Radiology

The Image Biomarker Standardization Initiative:
Standardized Convolutional Filters for Reproducible Radiomics and Enhanced Clinical Insights

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

Filters are commonly used to enhance specific structures and patterns in images, such as vessels or peritumoral regions, to enable clinical insights beyond the visible image using radiomics. However, their lack of standardization restricts reproducibility and clinical translation of radiomics decision support tools. In this special report, teams of researchers who developed radiomics software participated in a three-phase study (September 2020 to December 2022) to establish a standardized set of filters. The first two phases focused on finding reference filtered images and reference feature values for commonly used convolutional filters: mean, Laplacian of Gaussian, Laws and Gabor kernels, separable and nonseparable wavelets (including decomposed forms), and Riesz transformations. In the first phase, 15 teams used digital phantoms to establish 33 reference filtered images of 36 filter configurations. In phase 2, 11 teams used a chest CT image to derive reference values for 323 of 396 features computed from filtered images using 22 filter and image processing configurations. Reference filtered images and feature values for Riesz transformations were not established. Reproducibility of standardized convolutional filters was validated on a public data set of multimodal imaging (CT, fluorodeoxyglucose PET, and T1-weighted MRI) in 51 patients with soft-tissue sarcoma. At validation, reproducibility of 486 features computed from filtered images using nine configurations × three imaging modalities was assessed using the lower bounds of 95% CIs of intraclass correlation coefficients. Out of 486 features, 458 were found to be reproducible across nine teams with lower bounds of 95% CIs of intraclass correlation coefficients greater than 0.75. In conclusion, eight filter types were standardized with reference filtered images and reference feature values for verifying and calibrating radiomics software packages. A web-based tool is available for compliance checking.

© RSNA, 2024

Supplemental material is available for this article.

Abbreviation

IBSI = Image Biomarker Standardization Initiative

Summary

Standardizing convolutional filters that enhance specific structures and patterns in medical imaging enables reproducible radiomics analyses, improving consistency and reliability for enhanced clinical insights.

Essentials

- Fifteen international teams who developed radiomics software defined and standardized eight convolutional filter types for radiomic analyses: mean, Laplacian of Gaussian, Laws and Gabor kernels, separable and nonseparable wavelets (undecomposed, decomposed forms).
- Thirty-three reference filtered images and 323 reference feature values computed from filtered images were established to standardize radiomics analyses across various imaging modalities.
- A website-based tool is available for checking compliance of radiomics software.

Radiomics involves the high-throughput extraction of quantitative features from medical images to support clinical decision making (1,2). Relatively few radiomics decision support tools have entered the clinic because their clinical translation is restricted by both the lack of standardization of the extraction process and by lack of quality clinical evidence for their efficacy (3). Focusing on software-related aspects of the extraction process, the Image Biomarker Standardization Initiative (IBSI) previously established modality-independent standards for digital image processing and computation of handcrafted, quantitative radiomic features (4). This improved reproducibility and interchangeability of IBSI-compliant radiomics software packages, provided that the extraction process is configured the same between packages (5,6).

Filters (Table) are frequently used in radiomics analyses to enhance and quantify potentially clinically relevant characteristics and textures in medical images, such as the peritumoral region, blood vessels, contrast agent uptake, degree of calcification, or fibrosis, among others (7) (Appendix S1). For example, Beuque et al (8) applied a Laplacian of Gaussian filter to contrast-enhanced mammography to classify breast lesions into benign and malignant cases. The Laplacian of Gaussian filter enhanced the regions with contrast agent uptake, amplifying the signal, and therefore was found to be important for classifying lesion malignancy. Many filters, including the Laplacian of Gaussian filter used by Beuque et al, rely on convolution. Convolution is a mathematical operation where a filter (here an array of numbers) is systematically slid across the entire image (Fig 1). This process yields a filtered image that enhances and spatially locates potentially relevant image characteristics. However, the computational implementation of these filters has not been standardized, and quantitative features extracted from regions of interest in the filtered images were found to be poorly reproducible between radiomics software packages (9) (Fig 2). Consequently, radiomics decision support tools that incorporate features computed from regions of interest inside filtered images may be difficult to reproduce, validate, and translate clinically.

Because convolutional filters are both important and commonly used, the IBSI aimed to improve reproducibility of radiomics decision support tools involving these filters and to facilitate their clinical translation through a modality-independent software standardization process by establishing definitions for convolutional filters, including commonly used ones such as wavelets and Laplacian of Gaussian filters; by integrating convolutional filters into the previously established general radiomics image processing scheme (4); and by providing data sets, associated reference filtered images, reference feature values, and tools for verification and calibration of radiomics software packages.

Materials and Methods

Study Design

This standardization effort was divided into three phases (Fig 3) and was conducted between September 2020 and December 2022. During the first two phases the implementation and use of convolutional filters were standardized. Phase 1 concerned the creation of reference filtered images (ie, 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 filtered images. In phase 3, we assessed whether standardization of convolutional filters resulted in reproducible feature values. A website (*https://ibsi.radiomics.hevs.ch/*; Appendix S2) was created to coordinate the study.

Convolutional Filters

Convolutional filters transform an image to a filtered image by convolution. These filters consist of numerical weights that are predefined or parameterized in the spatial domain or in the frequency (Fourier) domain. Several convolutional filters were assessed (ie, mean filter, Laplacian of Gaussian filter, Laws kernels, Gabor kernels, separable and nonseparable wavelets, and Riesz transformations of convolutional filters; Fig 1). Details are supplied in Appendix S1 and in Depeursinge et al (10).

Participating Teams

Teams of radiomics researchers were invited to participate in this study. In addition to all teams that had previously participated in the IBSI (4), invitations were extended to any other team that indicated their desire to participate by using the main IBSI website (*https://theibsi.github.io/*) and by forms of personal communication. Participation was voluntary and open for the duration of the study. Teams were eligible to participate if they developed their own radiomics software and their software was compliant with the previous IBSI reference standard. Teams were not required to participate in all phases of the study.

Phase 1: Establishing Reference Filtered Images

In phase 1, five digital three-dimensional phantoms were used (Appendix S3), as follows: an orientation phantom to verify consistency of image orientation within the software of each team; an impulse phantom with a single, central, active voxel; a sphere phantom consisting of concentric spherical shells; a phantom with a checkerboard pattern; and a phantom with line patterns. Thirty-six convolutional filter configurations were defined to
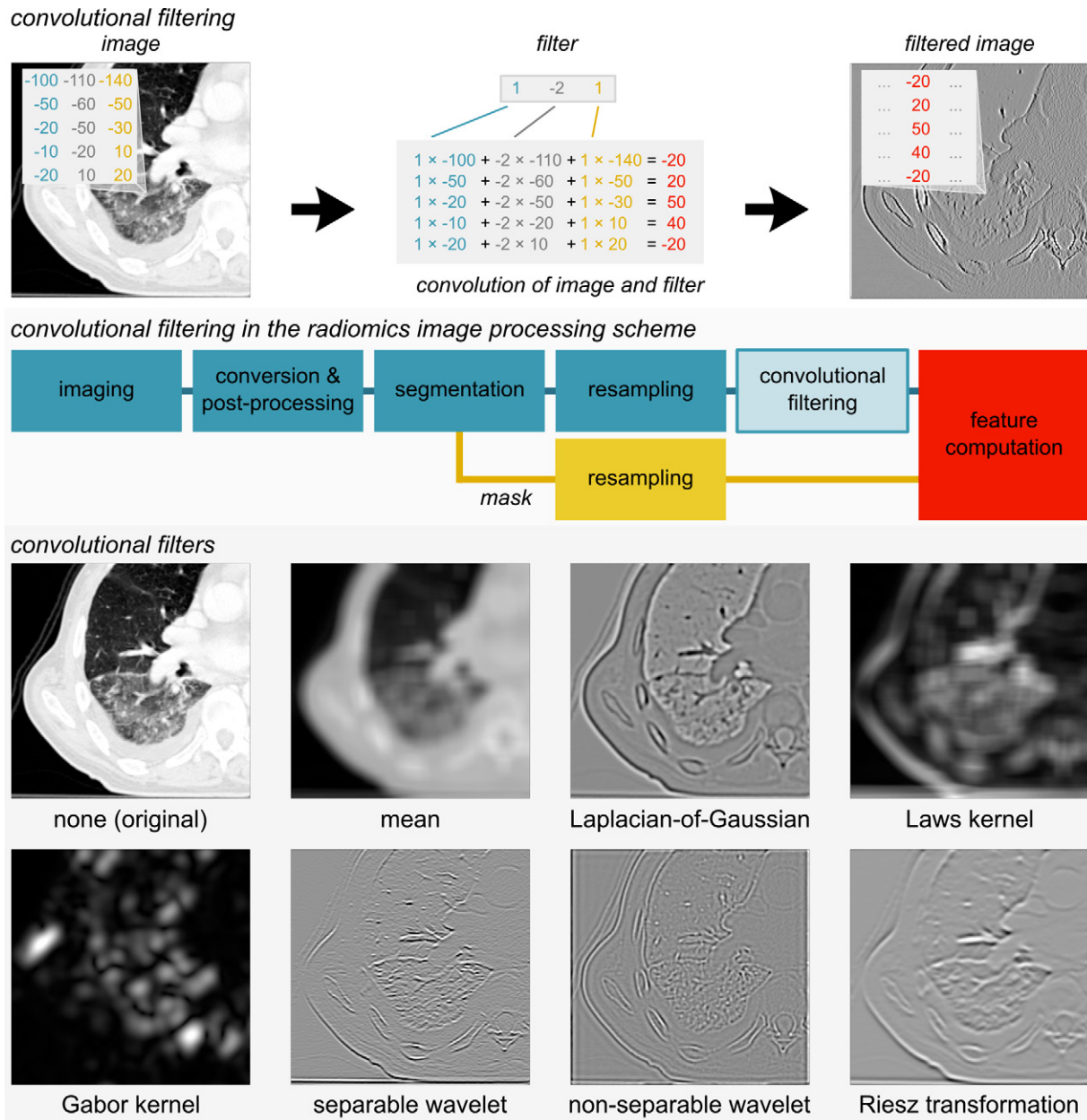
Figure 1: Overview of convolutional filters. An image is filtered using convolution to create a filtered image (top). Each image consists of values. At convolution, a filter with three weights (1.0, −2.0, 1.0) is slid across the image, and adjacent image values are multiplied with the corresponding filter values and summed to create a response value for each position in the image. Convolutional filtering is positioned after resampling in the overall radiomics image processing scheme (center). This workflow starts with an image obtained from a repository or archiving system in a digital format. The image is optionally converted (eg, from PET activity to standardized uptake values) and undergoes postprocessing (eg, MRI bias-field correction). Segmentation masks are either loaded in a digital format or automatically created. Both image and segmentation masks are optionally resampled. Filtered images are created by filtering the image. Filtered images and segmentation masks are then used to compute radiomic features. This study attempts to standardize several types of convolution filters (bottom). The original CT image is shown for reference.

establish reference filtered images (Appendix S4). Teams computed filtered images for each filter configuration and uploaded these to the study website.

The level of consensus for each filtered image was assessed using the same metrics as were used previously (4), as follows: by the number of teams that matched the tentative reference filtered image (Appendix S5) (ie, had filtered images with voxel-wise differences with the tentative reference filtered image that were less than 1% of the intensity range of the tentative reference filtered image for all voxels) and the previous number divided by the number of teams that contributed a filtered image. The levels of consensus were as follows: none, if the tentative reference filtered image was not produced by more than 50% of contributing teams; weak, match between fewer than three teams; moderate, match between three and five teams; strong, match between six and nine teams; and very strong, match between at least 10 teams.

Phase 2: Defining Feature Reference Values
Convolutional filtering was integrated into the general radiomics image processing scheme (Fig 1). Image processing and

Figure 2: Three filters are used to quantify different characteristics of the peritumoral region in a chest CT, with an out-of-plane tumor. For each filter, mean and maximum intensity are computed within the segmentation masks in three filtered images. The standardized filtered image was created by applying a standardized filte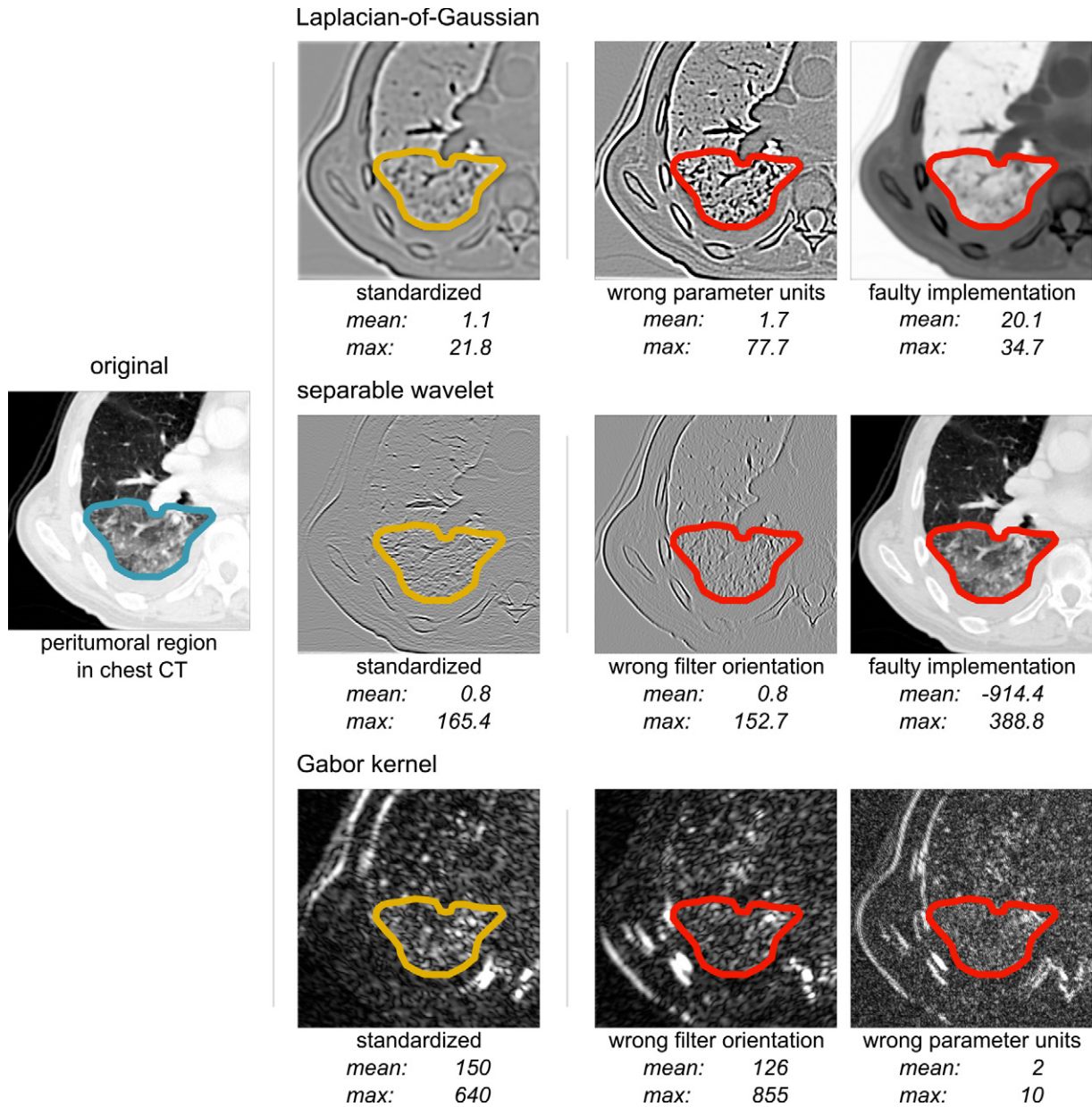r to the original image. The other two filtered images resulted from filter implementations that were not standardized. The Laplacian of Gaussian filter is used to quantify the presence of edges and highlight fine details. The scale of the filter is 2.0 mm, and it is truncated at 8.0 mm. The nonstandardized filters use 2.0 voxels and truncate at one filter scale (2.0 mm). Separable wavelets are designed to quantify image contents for different frequency bands, though in many radiomics analyses they are used to quantify edges. A pair of low-pass and high-pass wavelet kernels is used to filter the image, highlighting edges in the lateral direction. The nonstandardized filters either use an incorrect orientation (ie, low-pass and high-pass kernels were swapped) or are faulty because the first kernel is used for all directions (ie, a pair of low-pass-low-pass wavelet kernels). Gabor filters are used to quantify directional structures (eg, fibrosis and bronchi). The standardized filter used scale and wavelength parameters of 2.0 mm and was oriented under 30°. The nonstandardized filters use an incorrect orientation or express parameters in 2.0 voxels.

convolutional filter configurations were then defined for each filter. Both two- and three-dimensional filter configurations were created, yielding 22 configurations in total (Appendix S4). Teams computed a filtered image for each configuration from a publicly available chest CT image of a patient with lung cancer (11). Eighteen intensity-based features were computed from the gross tumor volume region of interest in each filtered image (Appendix S6). Thus, a total of 396 features could be computed (18 features × 22 configurations). After computing feature values,

teams uploaded their results to the study 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 (Appendix S6).

Phase 3: Validation
After completing phases 1 and 2, teams were asked to compute intensity-based features from the gross tumor volume segmentation

phase 1: establishing reference filtered images
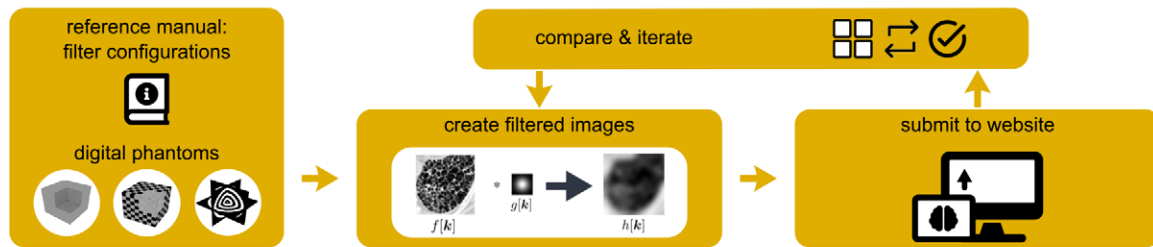
phase 2: defining feature reference values
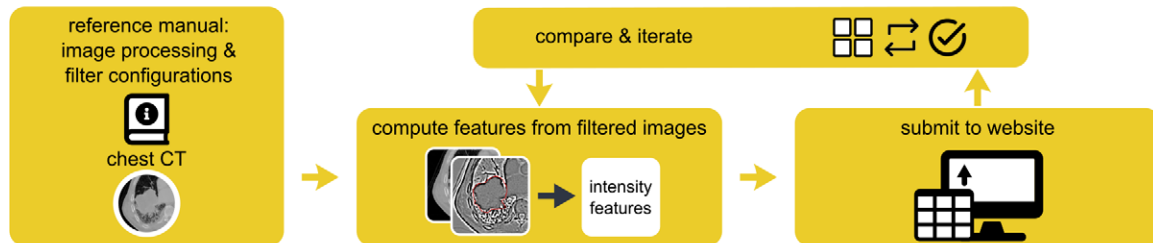
phase 3: validation
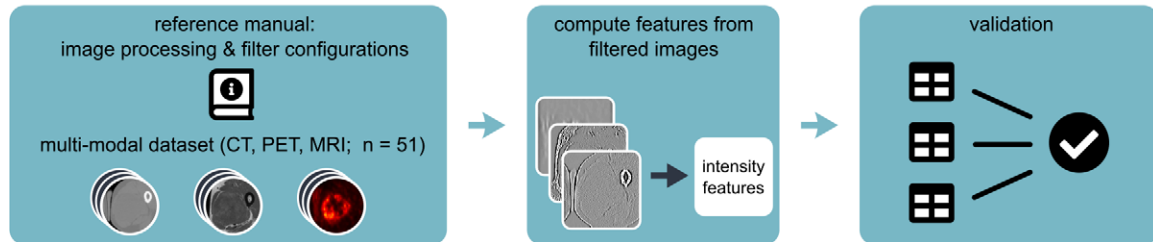
Figure 3: Study overview. Several figure elements adapted, under a CC BY 4.0 license, from reference 10.

in filtered images of a multimodality imaging cohort (co-registered CT, fluorine 18 fluorodeoxyglucose PET, and T1-weighted MRI). This cohort consisted of 51 patients with soft-tissue sarcoma obtained from the Cancer Imaging Archive (12–14). PET and MRI were preprocessed to ensure that conversion of PET activity concentration to standardized uptake value and MRI bias field correction and normalization could not affect validation results (Appendix S4). Nine image processing and convolutional filter configurations were specified for each modality. Thus, a total of 486 features (18 features × nine configurations × three image modalities) could be computed. Teams were blinded to the results submitted by other teams. After submitting results, obvious configuration errors were reported back to the submitting team.

Statistical Analysis

Reproducibility of each of the 486 features computed in the validation phase was assessed using two-way random effects single-rater intraclass correlation coefficient for absolute agreement between teams (15). Based on Koo and Li (16), reproducibility of each feature was assigned to one of the following categories, based on the lower bound of the 95% CI of the intraclass correlation coefficient (17), as follows: 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. Intraclass correlation coefficient and their CI were computed in R version 4.2.1 (R Foundation for Statistical Computing).

Code

Analysis and results for phase 1 were scripted in Matlab (version 2020b and later; MathWorks). Analysis and results for phases 2 and 3, the figures and tables pertaining to the results, and the analysis presented in Appendix S5 were scripted and created in R (version 4.2.1 and later; R Foundation for Statistical Computing). All code is publicly available at *https://github.com/theibsi/ibsi_2_data_analysis (commit fde70ca)*.

Results

Characteristics of Participating Teams

Fifteen teams from seven countries participated in the first phase, 11 teams participated in the second phase, and nine teams participated in the validation phase. Twelve teams had developed publicly available software: Cancer Imaging Phenomics Toolkit (known as CaPTk), Computational Environment for Radiological Research (known as CERR), FAST, Local Image Feature Extraction (known as LIFEx), Multimodality Imaging for Radiomics Software (known as MIRAS), Medical Image Radiomics Processor (known as MIRP), moddicom, Standardized Imaging Biomarker Explorer (known as S-IBEX), SPAARC Pipeline for Automated Analysis and Radiomics Computing (known as SPAARC), Visualized and Standardized Environment for Radiomics Analysis (known as ViSERA), and the McGill and Université de Sherbrooke teams (Appendix S7).
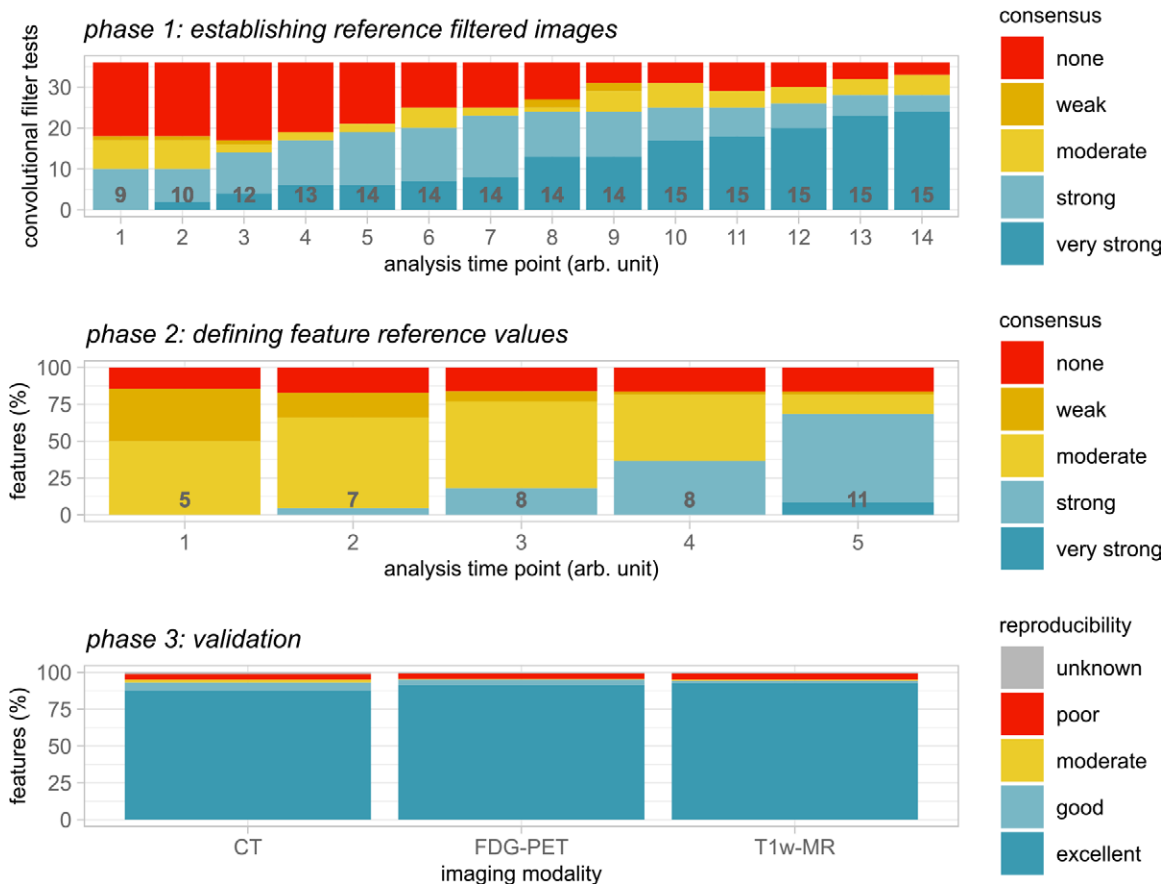
Figure 4: Results overview. In phase 1, participating teams computed 36 filtered images of convolutional filters according to predefined configurations. These filtered images were compared, and consensus was measured. Teams updated their implementations iteratively, which led to an improvement of consensus over time (arbitrary [arb.] unit; 27 months). Consensus strength was based on matching the voxel-wise difference between filtered images and the tentative reference filtered image within a tolerance. The number of participating teams at each point is shown. In phase 2, participating teams computed 396 features from filtered images of convolutional filters according to predefined filter and image processing configurations. As in phase 1, teams updated their implementations iteratively. Unlike phase 1, improvement in consensus was mostly because of more teams enrolling over time (arbitrary unit; 15 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 filtered images was validated. Teams computed 486 features from a public data set of 51 patients with soft-tissue sarcoma that were scanned using CT, fluorine 18 fluorodeoxyglucose (FDG) PET, and T1-weighted (T1w) MRI. Reproducibility was assessed using the lower bound of the 95% CI of the 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.

First Phase Results

Of the 36 filtered images that were assessed in the first phase, moderate or better consensus was found for 17 (47%) at the initial point (Fig 4). At the final point, moderate or better consensus was achieved for 33 (92%) configurations, of which 24 (67%) were very strong. Thus, 33 reference filtered images were established. Full consensus was reached for configurations corresponding to mean filters, Laplacian of Gaussian filters, Laws kernels, Gabor kernels, and separable and nonseparable wavelets (including decomposed forms). Weak or no consensus was achieved for three (8%) configurations, corresponding to configurations involving Riesz transformations (Fig S1).

Second Phase Results

At the initial time of the second phase, moderate or better consensus was achieved for 198 (50.0%) of 396 features, aggregated for 22 different filter configurations (Fig 4). At the final point, 323 (81.6%) features had at least moderate consensus. Again, full consensus was reached for features computed from filtered images of mean filters, Laplacian of Gaussian filters, Laws kernels, Gabor kernels, and separable and nonseparable wavelets (including decomposed forms), except for the quantile coefficient of dispersion feature for three-dimensional nonseparable wavelets. No consensus was established for features based on Riesz transformations (Fig S2) because too few teams submitted values for these features.

Validation Results

In summary, eight types of convolutional filters were standardized in the first two phases. The reproducibility of features from filtered images created by these filters was assessed in the third phase. Here, 458 (94.2%) of 486 features were found to have good to excellent reproducibility (intraclass correlation coefficient 95% CI lower bound, >0.75; Fig 4). Overall, 19 (3.9%; 19 of 486) features were poorly reproducible (intraclass correlation coefficient 95% CI lower bound, <0.50), and were found for

Glossary of Terms

Term	Definition
Standardization	The process of establishing uniform guidelines and protocols to ensure consistency and reproducibility.
Convolutional filter	A filter consisting of fixed or parameterized numerical values, that is slid (convolved) over an image to enhance potentially relevant characteristics, such as normal tissue-tumor boundaries, blood vessels, texture, and fibrosis.
Filtered image	The image produced by applying a (convolutional) filter to an image.
Low-pass filter	A filter that suppresses noise and other sharp patterns in an image and enhances smooth aspects.
High-pass filter	A filter that suppresses smooth aspects of an image and enhances details and sharp image patterns.
Band-pass filter	A filter that suppresses both smooth aspects of an image as well as sharp image patterns and enhances intermediate details.
Reference filtered image	An established filtered image representing the expected output of a specific convolutional filter applied to a specific image, that serves as a benchmark for verification and calibration.
(Radiomics) feature	A quantitative measure that is computed from a region of interest in a (filtered) image. The computation of common features was previously standardized by the image biomarker standardization initiative (4).
Reference feature value	An established expected value when computing a feature from a specific region of interest in a specific (filtered) image, that serves as a benchmark for verification and calibration.
Radiomics software	A software package that (at least) processes medical imaging and computes radiomics features.
Radiomics decision support tool	A computer application that provides clinical decision support based on radiomics features.
Mean filter	A filter that computes the average value within a neighborhood of voxels.
Laplacian of Gaussian filter	A filter used to detect edges and highlight fine details in an image.
Laws kernels	Sets of predefined filters used for highlighting various patterns in images, such as ripples.
Gabor kernels	Filters used for detection of directional patterns.
Wavelets	Sets of filters used to decompose images into different spatial frequency ranges.
Riesz transformation	A mathematical operation on filters that enhances edges and directional patterns in an image.

Laplacian of Gaussian and separable and nonseparable wavelet filters. Most of these features were either coefficient of variation or quartile coefficient of dispersion features that represented eight and nine of 19 features, respectively. A list of poorly reproducible features is provided in Table S1. Intraclass correlation coefficient values and their 95% CIs are listed in Tables S2–S10. No dependence on imaging modality could be observed.

Discussion

Convolutional filters enhance specific structures and patterns in medical images and are commonly used in radiomics analyses. However, because of a lack of proper consensus-based reference implementations, features computed from filtered images provided by these filters were difficult to reproduce (9). In our study, 15 teams from seven countries collaborated to remedy this situation by providing reference filtered images, reference feature values, and reference documentation. We were able to standardize and validate eight different filter types: mean, Laplacian of Gaussian, Laws and Gabor kernels, and separable and nonseparable wavelet filters in both undecomposed and decomposed forms. Thirty-three reference filtered images and 323 reference feature values, computed from filtered images, were established to standardize radiomics analyses across various imaging modalities.

Our results complement the previous results of the IBSI (4). The IBSI 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 methodologic details. This study adds to the previous work by standardizing the use of convolutional filters frequently used in radiomics.

Despite the overall success of the standardization process, there were two instances in which we did not achieve the desired level of success. First, we were unable to standardize Riesz transformations that, despite their attractive characteristics from a signal processing perspective, were not easy to implement. Thus, too few teams contributed data for Riesz transformations, and we could not establish their reference filtered images and reference values. Because Riesz transformations are rarely used in radiomics studies, the impact should be minimal. Second, several features could not always 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 filtered image with a mean intensity of zero. In the presence of high- and band-pass convolutional filters, the mathematical division operation present in both features led to otherwise negligible numeric differences between teams becoming relevant, resulting in poor reproducibility. Therefore, these features should not be used in combination with high- and band-pass filters.

Our work has several implications. First, we found that reproducible implementation of most types of convolutional filters across different radiomics software is not straightforward, as evidenced by the initial lack of consensus on reference filtered images in phase 1 (Appendix S8). Therefore, we assume that existing clinical or research radiomics software, which incorporates convolutional filters in advanced image analysis workstations,

may yield feature values that are not externally reproducible. This might impede external validation and subsequent clinical translation until the software is made to be compliant.

The second implication is that software labeled as IBSI-compliant is expected to reproduce the reference filtered images and reference feature values found in our study, insofar as convolutional filters are available in the software, in addition to the existing reference feature values (4). Developers of radiomics software supporting convolutional filters should aim to make their software compliant to improve reproducibility of radiomics analyses and allow for translation of enhanced clinical insights offered by convolutional filters. Developers should then clearly label their software as IBSI-compliant, to make it easier for users to identify and use their software for research and/or clinical purposes (with regulatory approval). Compliance may be checked using website-based tools *(https://ibsi.radiomics.hevs.ch/)*, or by manually comparing the produced filtered images and feature values against the provided reference data. Compliant software is expected to produce filtered images on which every voxel deviates from the reference filtered image by at most 1% of the range of intensity values of the reference filtered image (Appendix S5). Similarly, feature values must be within the specified tolerance margin around their reference feature values.

The third implication is that even though we contextualized our efforts within radiologic imaging, our work is relevant for quantitative image analysis in general, including digital pathology. Like our previous study (4), this study is anticipated to improve reproducibility of radiomics analyses beyond the modalities (digital phantoms, chest CT) and settings (non–small cell lung cancer) examined during the initial two phases of this study. To provide supporting preliminary evidence, we conducted validation using a publicly available data set composed of patients with soft-tissue sarcoma and multiple imaging modalities. The outcomes of the validation phase reinforce the potential applicability of our work in diverse settings.

Our study had limitations. First, its scope was restricted. Compliance with IBSI reference values helps to improve reproducibility of radiomic features (5,6). However, the results of a radiomics analysis also depend on image acquisition, reconstruction, segmentation, and data analysis steps (18,19), 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 (20). Such effects can be reduced by harmonization and cross-calibration of scanners and protocols (21) and post hoc techniques such as perturbation (22,23), batch normalization (24), and other methods (25). Second, participation in the IBSI does not guarantee that a particular software package is compliant with the IBSI reference standard. Changes introduced in software (5) or design choices may limit compliance (26). Third, we standardized intensity-based statistical features computed from filtered images but no other types of features. Morphologic features are mostly redundant because 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 when computed from filtered images. Their use may add hundreds or thousands of features to a radiomics analysis, which complicates the process of creating generalizable and interpretable radiomics models in the typical setting where at most a few hundred images are available for analysis. Finally, the IBSI has focused on radiomics using handcrafted features, and with this work offers a comprehensive reference standard for their computation. However, 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 conclusion, we standardized eight types of convolutional filters for radiomics to ensure that the enhanced clinical insights that can be gained through their use can be validated and reproduced. Going forward, developers should ensure compliance of their software with the proposed reference standards, and users are encouraged to use compliant software. A web-based tool is available for compliance checking. In the future, the Image Biomarker Standardization Initiative will focus on deep learning applications of radiomics, with the goal to provide reference standards for image preprocessing.

Author contributions: Guarantors of integrity of entire study, **A.Z., I.B., A.I., O.M., S.M.R., I.S., T.U., A.D.**; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; agrees to ensure any questions related to the work are appropriately resolved, all authors; literature research, **P.W., A.Z., A.P.A., S.B., O.M., A.R., S.M.R., C.G.R., M.R.S., I.S., F.Y., H.Z., H.M., M.V., A.D.**; clinical studies, **S.B., L.B., G.J.R.C., V.G., O.M., S.M.R., M.R.S., T.U., V.V.**; experimental studies, **P.W., A.Z., V.A., R.S., A.P.A., A.A., B.B., S.B., A.B., I.B., G.J.R.C., F.D., N.D., H.S.G., M. Hosseinzadeh, A.I., M.A.L.L., O.M., C.N., F.O., S.P., A.R., S.M.R., C.G.R., M.R.S., A.S., S.T.L., T.U., J.J.M.v.G., F.Y., H.Z., H.M., M.V., A.D.**; statistical analysis, **P.W., A.Z., A.P.A., S.B., M. Hosseinzadeh, J.L., F.M., O.M., S.M.R., C.G.R., M.R.S., I.S., T.U., H.M., M.V., A.D.**; and manuscript editing, **P.W., A.Z., R.S., A.P.A., B.B., S.B., R.B., I.B., G.J.R.C., N.D., H.S.G., M.G., M. Hatt, A.I., J.L., S.L., O.M., A.R., S.M.R., C.G.R., M.R.S., I.S., E.S., F.T., T.U., J.J.M.v.G., H.Z., H.M., M.V., A.D.**

References

1. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. Radiology 2016;278(2):563–577.
2. Tomaszewski MR, Gillies RJ. The Biological Meaning of Radiomic Features. Radiology 2021;298(3):505–516.
3. Huang EP, O'Connor JPB, McShane LM, et al. Criteria for the translation of radiomics into clinically useful tests. Nat Rev Clin Oncol2022;20(2):69–82.
4. 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.
5. 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.
6. Bettinelli A, Marturano F, Avanzo M, et al. A Novel Benchmarking Approach to Assess the Agreement among Radiomic Tools. Radiology 2022;303(3):533–541.
7. Depeursinge A, Al-Kadi OS, Ross Mitchell J. Biomedical Texture Analysis: Fundamentals, Tools and Challenges. Academic Press, 2017.
8. Beuque MPL, Lobbes MBI, van Wijk Y, et al. Combining Deep Learning and Handcrafted Radiomics for Classification of Suspicious Lesions on Contrast-enhanced Mammograms. Radiology 2023;307(5):e221843.
9. 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.
10. Depeursinge A, Andrearczyk V, Whybra P, et al. Standardised convolutional filtering for radiomics. arXiv [preprint] https://arxiv.org/abs/2006.05470. Posted June 9, 2020. Accessed August 7, 2023.
11. Lambin P. Data from: Radiomics Digital Phantom. CancerData. Published 2016. Accessed August 7, 2023.
12. 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.
13. 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.
14. 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. Published 2015. Accessed August 7, 2023.
15. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979;86(2):420–428.
16. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016;15(2):155–163.
17. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. Psychol Methods 1996;1(1):30–46.
18. 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.
19. 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.
20. 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.
21. Sullivan DC, Obuchowski NA, Kessler LG, et al. Metrology Standards for Quantitative Imaging Biomarkers. Radiology 2015;277(3):813–825.
22. Zwanenburg A, Leger S, Agolli L, et al. Assessing robustness of radiomic features by image perturbation. Sci Rep 2019;9(1):614.
23. Teng X, Zhang J, Zwanenburg A, et al. Building reliable radiomic models using image perturbation. Sci Rep 2022;12(1):10035.
24. 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.
25. 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.
26. Wright DE, Cook C, Klug J, Korfiatis P, Kline TL. Reproducibility in medical image radiomic studies: contribution of dynamic histogram binning. arXiv [preprint] https://arxiv.org/abs/2211.05241. Posted November 9, 2022. Accessed August 7, 2023.