Supplement 4
SPP Radiomics – List of challenges

A SPP Radiomics Workflow Definition Supplement

Version: 2024/03/13

Scope: This is a supplement for the publication “Radiomics Workflow Definitions and Challenges of Implementation in Clinics: a Delphi-based Interdisciplinary Consensus” by the Scientific Priority Program Radiomics (DFG SPP2177) by the Germany Research Foundation. The supplement contains a list of all challenges regarding the translation and clinical application of radiomics extracted from literature or proposed by experts. The challenges proposed by the experts have an index marked with *.X*) and were collected in the third round of the Delphi process described in the publication.

Remark: Challenge F.1 “Lack of reproducibility of biomarkers” was merged into C.5 “Problems related to reproducibility /generalizability” for round 5 of the Delphi process as on this abstraction level F.1 can be seen as a subset of C.5.

A Lack of guidelines

This category contains challenges that originate in a lack of guidelines for important/relevant steps in the workflow.

A.1 Lack of processing guidelines

A.2 Lack of reporting guidelines

A.X1 Lack of quality ensuring guidelines for reviewers (and editors)
(e.g. stricter guidelines to ensure scientific quality and manage the increasing number of radiomics; preventing quantity rather than quality (which is pushed by “publish and perish))

B Lack of standardization

This category contains challenges that originate in a lack of standardization of aspects in the workflow .

B.1 Lack of homogenous evaluation criteria
(e.g. the way evaluations are conducted (i.a. metrics to use, baselines) is not standardized; reporting guidelines not applied)

B.2 Lack of standardized computation methods
(e.g. the feature computation is not conducted in a standardized way)

C Problems related to radiomics studies

This category contains challenges that originate in problems with the study design or the way the study is conducted.

C.1 Lack of definition of a "good" radiomics study

C.2 Problems related to use of routine data
(e.g. the routine data proofs not to be suitable to answer the research question (i.a. due to problems with quantity, quality or contained information); other problems with the data not covered elsewhere)

C.3 Problems related to study design
(e.g. Insufficient size of patient cohorts, Lack of statistical significance, Non-consistent conduct of radiomics studies, High false-positive rates, Class imbalance; Insufficient / irrelevant clinical contribution (problem addressed is not relevant in clinical routine, e.g. radiomics prediction of IDH mutation status in gliomas))

C.4 Problems related to study results
(e.g. Lack of validation of studies, Causality difficult to establish, Overly optimistic results, Problems related to reporting of studies, Insufficient reporting about patient data, Insufficient reporting about prediction models, Insufficient reporting about methodological information)

C.5 Problems related to reproducibility /generalizability
(e.g. Lack repeatability, Lack of generalizability, Lack of reproducibility of studies)

C.X1 Lack of evidence gained by prospective evaluation
(e.g. performances of retrospective results cannot be reproduced in a prospective setting)

D Problems related to radiomics pipelines

This category contains challenges that originate in problems with the way the processing is done or the used tooling.

D.1 Problems related to image acquisition
(e.g. Imaging inconsistency, Imaging at multiple time points, Use of not standardized image acquisition protocols, Differences between imaging modalities, Lack of standardization of acquisition parameters, Heterogeneity caused by variations in acquisition parameters, different sequences (MRT), Contrast agent application protocol, used reconstruction kernels)

D.2 Problems related to prediction models
(e.g. Choosing a suitable algorithm for model building, Underfitting, Overfitting, Lack of generalizability of Radiomics Models, Lack of clinical utility, Lack of reproducibility of prediction models, Lack of standardized performance evaluation)

D.3 Problems related to image processing methods
(e.g. Number of subsets, Gaussian filter width for post reconstruction smoothing, unspecified hyperparameters)

D.5 Problems related to segmentations
(e.g. Differences in segmentation methods / software, Intra-/ Interobservervariability of segmentation, Reproducibility of segmentation methods)

D.6 Inconsistent processing schemes

D.7 Inconsistent radiomic methods

D.8 Variations in mathematical definitions

D.9 Differences in radiomics toolboxes

D.10 Differences between observers

D.11 Experimental inconsistency

D.X1 Problem related to uncertainty/trustability of models
(e.g. Workflows are not use because the uncertainty of their results can not be investigated; Workflows cannot detect that input is OOD and react/escalate accordingly)

D.X2 Lacking workflow integration
(e.g. Workflows are not well integrated in current workflows; imposing extra steps or systems that hinder normal workflow; too much time consumption by needed interactions)

E Problems related to data management

This category contains challenges that originate in problems with handling the data used in the study (e.g. including additional data not to the correct samples).

E.1 Complex data management

E.2 Difficulties integrating other data sources

F Problems related to radiomics features

This category contains challenges that originate in problems with the used radiomics features (e.g. use different radiomics extraction pipelines that have different names for the same features).

F.1 Lack of reproducibility of biomarkers

F.2 Variations in feature nomenclature

F.3 Lack of standardized radiomic features

F.4 Variability in radiomic feature generation

G Problems related to data sharing

This category contains challenges that originate in problems with sharing or reusing data for a study (e.g. not sharing data of private data sources makes studies not reproducible).

G.1 Legal and privacy problems
(e.g. open questions regarding model training/sharing and GPDR; Data sharing and GPDR/(broad)consent management; implications of Medical Device Directive if research software is applied prospectively (often not bearable from research groups))

G.2 Political and academic value of data

G.3 Hazards to reputation

G.4 Data recording methods

G.5 Cultural and language difficulties

G.6 Insufficient time

G.7 Insufficient human resources