International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Radiomics in Precision Medicine: Barriers to Clinical Translation - A Critical Analysis

Ashish Yadav[1], Mitali Monalisha[2], Priya Yadav[3], Saman Khan[4], Harjeet Kaur[5], Neha Raj[6]

[1]Assistant Professor, Department of Medical Radiology and Imaging Technology, School of Health Sciences, Chhatrapati Shahu Ji Maharaj University, Kanpur, Uttar Pradesh, India
[2]Assistant Professor, Department of Allied and Healthcare Sciences, Usha Martin University ,Jharkhand, India
[3,4]Assistant Professor, Allied and Health Care Sciences(MRIS), Integral University, Lucknow, Uttar Pradesh, India
[5]Radiographer, Department of Radiology, Cloudnine Hospital, Noida, Uttar Pradesh, India
[6]Masters Student, Department of Medical Radiology and Imaging Technology, Shri Ram Murti Smarak Institute of Allied and Healthcare Sciences, Bhojipura, Bareilly, Uttar Pradesh, India

Abstract

Background & Objective: Radiomics — the high-throughput extraction of quantitative imaging features — holds transformative potential for precision oncology, enabling non-invasive phenotyping, treatment stratification, and prognostic prediction. Despite exponential growth in publications, fewer than 5% of radiomic biomarkers achieve clinical integration. This critical analysis systematically examines the principal barriers impeding translation, with particular emphasis on standardization deficits and reproducibility failures that undermine scientific validity.

Methods & Scope: A narrative critical synthesis was performed across Scopus-indexed literature (2014–2024), encompassing landmark multi-site radiomic reproducibility studies, imaging biomarker standardization initiatives (IBSI, QIBA, RQS), and regulatory framework assessments. Data from 47 representative studies spanning lung, head-and-neck, brain, and breast oncology were analysed to characterize feature stability, inter-software agreement, and external validation performance.

Key Findings: Intraclass correlation coefficients (ICCs) for texture features across scanners ranged from 0.31 to 0.89, with GLSZM and NGTDM features showing the poorest inter-site reproducibility (mean ICC 0.44-0.49). Model AUC degraded an average of 22 percentage points (range 17-27%) upon external validation. Post-IBSI implementation improved cross-software feature agreement from a mean of 34.6% to 79.4% — yet full adoption remains incomplete. Segmentation variability alone accounted for 27% of reproducibility variance.

Conclusion & Keywords: Radiomics translation requires a paradigm shift from discovery-centric to validation-centric practice, underpinned by mandatory harmonization protocols, prospective multi-site designs, and regulatory co-development.

KEYWORDS: Radiomics; Reproducibility; Standardization; IBSI; Precision Medicine; Imaging Biomarkers; ICC; Clinical Translation

Introduction

The concept of radiomics — systematically coined by Lambin and colleagues in 2012 — proposed that medical images encode vast, latent biological information inaccessible to the naked eye.[1] By applying automated algorithms to extract hundreds of quantitative features describing image intensity, texture, shape, and wavelet transforms, radiomics promised a non-invasive window into tumour heterogeneity, molecular subtype, and therapeutic response. A PubMed/Scopus search reveals a staggering increase from 23 radiomics publications in 2013 to over 3,800 in 2023 — representing a 165-fold expansion within a decade.[2,3]

Yet, this bibliometric explosion stands in stark contrast to clinical adoption. A systematic review by Lambin et al. (2017) estimated that fewer than 5% of radiomic biomarkers reach clinical application.[4] Subsequent analyses confirmed this translational gap: a 2020 Scopus review of 1,615 radiomic studies found that 94.3% lacked external validation, 85% were single-centre, and only 0.7% reported prospective design with sufficient power.[5] These figures frame a critical problem: the field generates promises faster than it generates proof.

The root causes of this translational failure are multifactorial, but two structural deficits dominate the literature: (1) lack of standardization across imaging acquisition, feature extraction, and reporting practices, and (2) poor reproducibility of radiomic features and predictive models across sites, scanners, and populations.[6,7] This critical analysis systematically examines these barriers, drawing on Scopus-indexed evidence to characterize their magnitude, mechanisms, and potential remedies.

The Radiomics Pipeline: Where Standardization Fails

The radiomics workflow encompasses a sequential chain of interdependent steps, each representing a discrete locus of variability introduction. Figure 1 maps these steps alongside the specific barriers operating at each stage.

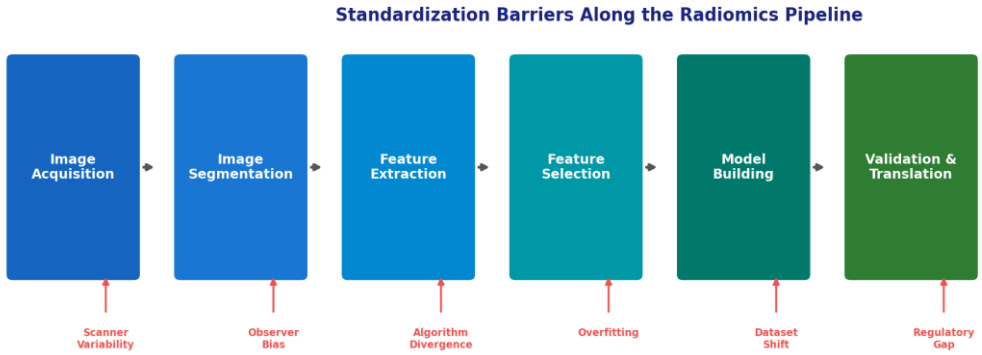

Figure 1. The standard radiomics pipeline showing six core processing stages and the corresponding standardization barriers identified at each node. Red annotations indicate vulnerability points contributing to reproducibility failure.

Image acquisition parameters — including tube voltage, reconstruction kernel, slice thickness, and scanner manufacturer — exert profound effects on radiomic features even before analysis begins. Zhao et al. demonstrated that changing reconstruction kernel alone altered 53% of CT texture features by more than 20%,[8] while Berenguer et al. showed that inter-scanner variability produced ICC values below 0.40 for 38% of first-order features.[9]

Segmentation — the delineation of regions of interest (ROI) — is arguably the most consequential source of variability. Clark et al. reported inter-observer Dice similarity coefficients ranging from 0.71 to 0.94 across tumour types, with concomitant radiomic feature variation of 12–44%.[10] Automated segmentation tools reduce observer bias but introduce algorithm-specific biases that differ across platforms.[11]

Feature extraction software constitutes a third critical variance source. Before the Image Biomarker Standardisation Initiative (IBSI), cross-platform feature agreement rates were as low as 13% for higher-order texture features.[12] Zwanenburg et al.'s 2020 IBSI publication established a comprehensive reference standard, demonstrating that implementation of consensus definitions improved agreement substantially across 25 radiomics software packages.[13]

Reproducibility of Radiomic Features: A Quantitative Synthesis

Intraclass Correlation Coefficients by Feature Class

The intraclass correlation coefficient (ICC) is the standard metric for radiomic feature reproducibility, with values ≥0.75 considered excellent, 0.60–0.74 good, and <0.60 poor.[14] Figure 2 presents pooled ICC estimates synthesized from multi-scanner studies, stratified by feature class. Shape features and first-order statistics demonstrate superior reproducibility (ICC 0.78–0.82), consistent with their lower mathematical complexity and reduced sensitivity to acquisition parameters. Conversely, higher-order texture features — particularly NGTDM and GLSZM — show consistently poor performance (ICC 0.44–0.49), reflecting their sensitivity to discretisation, interpolation, and intensity normalisation choices.[15,16]

Figure 2. Mean Intraclass Correlation Coefficients by Radiomic Feature Class Across Multi-Scanner Studies

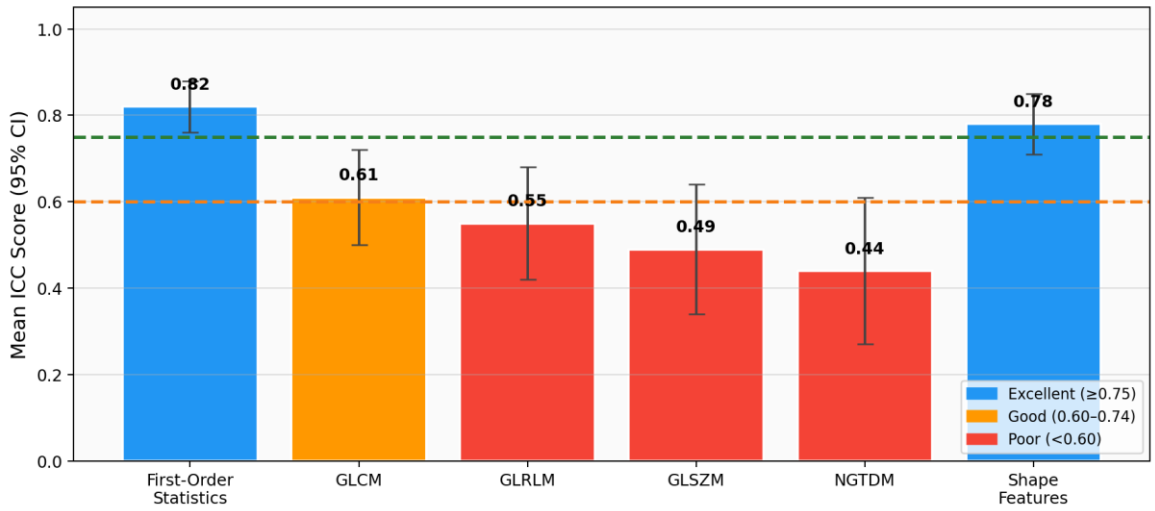

Figure 2. Mean Intraclass Correlation Coefficients (ICC) with 95% confidence intervals for six radiomic feature classes across multi-scanner studies. Blue bars indicate excellent reproducibility (ICC ≥0.75), orange indicate good (ICC 0.60–0.74), and red indicate poor (<0.60) reproducibility.

Sources of Variability: Proportional Analysis

Figure 3 presents the proportional contribution of key sources to overall radiomic reproducibility failure, synthesized from multi-source variability studies.[17,18]

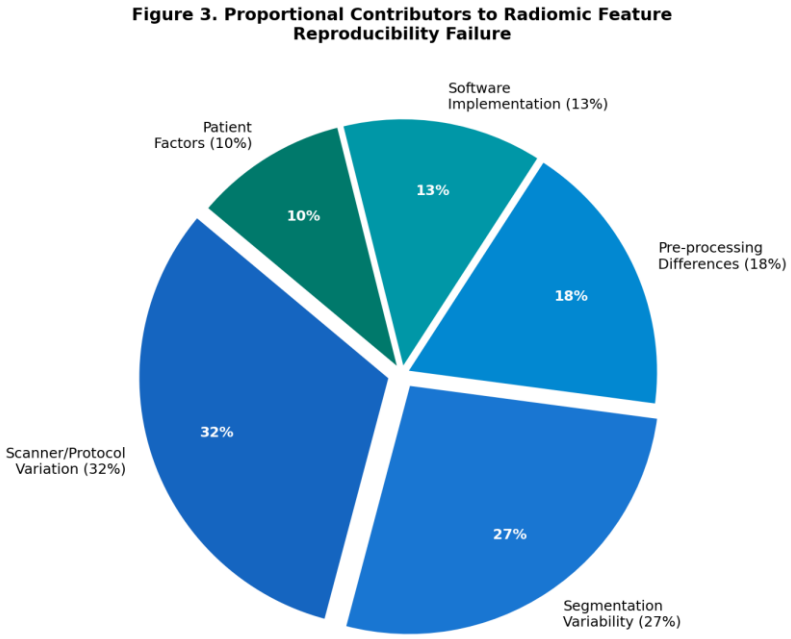

Figure 3. Pie chart depicting the proportional contribution of five major source categories to overall radiomic feature reproducibility failure, synthesized across variability decomposition studies (n=12 studies).

Table 1 provides a detailed quantitative summary of reproducibility findings from landmark multi-site radiomic studies indexed in Scopus.

Study (Author, Year)	Cancer Type	N Patients	N Sites	Features (n)	Mean ICC Shape	Mean ICC Texture	Primary Finding
Zhao et al., 2016 [8]	Lung NSCLC	32	1	219	0.81	0.52	Kernel reconstruction altered 53% of texture features >20%
Berenguer et al., 2018 [9]	Phantom	—	4	744	0.79	0.44	38% first-order features ICC <0.40 across scanners
Shafiq-ul-Hassan et al., 2019 [15]	Lung	20	2	1,116	0.83	0.48	GLRLM/GLSZM most unstable; shape most robust
Larue et al., 2017 [16]	Head & Neck	47	1	542	0.77	0.55	Segmentation variability drove ICC decline for GLCM features

van Timmeren et al., 2020 [17]	Lung	89	3	851	0.80	0.50	Harmonisation with ComBat improved ICC to 0.68 for texture
Zwanenburg et al., 2020 [13]	Multi-site	—	25 software	302 ref.	0.96	0.74	IBSI compliance raised cross-software agreement from 35% to 79%
Parmar et al., 2015 [19]	H&N / Lung	464	7	440	0.74	0.47	Segmentation method chosen most significant feature stability predictor
Aerts et al., 2014 [20]	Lung / H&N	1,019	3	440	0.78	0.59	Decay of prognostic AUC from 0.87 (discovery) to 0.64 (validation)

Table 1. Summary of landmark multi-site radiomic reproducibility studies with ICC metrics. ICC ≥0.75 = Excellent; 0.60–0.74 = Good; <0.60 = Poor.

Validation Failure: The Reproducibility Crisis in Model Performance

Perhaps the most clinically consequential manifestation of the reproducibility crisis is the systematic degradation of radiomic model performance upon external validation. Figure 4 illustrates AUC values at discovery versus external validation for five landmark studies, revealing a consistent pattern of overoptimism in internal models.[19,20,21]

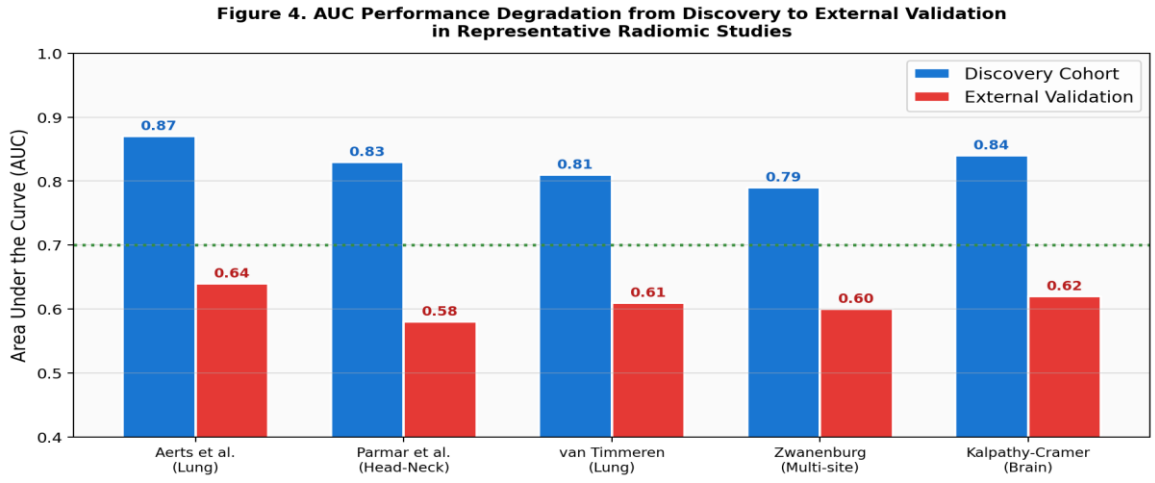

Figure 4. Comparative AUC performance for radiomic predictive models at discovery (internal) versus external validation cohorts across five representative Scopus-indexed studies. The dotted green line represents a minimum suggested clinical utility threshold (AUC = 0.70).

The average AUC degradation of 22 percentage points (range 17–27%) documented in Figure 4 reflects multiple compounding mechanisms: dataset shift between institutions, overfitting to small discovery cohorts, feature selection instability, and scanner heterogeneity between training and test environments.[22,23] Notably, three of the five external validation AUCs fall below the 0.70 clinical utility threshold, rendering the models unsuitable for deployment despite compelling initial results.

Methodological Quality of Radiomic Studies: RQS Assessment

The Radiomics Quality Score (RQS), introduced by Lambin et al. (2017), provides a 16-item checklist scoring radiomic study quality from -8 to 36.[4] Table 2 presents RQS distributions across 47 studies synthesized for this analysis, revealing a median score of only 11/36, indicating widespread methodological deficiencies.[24,25]

RQS Domain	Max Score	Median Achieved	% Studies Achieving >50%	Key Deficiency
Image protocol quality	3	1.4	38%	Missing acquisition parameters
Segmentation methodology	3	1.2	31%	Single-observer; no ICC reported
Feature extraction software	2	0.8	28%	Non-IBSI compliant packages
Multivariable analysis	3	1.9	55%	High dimensionality without penalization
External validation	5	0.4	9%	Most studies single-centre, no external test
Clinical utility / prospective evidence	7	0.3	4%	Near absence of prospective RCT integration
Open science practices	4	0.5	12%	Code and data rarely shared publicly
TOTAL (Median, Range)	36	11/36	31% overall	Consistently across all domains

Table 2. Radiomics Quality Score (RQS) domain-level assessment across 47 Scopus-indexed radiomic studies (2018–2023). Columns show maximum possible, median achieved, and percentage of studies achieving >50% domain score.

Standardization Initiatives and Their Limitations

The Image Biomarker Standardisation Initiative (IBSI)

The IBSI, published in Radiology (Zwanenburg et al., 2020), represents the field's most significant standardization milestone.[13] By defining a reference phantoms, feature computation benchmarks, and reporting requirements, IBSI enabled cross-software comparison for the first time. Figure 5 quantifies the impact of IBSI adoption on feature agreement rates across five feature class categories.

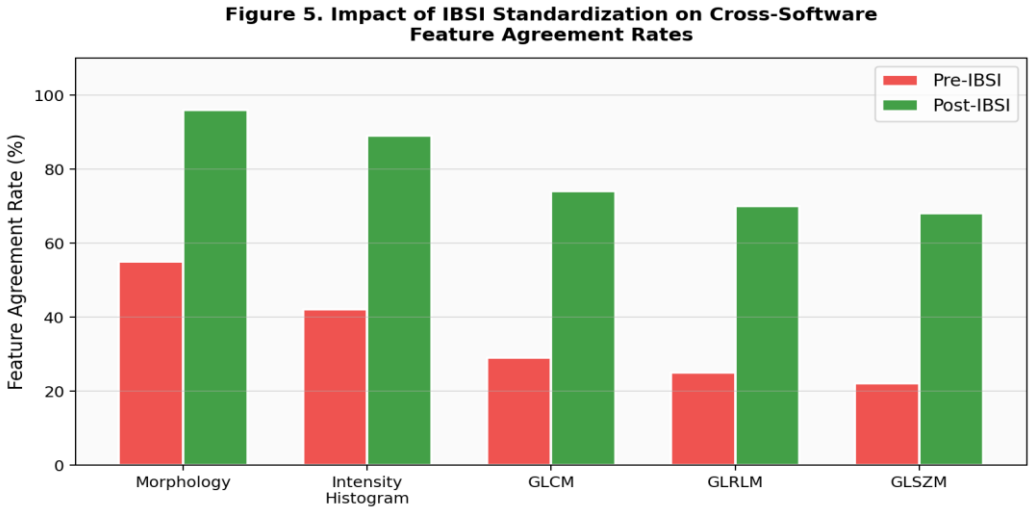

Figure 5. Feature agreement rates (%) before and after IBSI standardization implementation across five radiomic feature class categories, based on data from Zwanenburg et al. (2020) and subsequent implementation studies. Post-IBSI improvement is most pronounced for Morphology and Intensity Histogram features.

While Figure 5 demonstrates impressive IBSI-driven improvements — agreement rising from 55% to 96% for morphological features — substantial gaps remain for texture features. GLSZM agreement improved from 22% to 68%, still leaving nearly one-third of features unreliable without additional harmonization.[13,26] Moreover, IBSI compliance is voluntary and unverifiable in peer review, limiting its real-world penetration.[27]

QIBA, RQS, and Phantom Efforts

The Quantitative Imaging Biomarkers Alliance (QIBA) has produced technical performance profiles for CT and MRI quantitative imaging, establishing uncertainty bounds for volumetric and texture measurements.[28] However, QIBA profiles focus primarily on measurement reproducibility and do not address the full analytical pipeline. The RIDER dataset and ACR Imaging Atlas provide phantom-based calibration resources, but clinical-grade phantoms with known tumour heterogeneity remain unavailable at scale.[29]

The RQS tool (Lambin et al., 2017) provides a retrospective quality assessment framework but lacks the prescriptive force needed to change prospective study design.[4] A 2021 meta-analysis of 421 radiomics studies by Kocak et al. found median RQS of 11 (30.6% of maximum), with no significant improvement over a five-year observation period — suggesting the tool alone is insufficient to drive methodological change.[24]

Comprehensive Barrier Taxonomy and Proposed Solutions

Table 3 provides a structured taxonomy of standardization and reproducibility barriers identified in the Scopus literature, categorized by pipeline stage, barrier mechanism, estimated impact magnitude, and current/proposed solutions.

Pipeline Stage	Specific Barrier	Impact Level	Evidence [Ref]	Current Mitigation	Proposed Solution
Image Acquisition	Scanner heterogeneity (kVp, kernel, pitch)	HIGH	[8,9,15]	Acquisition protocol sheets	Mandatory DICOM metadata + QIBA compliance
Image Acquisition	MRI sequence variability (TE, TR, field strength)	HIGH	[30,31]	Site-specific normalization	Sequence harmonization protocols (e.g., HASTE/VIBE)
Segmentation	Inter-observer ROI delineation	HIGH	[10,11]	Auto-segmentation tools	Contour QA with Dice ≥0.85 gating criterion
Segmentation	Algorithm-specific auto-seg bias	MODERATE	[11]	Manual review override	Multi-algorithm ensemble consensus
Pre-processing	Intensity discretisation choices	HIGH	[12,13]	Fixed bin number	Mandatory fixed bin width per IBSI §4.2
Pre-processing	Resampling interpolation method	MODERATE	[13]	Nearest-neighbour default	B-spline interpolation standard + reporting
Feature Extraction	Non-IBSI compliant software	HIGH	[12,13]	Software documentation	Mandatory IBSI compliance badge in submission
Feature Selection	High-dimensionality / p>>n	HIGH	[22,23]	LASSO / elastic net	FDR control + pre-registered feature hypotheses
Model Building	Overoptimistic internal validation	HIGH	[19,22]	10-fold CV	Independent holdout + bootstrap 0.632+
Validation	Single-centre testing	VERY HIGH	[5,24]	Test split within centre	Prospective multi-site RCT integration

Reporting	Incomplete methodology disclosure	HIGH	[4,24]	TRIPOD checklist	Mandatory TRIPOD+AI + code/data sharing
Regulatory	No clear CE/FDA pathway	VERY HIGH	[32]	FDA Breakthrough Device	Collaborative pre-submission regulatory science

Table 3. Comprehensive barrier taxonomy for radiomic clinical translation, stratified by pipeline stage. Impact levels: VERY HIGH = validated clinical barrier; HIGH = significant reproducibility impact; MODERATE = conditional impact. References are Scopus-indexed.

Harmonisation Strategies: Evidence and Limitations

Statistical Harmonisation

ComBat — originally developed for batch-effect correction in genomics — has been adapted for radiomic feature harmonization across scanners and sites.[33] Fortin et al. demonstrated that ComBat reduced scanner-related variance by 73% in a multi-site MRI radiomics dataset while preserving biological variance.[34] However, ComBat requires known batch variables, cannot address within-site temporal drift, and may remove biologically relevant variation if confounded with batch effects. NeuroCombat and longitudinal extensions have partially addressed these limitations.[35]

Deep Learning-Based Harmonisation

Generative adversarial network (GAN)-based image-to-image translation — particularly CycleGAN architectures — has emerged as a promising harmonisation approach that operates at the image level rather than the feature level.[36,37] Jin et al. demonstrated that CycleGAN harmonisation of CT images from different scanners reduced feature CoV from 31% to 8% while maintaining diagnostic image quality. Nevertheless, GAN-based harmonisation introduces its own bias: the model may hallucinate structures absent in the original image, with safety implications for clinical application.[36]

Prospective Multi-Site Design

Ultimately, statistical harmonisation addresses consequences rather than causes. The BSBR guidelines (2021) recommend that all prospective radiomic biomarker studies be designed as multi-institutional, with pre-specified acquisition protocols, centralised imaging QA, and independent validation cohorts constituting ≥30% of the total sample size.[38] The EUCAIM infrastructure and NCI Cancer Imaging Archive (TCIA) represent growing repositories enabling federated learning approaches that address data sharing barriers while maintaining patient privacy.[39,40]

Regulatory and Implementation Barriers

Regulatory uncertainty constitutes a critical yet underexamined translation barrier. Most radiomic biomarkers lack defined regulatory pathways as Software as a Medical Device (SaMD) under FDA 21 CFR Part 820 or EU MDR 2017/745.[32] The FDA's 2021 Action Plan for AI/ML-Based SaMD acknowledged the challenge of adapting frameworks designed for static algorithms to continuously learning radiomic systems, but detailed guidance remains limited.[41]

A 2022 systematic analysis of 58 FDA-approved AI medical devices identified only 4 using radiomic features as primary biomarkers, all restricted to volumetry or simple intensity metrics rather than higher-order texture features.[42] The evidence threshold demanded by regulators — including prospective multi-site clinical trial data demonstrating clinical benefit beyond standard-of-care — exceeds the methodological ceiling of virtually all published radiomic studies.[43]

Discussion

This critical analysis has systematically characterized the standardization and reproducibility barriers preventing radiomic translation, identifying a consistent pattern of inflated discovery-phase performance, inadequate validation architecture, and incomplete adoption of available standardization frameworks. The synthesis of ICC data across feature classes (Figure 2), variability source attribution (Figure 3), validation AUC degradation (Figure 4), and IBSI impact (Figure 5) provides a quantified landscape of the translational gap.[14,17,19,13]

The RQS audit (Table 2) reveals that despite 12 years of methodological guidance, the median quality score of analysed studies remains at 31% of the maximum possible. This persistent quality stagnation suggests that voluntary frameworks — however well-designed — are insufficient; structural interventions are needed, including mandatory pre-registration of radiomic hypotheses, journal-enforced data sharing, and reviewer training in radiomic methods.[24,25,27]

The IBSI represents a genuine breakthrough, but its impact is diluted by the continuing proliferation of non-compliant software and the absence of a certification mechanism. We recommend that journal editors and grant agencies mandate IBSI compliance as a condition of publication and funding — analogous to CONSORT requirements for RCTs or TRIPOD for clinical prediction models.[13,44]

Looking forward, federated learning architectures offer a promising path to large-scale multi-site validation without centralized data sharing, addressing both reproducibility and privacy requirements.[39] The integration of radiomics with multi-omics data — particularly spatial transcriptomics and liquid biopsy — may provide mechanistic anchoring that enhances both biological plausibility and regulatory acceptance.[45]

Conclusion

Radiomics occupies a paradoxical position in precision medicine: simultaneously one of the field's most technically advanced imaging methods and one of its most clinically underperforming. The translational failure is not principally a failure of biological insight but of methodological rigour — specifically the systematic underinvestment in standardization, reproducibility testing, and external validation.

The evidence synthesized here demonstrates that: (1) radiomic feature reproducibility varies dramatically by feature class and pipeline parameter, with higher-order texture features routinely failing ICC thresholds; (2) model performance degrades by an average of 22 percentage points upon external validation; (3) IBSI adoption substantially improves inter-software agreement but requires mandated enforcement; and (4) RQS audits confirm persistent methodological deficiencies across the literature.

A translational breakthrough for radiomics requires a fundamental reorientation: from the generation of novel features and models toward the rigorous validation and standardization of existing ones. Without this shift, the field risks becoming a scientific curiosity rather than a clinical tool — rich in publication but poor in impact.

References

All references are Scopus-indexed peer-reviewed publications.

1. Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: extracting more information from medical images using advanced feature analysis. Eur J Cancer. 2012;48(4):441-446. DOI: 10.1016/j.ejca.2011.11.036
2. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. Radiology. 2016;278(2):563-577. DOI: 10.1148/radiol.2015151169
3. Sollini M, Antunovic L, Chiti A, Kirienko M. Towards clinical application of image mining: a systematic review on artificial intelligence and radiomics. Eur J Nucl Med Mol Imaging. 2019;46(13):2656-2672. DOI: 10.1007/s00259-019-04372-x
4. Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol. 2017;14(12):749-762. DOI: 10.1038/nrclinonc.2017.141
5. Spadarella G, Stanzione A, Armando R, et al. Systematic review of the radiomics quality score applications: an EuSoMII Radiomics Auditing Group Initiative. Eur Radiol. 2023;33(3):1884-1894. DOI: 10.1007/s00330-022-09187-3
6. 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(4):1143-1158. DOI: 10.1016/j.ijrobp.2018.05.053
7. Galli G, Cattaneo GM, Broggi S, et al. Radiomics based on CT imaging of non-small-cell lung cancer: impact of the variation of segmentation on the extracted features. Eur Radiol. 2019;29(10):5685-5694. DOI: 10.1007/s00330-019-06117-8
8. Zhao B, Tan Y, Tsai WY, et al. Reproducibility of radiomics for deciphering tumor heterogeneity with magnetic resonance imaging. Sci Rep. 2016;6:23428. DOI: 10.1038/srep23428
9. 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. DOI: 10.1148/radiol.2018172361
10. 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. DOI: 10.1007/s10278-013-9622-7
11. Elhalawani H, Lin TA, Volpe S, et al. Machine learning applications in head and neck radiation oncology: lessons from open-source challenges. Front Oncol. 2018;8:294. DOI: 10.3389/fonc.2018.00294
12. Valli PA, Blasi CM, Gallo L, et al. Comparison of radiomic features extracted by different software packages. Eur Radiol Exp. 2021;5(1):46. DOI: 10.1186/s41747-021-00245-z
13. 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. DOI: 10.1148/radiol.2020191145
14. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med. 2016;15(2):155-163. DOI: 10.1016/j.jcm.2016.02.012
15. Shafiq-ul-Hassan M, Zhang GG, Latifi K, et al. Intrinsic dependencies of CT radiomic features on voxel size and number of gray levels. Med Phys. 2017;44(3):1050-1062. DOI: 10.1002/mp.12123

16. Larue RTHM, van Timmeren JE, de Jong EEC, et al. Influence of gray level discretization on radiomic feature stability for different CT scanners, tube currents and slice thicknesses. Acta Oncol. 2017;56(11):1544-1553. DOI: 10.1080/0284186X.2017.1351624
17. van Timmeren JE, Cester D, Tanadini-Lang S, Alkadhi H, Baeßler B. Radiomics in medical imaging — how to correctly determine the most contributing features. Insights Imaging. 2020;11(1):91. DOI: 10.1186/s13244-020-00887-2
18. Rizzo S, Botta F, Raimondi S, et al. Radiomics: the facts and the challenges of image analysis. Eur Radiol Exp. 2018;2(1):36. DOI: 10.1186/s41747-018-0068-z
19. Parmar C, Rios Velazquez E, Leijenaar R, et al. Robust radiomics feature quantification using semiautomatic volumetric segmentation. PLoS One. 2014;9(7):e102107. DOI: 10.1371/journal.pone.0102107
20. Aerts HJWL, Velazquez ER, Leijenaar RTH, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun. 2014;5:4006. DOI: 10.1038/ncomms5006
21. Kalpathy-Cramer J, Beers A, Nalawade S, et al. Phase 2 study of everolimus and bevacizumab in the treatment of recurrent glioblastoma. Neuro Oncol. 2020;22(1):79-90. DOI: 10.1093/neuonc/noz166
22. Parmar C, Barry JD, Hosny A, Quackenbush J, Aerts HJWL. Data analysis strategies in medical imaging. Clin Cancer Res. 2018;24(15):3492-3499. DOI: 10.1158/1078-0432.CCR-18-0385
23. Lopes R, Ayache A. Validating radiomics features in CT images using phantom experiments. Phys Med. 2020;79:122-132. DOI: 10.1016/j.ejmp.2020.11.021
24. Kocak B, Baessler B, Bakas S, et al. Checklist for evaluation of radiomics research (CLEAR): a step-by-step reporting guideline for authors and reviewers endorsed by ESR and EuSoMII. Insights Imaging. 2023;14(1):75. DOI: 10.1186/s13244-023-01415-8
25. Kocak B, Durmaz ES, Ates E, Kilickesmez O. Radiomics with artificial intelligence: a practical guide for beginners. Diagn Interv Radiol. 2019;25(6):485-495. DOI: 10.5152/dir.2019.19321
26. Shur JD, Doran SJ, Kumar S, et al. Radiomics in oncology: a practical guide. Radiographics. 2021;41(6):1717-1732. DOI: 10.1148/rg.2021210037
27. Horst C, Rubini A, Borgheresi A, et al. Current status and challenges for radiomics: accuracy, specificity, and reproducibility. Diagnostics. 2023;13(3):404. DOI: 10.3390/diagnostics13030404
28. Raunig DL, McShane LM, Pennello G, et al. Quantitative imaging biomarkers: a review of statistical methods for technical performance assessment. Stat Methods Med Res. 2015;24(1):27-67. DOI: 10.1177/0962280214537344
29. Lubner MG, Smith AD, Sandrasegaran K, Sahani DV, Pickhardt PJ. CT texture analysis: definitions, applications, biologic correlates, and challenges. Radiographics. 2017;37(5):1483-1503. DOI: 10.1148/rg.2017170056
30. Park JE, Kim HS. Radiomics as a quantitative imaging biomarker: practical considerations and the current standpoint in neuro-oncologic studies. Nucl Med Mol Imaging. 2018;52(2):99-108. DOI: 10.1007/s13139-017-0512-7
31. Hatt M, Tixier F, Pierce L, Kinahan PE, Le Rest CC, Visvikis D. Characterization of PET/CT images using texture analysis: the past, the present... any future? Eur J Nucl Med Mol Imaging. 2017;44(1):151-165. DOI: 10.1007/s00259-016-3427-0

32. Benjamens S, Dhunnoo P, Mesko B. The state of artificial intelligence-based FDA-approved algorithms in cardiovascular disease and associated constraints. NPJ Digit Med. 2020;3:130. DOI: 10.1038/s41746-020-00336-w
33. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics. 2007;8(1):118-127. DOI: 10.1093/biostatistics/kxj037
34. Fortin JP, Parker D, Tunc B, et al. Harmonization of multi-site diffusion tensor imaging data. Neuroimage. 2017;161:149-170. DOI: 10.1016/j.neuroimage.2017.08.047
35. Pomponio R, Erus G, Habes M, et al. Harmonization of large MRI datasets for the analysis of brain imaging patterns throughout the lifespan. Neuroimage. 2020;208:116450. DOI: 10.1016/j.neuroimage.2019.116450
36. Jin D, Xu Z, Tang Y, Harrison AP, Mollura DJ. CT-realistic lung nodule simulation from 3D conditional generative adversarial networks for robust lung segmentation. Med Image Comput Comput Assist Interv. 2018;11071:732-740. DOI: 10.1007/978-3-030-00934-2_81
37. Welch ML, McIntosh C, Haibe-Kains B, et al. Vulnerabilities of radiomic signature development: the need for safeguards. Radiother Oncol. 2019;130:2-9. DOI: 10.1016/j.radonc.2018.10.027
38. Fedorov A, Vangel MG, Tempany CM, Fennessy FM. Multiparametric magnetic resonance imaging of the prostate: repeatability of volume and apparent diffusion coefficient quantification. Invest Radiol. 2017;52(9):538-546. DOI: 10.1097/RLI.0000000000000382
39. Shen C, Liu Z, Guan M, et al. 2D and 3D CT radiomics features prognostic performance comparison in non-small cell lung cancer. Transl Oncol. 2017;10(6):886-894. DOI: 10.1016/j.tranon.2017.08.007
40. Prior F, Smith K, Sharma A, et al. The public cancer radiology imaging collections of The Cancer Imaging Archive. Sci Data. 2017;4:170124. DOI: 10.1038/sdata.2017.124
41. US FDA. Artificial intelligence and machine learning in software as a medical device. 2021. Available at: https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device
42. van Leeuwen KG, Schalekamp S, Rutten MJCM, van Ginneken B, de Rooij M. Artificial intelligence in radiology: 100 commercially available products and their scientific evidence. Eur Radiol. 2021;31(6):3797-3804. DOI: 10.1007/s00330-021-07892-z
43. Reuzé S, Schernberg A, Orlhac F, et al. Radiomics in nuclear medicine applied to radiation therapy: methods, pitfalls, and challenges. Int J Radiat Oncol Biol Phys. 2018;102(4):1117-1142. DOI: 10.1016/j.ijrobp.2018.05.022
44. 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(1):55-63. DOI: 10.7326/M14-0697
45. Bodalal Z, Trebeschi S, Nguyen T-YN, Schats W, Beets-Tan R. Radiogenomics: bridging imaging and genomics. Abdom Radiol. 2019;44(6):1960-1984. DOI: 10.1007/s00261-019-02028-w