensity-based statistical features computed from response maps, but not other types of features. Particularly, morphological features are mostly redundant as these are based on segmentation masks that are explicitly not altered by convolutional filtering. Most texture features, in our estimation, would be too abstract to allow for interpretation. We recommend computing only intensity-based features from response maps, as otherwise hundreds to thousands of features might be added to a radiomics analysis. This makes the process of creating a generalizable and interpretable radiomics model more difficult in the typical setting where at most a few hundred images are available for analysis (13).

A limitation of this work is its necessarily restricted scope. Compliance with IBSI reference values helps to improve reproducibility of radiomic features (3,4). However, a radiomics study also comprises image acquisition, reconstruction, segmentation, and data analysis steps (14,15), which we did not address here or in our previous work. Differences in, for example, image acquisition protocols are known to affect the appearance of an image, and therefore also reproducibility of radiomic features (16). Such effects can be reduced by harmonization and cross-calibration of scanners and protocols (17) and post-hoc techniques such as perturbation (18,19), batch normalization (20), and other methods (21).

Another limitation is that participation in the IBSI does not guarantee that a particular software package is compliant with the IBSI reference. Changes introduced in software (3), or design choices may limit compliance (22). It is therefore important that developers integrate comparison with the reference standard into their testing framework, and where absent, for users to assess whether the software is compliant.

The IBSI has so far focused on radiomics using hand-crafted features, and with this work offers a comprehensive reference standard for their computation. We recognize that there are more features and other filters than the ones we have standardized so far. These are not implemented often and will be hard to standardize for that reason. In the future, the IBSI will focus on deep learning applications of radiomics, with an aim to provide a reference standard for image pre-processing.

In conclusion, we standardized the use of eight types of convolutional filters for radiomics, and produced reference response maps, reference feature values and tools for verification and calibration of radiomics software packages.

# References


1. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. Radiology. 2016;278(2):563–577.

2. Zwanenburg A, Vallières M, Abdalah MA, et al. The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping. Radiology. 2020;295(2):328–338.

3. Fornacon-Wood I, Mistry H, Ackermann CJ, et al. Reliability and prognostic value of radiomic features are highly dependent on choice of feature extraction platform. Eur Radiol. 2020;30(11):6241–6250.

4. Bettinelli A, Marturano F, Avanzo M, et al. A Novel Benchmarking Approach to Assess the Agreement among Radiomic Tools. Radiology. 2022;211604.

5. Bogowicz M, Leijenaar RTH, Tanadini-Lang S, et al. Post-radiochemotherapy PET radiomics in head and neck cancer - The influence of radiomics implementation on the reproducibility of local control tumor models. Radiother Oncol. 2017;125(3):385–391.

6. Lambin P. Data from: Radiomics Digital Phantom. CancerData; 2016. doi: 10.17195/candat.2016.08.1.

7. Clark K, Vendt B, Smith K, et al. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. J Digit Imaging. 2013;26(6):1045–1057.

8. Vallières M, Freeman CR, Skamene SR, El Naqa I. A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities. Phys Med Biol. 2015;60(14):5471–5496.

9. Vallières M, Freeman CR, Skamene SR, El Naqa I. Data from: A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities. The Cancer Imaging Archive; 2015. doi: 10.7937/K9/TCIA.2015.7GO2GSKS.

10. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull. 1979;86(2):420–428.

11. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med. 2016;15(2):155–163.

12. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. Psychol Methods. 1996;1(1):30–46.

13. Demircioğlu A. The effect of preprocessing filters on predictive performance in radiomics. Eur Radiol Exp. 2022;6(1):40.

14. Zwanenburg A. Radiomics in nuclear medicine: robustness, reproducibility, standardization, and how to avoid data analysis traps and replication crisis. Eur J Nucl Med Mol Imaging. 2019;46(13):2638–2655.

15. 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(1):91.

16. Berenguer R, Pastor-Juan MDR, Canales-Vázquez J, et al. Radiomics of CT Features May Be Nonreproducible and Redundant: Influence of CT Acquisition Parameters. Radiology. 2018;288(2):407–415.

17. Sullivan DC, Obuchowski NA, Kessler LG, et al. Metrology Standards for Quantitative Imaging Biomarkers. Radiology. 2015;277(3):813–825.

18. Zwanenburg A, Leger S, Agolli L, et al. Assessing robustness of radiomic features by image perturbation. Sci Rep. 2019;9(1):614.

19. Teng X, Zhang J, Zwanenburg A, et al. Building reliable radiomic models using image perturbation. Sci Rep. 2022;12(1):1–10.

20. Orlhac F, Frouin F, Nioche C, Ayache N, Buvat I. Validation of A Method to Compensate Multicenter Effects Affecting CT Radiomics. Radiology. 2019;291(1):53–59.

21. Mali SA, Ibrahim A, Woodruff HC, et al. Making Radiomics More Reproducible across Scanner and Imaging Protocol Variations: A Review of Harmonization Methods. J Pers Med. 2021;11(9):842.

22. Wright DE, Cook C, Klug J, Korfiatis P, Kline TL. Reproducibility in medical image radiomic studies: contribution of dynamic histogram binning. arXiv [eess.IV]. 2022. http://arxiv.org/abs/2211.05241.