

# Radiomics Beyond the Hype: A Critical Evaluation Toward Oncologic Clinical Use

Natally Horvat, MD, PhD • Nikolaos Papanikolaou, PhD • Dow-Mu Koh, MD

From the Department of Radiology, Memorial Sloan-Kettering Cancer Center, New York, NY (N.H.); Department of Radiology, University of São Paulo, São Paulo, Brazil (N.H.); Computational Clinical Imaging Group, Champalimaud Foundation, Portugal (N.P.); and Department of Radiology, Royal Marsden Hospital, Downs Rd, Sutton SM2 5PT, United Kingdom (N.P., D.M.K.). Received October 8, 2023; revision requested November 26; revision received April 14, 2024; accepted April 22. **Address correspondence to** D.M.K. (email: [dowmukoh@icr.ac.uk](mailto:dowmukoh@icr.ac.uk)).

N.H. supported in part by the NIH/NCI Cancer Center Support Grant (P30 CA008748). D.M.K. supported by the National Institute for Health and Care Research (NIHR) Biomedical Research Centre at the Royal Marsden NHS Foundation Trust and The Institute of Cancer Research, London.

Conflicts of interest are listed at the end of this article.

*Radiology: Artificial Intelligence* 2024; 6(4):e230437 • <https://doi.org/10.1148/ryai.230437> • Content codes: **IN AI**

Radiomics is a promising and fast-developing field within oncology that involves the mining of quantitative high-dimensional data from medical images. Radiomics has the potential to transform cancer management, whereby radiomics data can be used to aid early tumor characterization, prognosis, risk stratification, treatment planning, treatment response assessment, and surveillance. Nevertheless, certain challenges have delayed the clinical adoption and acceptability of radiomics in routine clinical practice. The objectives of this report are to (a) provide a perspective on the translational potential and potential impact of radiomics in oncology; (b) explore frequent challenges and mistakes in its derivation, encompassing study design, technical requirements, standardization, model reproducibility, transparency, data sharing, privacy concerns, quality control, as well as the complexity of multistep processes resulting in less radiologist-friendly interfaces; (c) discuss strategies to overcome these challenges and mistakes; and (d) propose measures to increase the clinical use and acceptability of radiomics, taking into account the different perspectives of patients, health care workers, and health care systems.

©RSNA, 2024

**R**adiomics is a promising and fast-developing field in oncology that leverages advanced computational techniques to mine high-dimensional quantitative data from medical images. The premise of radiomics is that medical images hold valuable information regarding tumor biology, behavior, and pathophysiology that are not readily discernible through conventional radiologic visual assessment (1).

Given the ability of radiomics to provide high-dimensional quantitative data from medical images, it has the potential to transform cancer management, whereby radiomics data can be used to aid early tumor characterization, prognosis, risk stratification, treatment planning, response assessment, and surveillance (1). Correspondingly, since its introduction in the field of oncology not more than a decade ago, there has been an exponential growth in radiomics studies within oncology. Nevertheless, despite the substantial body of radiomics research that has been established in the literature to date, certain challenges have delayed the translation and acceptability of radiomics in routine clinical practice. To fully unlock the potential of radiomics as a clinical tool, improvements must yet be made to address these challenges, including data standardization, infrastructure support, and radiomics signature reproducibility, transparency, validation, usability, and trustworthiness (2,3).

In this report, our objectives are to (a) provide our perspective on the translational potential and potential impact of radiomics in oncology; (b) explore frequent challenges and mistakes in its derivation, encompassing study design, technical requirements, standardization, model reproducibility, transparency, data sharing, privacy concerns, quality control, as well as the complexity of multistep processes

resulting in less radiologist-friendly interfaces; (c) discuss strategies to overcome these challenges and mistakes; and (d) propose measures to increase the clinical use and acceptability of radiomics, taking into account the different perspectives of patients, health care workers, and health care systems.

## Radiomics and Oncology

Imaging remains integral in patients with cancer, providing critical information to oncologists, surgeons, and other members of the multidisciplinary patient management team who must discern the most appropriate treatment road map for each patient. However, there is a continuing shortage and unequal distribution of radiologists and other members of the oncology workforce throughout the world (4), even as the global cancer burden is projected to surge by 47% between 2020 and 2040, surpassing 28 million cases worldwide (5). This highlights a compelling need for innovative approaches not only to maximize the information provided by imaging but also to address the clinical and economic challenges posed by the growing cancer burden worldwide.

Radiomic features can be divided into two broad categories: (a) engineered or handcrafted features and (b) deep learning-derived features. The former refers to intensity, shape, and texture-related information that can be captured after applying predefined mathematical transformations at a pixel or voxel level, constituting traditional, manually crafted elements. On the other hand, the latter refers to advanced computational techniques and features that are learned in an end-to-end manner, mostly using convolutional neural networks. Throughout this report, we

## Abbreviations

ANOVA = analysis of variance, ComBat = combat batch effect adjustment, LASSO = least absolute shrinkage and selection operator

## Summary

To fully unlock the potential of radiomics as a clinical tool, improvements must be made to address multiple challenges, including challenges in data standardization, infrastructure support, and radiomics signature reproducibility, transparency, validation, usability, and trustworthiness.

## Key Points

- Radiomics can improve cancer imaging assessment by extracting quantitative data from routinely used medical images in the management of oncologic patients.
- Unaddressed challenges in radiomics hinder its clinical utility, leading skeptics to doubt its usefulness due to pipeline limitations.
- To maximize radiomics' clinical potential, it is essential to enhance multidisciplinary integration and address concerns regarding data standardization, infrastructure support, reproducibility, transparency, validation, usability, and trustworthiness.

## Keywords

Radiomics, Oncology, Cancer Management, Artificial Intelligence

focus on the engineered features. Radiomics holds the potential to augment what is currently provided by imaging as well as to provide a critical advantage in the setting of specialist shortages and rising cancer burden. The clinical translation potential of radiomics in oncology spans the entire patient journey, including early tumor detection (6), tumor characterization and subtyping (7), genetic profiling (8), prognostic assessment (9), risk stratification (10), assessment of response to treatment (11), assessment of treatment toxicity (12), surveillance (13), and assessment of survival outcomes (14). Fundamentally, radiomics can provide diagnostic or prognostic markers of clinically relevant oncologic outcomes, either alone or in combination with clinical data, laboratory tests, semantic radiologic features, and histologic and/or genetic findings. Ultimately, its goal is to function as a clinical decision support tool, providing the multidisciplinary teams with novel information for personalized and cost-effective treatment decisions.

Since radiomics was introduced to oncology in 2012, there has been almost 8000 studies evaluating the use of radiomics in oncology that have been indexed by PubMed up to December 6, 2023, as determined through a search employing the search terms: (“radiomics”) AND (“cancer” OR “tumor” OR “oncology”). Nevertheless, only a limited number of radiomics signatures have undergone rigorous external testing (15–25), and none have gained routine adoption in clinical practice. Clinical trials are essential for advancing medical knowledge, testing innovations, and improving patient outcomes through rigorous methodology and standardized data collection, ultimately leading to better treatment strategies. However, only a small number of radiomics studies have been conducted within a clinical trial setting, with fewer than 20 published studies incorporating clinical trial data (26–41). Notably, none of the published radiomics studies have prospectively implemented radiomics as a clinical decision support tool.

## Radiomics Workflow, Challenges, Common Mistakes, and Strategies to Overcome Them

Radiomics is a multistep process which involves several domains of expertise and elements. Given its interdisciplinary nature, the integration of these diverse elements can be prone to challenges and mistakes, which can be grouped into three main layers: (a) clinical relevance, (b) radiomics pipeline, and (c) publication bias. Figure 1 summarizes an ideal radiomics workflow from conception to clinical implementation.

### Clinical Relevance

**Rationale:** Similar to any other clinical research, the primary goal of radiomics research is to develop a model or signature to address unmet clinical needs and positively impact patient management. Defining the clinical relevance of a research study is the foundation not only in radiomics research, but also in all domains of medical research to ensure meaningful, impactful, and ethical research.

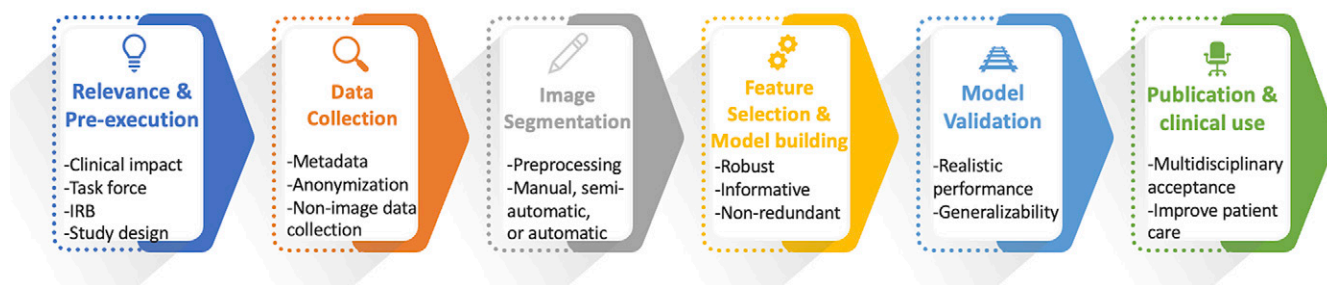
**Mistakes:** Based on our research experience, we believe that a frequent error in delineating the clinical relevance of a research study is the absence of integration and communication among various members of the multidisciplinary team during the initial stages of the study. Operating in isolation without a defined end point for clinical relevance and application, in addition to lacking potential benefits for patients and the health care community, stands out as a common and consequential mistake.

**Strategies to overcome mistakes:** To overcome this challenge, multidisciplinary stakeholders must be engaged at the project's outset and throughout the research process. This ensures a comprehensive understanding of clinical needs and real-world applications. Additionally, defining clear clinical utility (intended use) and potential benefits for patients and the community helps to align and prioritize research efforts. Finally, regular meetings and discussion among the team members helps to keep the group focused and facilitate dynamic exchanges and problem solving.

Another factor worth reflecting on is the likelihood of success. Although radiomics is a data-driven exercise, linking the radiomics findings across the spatial scale to a cellular or molecular event may be more challenging than providing radiomics quantification to a radiologic observation that can be discerned by the human eye or that is underpinned by a known biologic or pathologic link between the imaging and the outcome. Indeed, although radiomics signatures have been developed for molecular or genomic expression, none have been translated into a clinical tool or used in prospective studies. This reflects the challenges of filtering the relevant imaging signal from noise in the data when crossing spatial scales, which would impact the stability, reliability, and reproducibility of the observations.

### Radiomics Pipeline

**Pre-execution.— Rationale:** This step is required before conducting a study, primarily for study planning, assessing feasibility (sometimes necessitating pilot data gathering to determine the sample size), and evaluating the availability of high-quality data for model development and validation, including external



**Figure 1:** The figure summarizes an ideal radiomics workflow from conception to clinical implementation. IRB = institutional review board.

testing. Typically, this preliminary phase involves establishing the research team, designing the study, formulating hypotheses, setting study objectives, specifying inclusion and exclusion criteria, and estimating the expected sample size. It also entails deciding on the collection of relevant nonradiomics data (such as clinical, laboratory, and semantic quantitative data) and selecting the imaging modality to be evaluated. Moreover, it includes determining whether one or multiple timelines (eg, delta radiomics) will be incorporated (1,42). Finally, it includes seeking approval from the local ethics board, and if a multicentric study is planned, establishing data sharing agreements.

It is also crucial to carefully design the strategy for radiomics research. This involves constructing a causal framework by creating a direct acyclic graph and considering the development of the radiomics signature within this framework. Since radiomics relies solely on the collection of historical observational data that may be subject to biases, the modeling strategy plays a vital role in mitigating these biases and ensuring validity of the results.

**Mistakes:** Once again, the lack of cohesion among stakeholders can lead to misalignment in the overall study design and objectives. For instance, the absence of a comparison or an association between the radiomics signature and established imaging or clinical assessment tools can diminish its added clinical value. Additionally, the inclusion of a small sample or low-quality data can substantially weaken the reliability and quality of the radiomics signature being developed. As with other types of studies, data quality is intricately linked to study outcomes, such that the data integrated into the model affects the accuracy of the reference standard end point. Finally, it is a crucial mistake to overlook potential input bias which may be associated with sample selection, data gaps, imbalanced classes, socioeconomic disparities, imaging protocols and scanners as it can introduce bias into the output model and reduce generalizability of the results.

**Strategies to overcome mistakes:** It is a complex task to overcome the possible mistakes in this first step of the radiomics pipeline. This may require not only effort from the research team but also the institution and potentially societies and professional organizations.

From the research team, the following strategies should be considered: (a) adhere to radiomics specific checklists, such as CheckList for EvaluAtion of Radiomics research (CLEAR) (43), radiomics quality score (RQS) (44), and METHodological RadiomICs Score (METRICS), which was recently developed by the European Society of Medical Imaging Informatics with a

large international expert panel aiming at a transparent methodology from handcrafted radiomics to deep learning workflows (45) from the conceptualization phase to ensure comprehensive completion of the checklist; (b) increase integration at the inception of the research; (c) consider sample size and the need for multicentric collaborations; (d) implement data quality control to ensure high-quality input; and (e) actively identify and mitigate possible sources of bias. Regarding an initial sample size estimation, it is known that when performing feature reduction in radiomics studies, ensuring an adequate sample size for model validation is paramount. Typically, it is recommended to have around one-third of the training sample size for validation purposes. However, this requirement can vary depending on the balance in the dataset (eg, the number of “positive” or “negative” events in the dataset). In cases where the data are highly unbalanced, a larger sample size may be necessary to maintain model robustness and reliability. For instance, consider a scenario where a 10-feature radiomics model is being developed. To adhere to the one-third validation criteria, a minimum of 133 samples would be required, with 100 allocated for training and 33 for validation. Adequate sample sizes and careful consideration of data balance are crucial steps in ensuring the accuracy and potential generalizability of radiomics models (1,46).

In scenarios where dealing with imbalanced data are unavoidable, imbalanced data represent a potential source of bias. Acknowledging this challenge, the research team can employ various strategies to address it effectively. For instance, data preprocessing techniques, such as oversampling the minority class or undersampling the majority class, can be applied. Additionally, the extraction of relevant radiomics features that provide information for both classes is crucial. The choice of algorithms can be also helpful, and opting for those robust to imbalanced data, such as ensemble methods like random forest or those with cost-sensitive learning capabilities, is recommended. Use of cross-validation strategies, including k-fold, becomes essential to maintain the unbalanced distribution across different folds during both training and testing phases. Finally, continuous monitoring is also emphasized to ensure ongoing awareness and mitigation of any potential degradation in model performance.

Facilitating external testing across institutions and societies requires a thoughtful approach. Several key strategies can be considered to foster collaboration and ensure the reliability of research findings. First, promoting collaborative expertise among experts and encouraging the formation of multidisciplinary and multicentric teams can enhance the collective knowledge base.

Second, streamlining research processes and ethical approvals, alongside implementing secure data-sharing protocols that prioritize privacy measures, proves instrumental in facilitating external testing and supporting multicentric collaboration. Last, strengthening data management infrastructure is paramount. Allocating dedicated resources ensures the efficient handling of large datasets while also promoting seamless integration with existing clinical workflows and tools. These strategies collectively contribute to the establishment of a robust framework for external testing, fostering trust and collaboration.

**Data collection.— Rationale:** High-quality imaging and clinical data can effectively reflect cancer behavior and heterogeneity. This forms the basis for robust radiomics research, as the quality of the collected data directly influences the reliability and relevance of subsequent analyses.

**Mistakes:** Mistakes often stem from the absence of a streamlined data collection workflow, from data curation performed by individuals not specialized in the field and from heterogeneous imaging acquisition parameters. Image acquisition parameters greatly influence various stages in the radiomics workflow, from feature extraction to model building, thereby impacting the robustness and reproducibility of radiomic analyses. It is essential to consider these parameters during the data collection step to ensure that imaging parameters will be collected to be taken into consideration during model building and interpretability of the results. Conversely, depending on the radiomics signature and sample size, heterogeneity in the collected data can also present opportunities to enhance the generalizability of the radiomics signature being developed.

**Strategies to overcome mistakes:** Well-defined and optimized workflows that outline clear data collection processes, standardized clinical and radiologic reports, and standardized imaging protocols may be implemented to overcome mistakes pertaining to data collection. Notably, standardization across the radiomics pipeline has been recommended in different published guidelines, but there is no universal agreement as to how this should be achieved.

In addition, most radiomics signatures in published studies have been developed based on a “minimum” or limited dataset, which makes real-world clinical translation challenging when exposed to previously unseen data. Hence, data quantity and data diversity need to be addressed toward the creation of meaningful radiomics signatures that can be more widely generalized. The need for multicentric research and data curation at scale as well as the creation of well-curated imaging databases cannot be overemphasized. Large-scale data curation projects (eg, European Cancer Imaging Initiative) and existing cancer imaging databases (eg, The Cancer Imaging Archive) will play an increasingly important role in hosting relevant imaging data to drive radiomics research and clinical translation. Finally, engaging all stakeholders in data collection and instituting robust quality control programs are integral to the data collection process.

**Imaging segmentation and preprocessing.— Rationale:** The goal of image segmentation is to include a representative two-dimensional area (ie, region of interest) or three-dimensional

volume (ie, volume of interest) of the tumor, peritumoral region, or a tumor subregion, also known as a *tumor habitat*. Image segmentation can be performed using manual, semiautomatic, or automatic methods. Meanwhile, image preprocessing (such as filtering, normalization, and resampling) helps to enhance image quality and may simultaneously improve the standardization and reproducibility of radiomics features across scanners and imaging techniques. Accurate segmentation is extremely important before performing radiomics analysis to include both reproducible and representative areas of interest.

Combat batch effect adjustment (ComBat) is a statistical method initially designed to remove batch effects from microarray data. However, it has also been applied in radiomics research for image normalization and harmonization. ComBat uses an empirical Bayes model to adjust for batch effects by modeling the relationship between batches and gene expression levels. In the context of radiomics, ComBat can be used to adjust for acquisition parameters that may introduce systematic differences between datasets, allowing for more accurate comparisons and integrations of data from multiple sources or scanners. The application of ComBat in the radiomics workflow typically occurs during preprocessing or normalization stages before feature extraction and selection. By removing batch effects early on, subsequent steps in the pipeline, such as feature selection and machine learning model development, can benefit from more accurate and comparable data (47,48). It is essential to note that harmonization techniques like ComBat should be applied judiciously and only when necessary as they may introduce additional computational complexity and potential biases if not properly validated or used appropriately. Finally, it is our opinion that ComBat is not universally accepted due to its limitations with unseen data representations.

**Mistakes:** The most frequent shortcomings with image segmentation are related to segmentation errors, inter- and intra-reader variability, and insufficient integration with the clinical workflow or lack of functionality of the software used for image segmentation. Segmentation often requires analyzing multiple scans and employing techniques like comparing current scans with baseline images and using multiplanar reconstructions for precise delineation. When the segmentation software does not have full functionality like the picture archiving and communication system or when segmentation is performed without the availability of other image series on picture archiving and communication system workstations, it can hamper accuracy. Meanwhile, image preprocessing, although often necessary to harmonize data from different sources, can mask relevant radiomics signatures. Image preprocessing encompasses various steps, including spatial resampling and intensity discretization which, when applied without careful consideration, may inadvertently alter the inherent characteristics of the radiomics signatures. Spatial resampling, for instance, involves altering the voxel size, potentially leading to information loss or distortion. Similarly, intensity discretization, the process of categorizing pixel values into discrete bins, might oversimplify the intricate patterns present in the data. It is crucial to recognize these nuances and exercise caution during preprocessing to avoid unintended consequences that could compromise the authenticity and interpretability of radiomics signatures.



*Strategies to overcome mistakes:* To avoid these mistakes, presegmentation training sessions should be conducted to familiarize operators with the process and avoid pitfalls. The development and incorporation of automatic segmentation algorithms, when possible, can substantially reduce human fatigue and errors. Additionally, the implementation of stringent quality control measures as an integral part of the workflow can help ensure data reliability. In the context of heterogeneous tumors characterized by diverse components, different studies take different approaches to segment tumors. Some studies opt to exclude internal structures such as necrosis, calcification, and vessels from the region of interest or volume of interest, while others encompass these elements within the tumor segmentation. This variability underscores the need for a more standardized approach to tissue segmentation in every radiomics project. It is crucial to recognize that even the choice of segmentation tools can substantially impact the accuracy and reproducibility of radiomics results. For instance, a study by Owens et al (49) compared two segmentation approaches and revealed that radiomics features extracted from lesion sizing toolkit contours exhibited greater reliability across and among observers. The study further demonstrated that semiautomatic segmentation tools enabled observers without formal clinical training to achieve comparable results to physicians in evaluating tumor segmentation (49).

Regarding possible obscuration of relevant radiomics features during preprocessing, a meticulous and well-documented approach is essential. The following strategies can be considered: (a) clearly define and document each preprocessing step, (b) rigorously validate these techniques on representative datasets, (c) regularly review preprocessed data against original images, and (d) engage domain experts in decision-making.

#### **Feature extraction, feature selection, and model building.—**

*Rationale:* The goal of this step is to define the radiomics features that are both substantial and reproducible, with a focus on prioritizing those that are informative, stable, and nonredundant while also assessing their associations with biologic aspects of the tumor region or with outcome variables and to build a clinically meaningful and robust model.

*Mistakes:* Several mistakes can occur in this step, the most common ones being the following: (a) a lack of adherence to standardization, as recommended by the recently updated Image Biomarker Standardization Initiative (50), which can lead to inconsistencies in feature engineering and hinder comparison across studies and datasets; (b) feature extraction errors, such as selecting inappropriate algorithms, which can incur bias and diminish the quality of data representation; (c) cherry-picking features, which can lead to overfitting of the model; (d) multiple testing without proper correction, which can hinder the identification of potential false patterns that are due to chance rather than tumor characteristics; (e) lack of transparency in documenting the model-building process, including data preprocessing and algorithm selection, which can hinder model interpretability and reproducibility; and (f) insufficient small sample sizes, which can lead to overoptimistic model performance and limited generalizability.

*Strategies to overcome mistakes:* Well-suited feature extraction algorithms and robust statistical methods to strike the right balance between information retention and noise reduction should be employed. Feature selection methods should be implemented in a way that they do not expose information from the hold-out test set to the model and consequently lead to overoptimistic performance estimates. Rigorous and systematic approaches to model building include having a proper sample size, employing rigorous feature selection, implementing statistical correction for multiple tests, ensuring careful data handling to prevent leakage, and maintaining transparency throughout the process to ensure the model's reliability and usability.

In the context of small datasets, using nested cross-validation methods for feature selection can help mitigate bias and select features in an unbiased manner. Nested cross-validation involves an outer loop to evaluate the model's performance and an inner loop to select features within each outer fold. This approach can provide more reliable estimates of model performance and help prevent overfitting.

Notably, traditional wisdom in feature selection prioritizes features based on the level of statistical significance in their association with the outcome variable. This can result in the inclusion of features that lack explainability. Shape-based first-order features, and to some extent texture features (such as gray-level co-occurrence matrix [GLCM], gray-level run-length matrix [GLRLM], gray-level size-zone matrix [GLSZM], and neighboring gray-tone difference matrix [NGTDM]), can be more explainable compared with higher-order features (such as fractal dimension and Laplace of Gaussian) and may add confidence to the deployment and use of radiomics. While understanding the extraction processes and interpretability may not be essential for ensuring clinically useful and accurate models, prioritizing explainability in predictive modeling can substantially enhance trust, reliability, and ultimately the effectiveness of clinical translation and engagement. Moreover, legal considerations further underscore the importance of explainability in clinical contexts.

In the setting of large numbers of features, applying corrections for multiple testing, such as Bonferroni correction or false discovery rate correction, becomes crucial to mitigate the risk of spurious findings or false positives. While it may seem intuitive to perform corrections before feature selection to address this issue comprehensively, practical considerations may limit this approach's feasibility. The challenge lies in striking a balance between minimizing false positives and maintaining the power to detect relevant features. Therefore, a judicious application of correction methods, considering the specific stage of analysis and its computational demands, is essential for robust and reliable radiomics investigations.

Additionally, when working with a large number of texture features extracted from medical images, it is crucial to reduce their dimensionality to enhance model performance and prevent overfitting. To select the most important features, criteria such as relevance, redundancy, and generalization ability are among the most important to be considered. Relevance refers to how well each feature correlates with the target variable one aims to predict. Redundancy helps eliminate duplicate features that provide similar information. For example, strongly correlated features

are likely redundant. Although analysis of variance (ANOVA) is a commonly used feature selection method, its application in radiomics comes with major limitations due to the high correlation between input variables. This issue relates to the fact that ANOVA is insensitive to such correlations, leading to suboptimal results. Therefore, it is recommended to consider using alternative methods like least absolute shrinkage and selection operator (LASSO), minimum redundancy maximum relevance (mRMR), recursive feature elimination (RFE), and Boruta for feature selection before applying ANOVA. These methods are specifically designed to handle highly correlated radiomics features effectively and may provide more robust radiomics signatures. Additionally, they can be used in combination with ANOVA or other statistical methods to improve the overall performance of feature selection and ensure the most accurate and effective results for clinical applications. Features with consistent relevance across datasets have greater generalization ability. Regularized regression techniques like LASSO that penalize less useful features are commonly used in radiomics to automate feature selection based on relevance and redundancy criteria. LASSO is among the most popular automated methods. Ensemble approaches that aggregate rankings from multiple selection algorithms also enhance robustness. The dimensionality reduction goals guide the process, whether optimizing for model accuracy, efficiency, or other aims. There is no definitive rule for deciding the ideal number of features, but based on empirical evidence, it is generally advisable to have one feature for every 10 to 15 training samples (51).

Finally, in case one class is far more prevalent than the others, a common mistake is to use accuracy as an evaluation metric of the model's performance. Accuracy is defined as the proportion of correctly classified samples divided by the total number of samples. In an imbalanced dataset, a dummy classifier can achieve high accuracy by simply predicting the majority class. In such situations, the F1 score is a better choice. It is defined as the harmonic mean of precision and recall, where precision is the ratio of true positives to the sum of true positives and false positives, and recall is the ratio of true positives to the sum of true positives and false negatives. The F1 score takes into consideration both false positives and false negatives, making it less affected by class imbalance. It is a more robust measure of a model's capacity to classify instances in an imbalanced dataset correctly (52).

**Validation.**— *Rationale:* The goal of this step to validate generalizable radiomics signatures to predict the desired end points.

*Mistakes:* Improper data management, such as data leakage of information from the test set, influences the training set. Model validation should not be static given the various drifts that may happen at the image, target, or concept level.

*Strategies to overcome mistakes:* External test sets should be implemented to validate built models as they will reduce bias and lead to potentially more generalizable radiomics signatures. Assessing the temporal performance of a model through use of consecutive prospective patients for a period of time can give indications of model's robustness to external changes (ie, before and after a "silent" update in the scanner software). Ideally, models

should be deployed in a machine learning operations (MLOps) platform to be able to monitor their performance and check for input data drifts that may trigger retraining sessions of the model to keep it up to date. Model aging is a known effect that may result in gradual increases in model errors and variability, posing a potential risk for harming patients through erroneous predictions if left unattended.

One important limitation in the field of radiomics, as highlighted in recent studies (53), is the variability of radiomics features across different datasets. These features often demonstrate high performance only when applied to data similar to that on which they were originally trained. This limitation impacts the generalizability and reproducibility of radiomics signatures, making universal clinical application challenging. Beyond patient demographics, variations in data representations can stem from less apparent factors, including patient socioeconomic status, medical imaging vendors, model types, software releases, acquisition protocols, and specific local guidelines for patient selection for medical imaging examinations. This complexity underscores the need for standardization with image acquisition and feature extraction processes. One approach to mitigate these challenges is to retrain or fine-tune existing models using local data, creating model versions tailored to each hospital setting. This approach, however, poses substantial regulatory challenges. Additionally, the development of collaborative models trained on multi-institutional datasets may promote the creation of more robust and universally applicable radiomics signatures.

#### Publication Bias

Publication bias favors the publication of studies with positive outcomes over those reporting negative results, contributing to an overly optimistic view of radiomics that can have a detrimental impact on its reproducibility and hinder clinical acceptance. Moreover, numerous published studies have neglected to provide comprehensive details regarding imaging preprocessing and feature extraction methods, including aspects like voxel size resampling or bin width settings. This lack of transparency can impede the reproducibility of results. Addressing publication bias demands not only the encouragement of publishing negative or inconclusive findings but also advocating for the use of radiomics-specific checklists and quality scores to enhance the overall transparency and rigor of research methodologies.

#### Unmet Needs in Radiomics Research and Implementation

Despite more than 7000 published articles in radiomics research, there has been poor clinical translation of radiomics into patient care. This has led some skeptics to proclaim that radiomics is unlikely to be clinically useful due to the current limitations encountered across the entire radiomics pipeline. Optimists and supporters, however, hold firm that more needs to be done to create radiomics signatures that are more stable, repeatable, reproducible, interpretable, explainable, and generalizable, which can then provide unique insights into meaningful clinical outcomes. Perhaps the middle ground is to explore and address the current unmet needs

in radiomics development that are hindering the progress of radiomics in clinical translation.

Radiomics is data-driven research, and the paucity of high-quality well-annotated data are limiting the extent and depth of data exploration. Most radiomics signatures are built on limited datasets of under 1000 data points, which is a small number compared with the hundreds of radiomics features that are typically generated as part of the workflow. Despite our best efforts using current standardization and cross-validation techniques, there remains uncertainty in the feature selection process. Hence, collating larger data volume by the creation or linkage of image databases and repositories will be an important step toward the availability of data for radiomics research. Furthermore, as discussed, radiomics features often lack explainability, which can also hinder their clinical adoption. While interpretability may not be critical for ensuring clinically useful and accurate models, prioritizing explainable models can greatly enhance trust, reliability, and, ultimately, the effectiveness of clinical translation and engagement.

Part of the reason for limited data availability is because image segmentation and annotation is a time-consuming task that is reliant on trained radiologists and/or technicians to undertake. As the radiomics pipeline is not within the mainstream clinical workflow, these tasks are usually performed retrospectively. With the current workforce crisis in radiology, the ability to create data annotations prospectively is limited. Hence, artificial intelligence–assisted image annotation and segmentation holds the key to the automatic prospective creation of useful large datasets that can be fed into the radiomics workflow.

Although much has been discussed regarding standardization of the radiomics workflow, there is no universal agreement between published guidelines as to how this should be achieved. Due to the wide heterogeneity of protocols across vendor system and hospitals, standardization of scanning parameters is only pragmatic within a trial setting or across a few centers. However, the majority of our imaging data remain from the real world and are retrospective, which cannot be subject to such rigorous standardization. Hence, large datasets remain paramount in the search for meaningful radiomics signatures in heterogeneous datasets across different scanner systems. Recognizing the anticipated diversity in these datasets, the adoption of postextraction techniques, such as harmonization, emerges as a valuable strategy to improve model performance in the context of real-world cohorts. Even if we allow for the creation of a meaningful radiomics signature from clinical trial data, the validity of such a signature over time is uncertain as machine upgrades and performance drifts may affect the reliability of the signature.

The relevance of the radiomics use case cannot be overstated, as many published studies are of low-to-modest quality and may not address useful clinical questions. In the search for a relevant and widely generalizable radiomics signature, initial efforts may focus on an explainable signature where the outcome is more tightly linked to the imaging phenotype rather than across spatial scales (eg, gene expression). The latter may suffer from poorer signal-to-noise in the imaging data and potentially lead to spurious results without also considering all the potential causal inferences.

Undertaking the full radiomics pipeline and data analyses requires skills that are beyond the training of most radiologists. Hence, radiologists should be prepared and educated in those elements of the workflow that go beyond defining the research question, identifying the study population, performing image segmentation and annotation, and contributing to the interpretation of the results. The radiology department should recognize the investments needed to be ready for artificial intelligence and imaging data science. In particular, radiologists should be aware of the skill sets and new personnel required for future radiology departments, as well as the informatics infrastructure needed to operationalize a radiomics and artificial intelligence workflow. Figure 2 summarizes the most common mistakes within the radiomics workflow and suggests strategies to overcome them.

## **Radiomics: From Promises to Clinical Application**







Navigating the journey of radiomics is not solely reliant on technological advancements and scientific innovations. It also relies on the ability to convince end users of its value. In the setting of radiomics, we should consider the different perspectives of patients; health care providers, including radiologists and nonradiologists; and health care systems. Beyond the well-known importance of constructing an accurate, reproducible, and robust radiomics model, it is paramount to also comprehend the needs, trust, preferences, and behaviors of users in the radiomics clinical translation process (Fig 3).

### **Patients' Perspective**

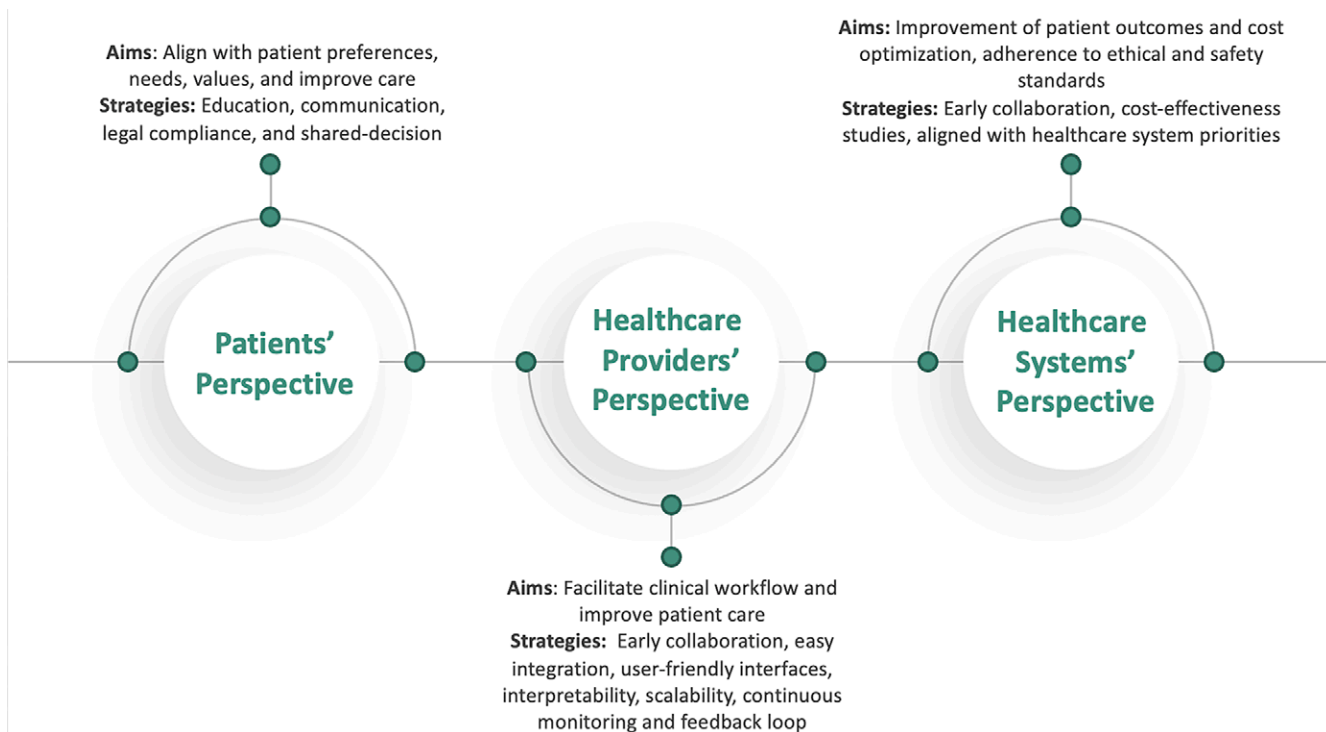
The traditional paternalistic approach of medicine is gradually evolving toward a more patient-centered paradigm, wherein patients are assuming an increasingly important role in shaping their health care decisions (54). As radiomics advances, it is essential to align its clinical application with patients' preference, needs, and values. A framework that includes education, communication, compliance with legal and regulatory standards, and the safeguarding of patient privacy may be considered in light of its proven benefits in other innovation initiatives. Informed consent and shared decision-making should be taken into consideration in the clinical translation process of radiomics (54,55). Although some technical definitions are difficult to understand, it is important for medical doctors to be able to explain the clinical significance to patients as much as possible. Patients should have access to information about how radiomics can contribute to their diagnosis and treatment, and they should feel empowered to ask questions, express their concerns, and actively participate in decisions about its use in their health care.

### **Health Care Providers' Perspective**

Health care providers, including radiologists, oncologists, and other specialists, will play a central role in employing the radiomics tools in the patient care workflow. The radiomics journey should include various stakeholders at the inception to build trust regarding tool reliability, reproducibility, and accuracy, since skepticism can be a barrier to adoption. Health care providers must receive comprehensive education to un-

	Common Mistakes	Strategies to overcome the mistakes
 <b>Relevance &amp; pre-execution</b>	Lack of integration among stakeholders, suboptimal sample size, overlook of potential bias	Early multidisciplinary engagement, critical study feasibility review, institutional and societal support for multicentric studies
 <b>Data collection</b>	Absence of streamlined workflow, low quality input data, heterogeneous data	Improved workflow, standardized protocols and reports, protected time to specialists involved on data collection
 <b>Image segmentation and preprocessing</b>	Segmentation errors, inter- and intra-reader variability, and lack of integration with PACS	Pre-segmentation training, automatic segmentation tools, quality control
 <b>Feature extraction, selection and model building</b>	Lack of adherence to standardization, feature extraction errors, incorrect feature selection (over- or underfitting)	Prefer standardized and robust feature extraction, careful feature selection, nested cross-validation from small data, consider explainability
 <b>Model validation</b>	Lack of transparency, data leakage to the validation set	Rigorous, systematic, and transparent approaches, careful data handling, external validation
 <b>Publication and clinical use</b>	Bias favoring positive results, focus on model accuracy rather than workflow implementation and reproducibility	Promote standardized guidelines, resources dedicated to improved workflow

**Figure 2:** Summary of the most common mistakes within the radiomics workflow and suggested strategies to overcome them. PACS = picture archiving and communication system.



**Figure 3:** Diagram shows different perspectives to be considered before radiomics is translated to clinical practice.

derstand radiomics' capabilities and limitations, critically assess its evidence base, and seamlessly integrate it into their clinical workflow. Simplicity, automatization, and easy integration between different systems are crucial factors. Health care providers already grapple with burnout, and the introduction of tools that increase the complexity of their daily work may exacerbate

anxiety and contribute to increasing burnout (56), ultimately adversely affecting patient outcomes (57). The goal of radiomics solutions should be to align with the preferences and workflows of health care professionals to facilitate their integration into clinical practice. Furthermore, health care providers must be able to assess radiomics' clinical utility in specific



patient populations, engage in shared decision-making with patients regarding its use, and collaborate across disciplines to maximize its impact. Programs of continuous learning are essential in this evolving field. Ultimately, radiomics should also align with patient-centered care, enhancing diagnostic accuracy and treatment planning while maintaining patient privacy and improving health care outcomes.

### Health Care Systems' Perspective

Despite the inherent complexity of health care systems, potential priorities can be summarized as follows: (a) patient outcomes and cost optimization (radiomics tools should prioritize the demonstration of clinical utility and cost-effectiveness with clear end points that align with the health care system's priorities to a given disease scenario), (b) feasibility (compatibility with existing systems and data sharing strategies is critical for seamless integration), and (c) regulatory and ethical considerations (close and upfront adherence to established standards and guidelines with a strong emphasis on patient privacy and data security are necessary to address health care system concerns). Therefore, involving health care system stakeholders and being aware of their priorities are important in shaping future radiomics tools and ensuring that they align with overall system goals and needs (58).

### Conclusion

In conclusion, this report provides a critical evaluation of the potential use of radiomics in oncology. While radiomics has the potential to add value to cancer management through the mining of quantitative high-dimensional data from medical images and much has been published on radiomics, important challenges remain unaddressed that hinder its clinical translation. To bridge the gap between the potential of radiomics and its clinical translation, this report highlights the current unmet needs, strategies that can address important mistakes in the radiomics pipeline, and the importance of including the perspective of multiple stakeholders to enhance the acceptability and widespread clinical use of radiomics in the oncologic setting.

**Acknowledgment:** The authors thank Joanne Chin, MFA, ELS, for editing this manuscript.

**Author contributions:** Guarantors of integrity of entire study, **N.H.**, **D.M.K.**; 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, all authors; experimental studies, **N.P.**; statistical analysis, **N.P.**; and manuscript editing, all authors

**Disclosures of conflicts of interest:** **N.H.** Consulting fees from Guerbet; payment from Bayer; travel support from Guerbet. **N.P.** Shareholder in MRIcons. **D.M.K.** Payment for speakers bureau for Bayer HealthCare and for consultancy to Guerbet.

### References

- Shur JD, Doran SJ, Kumar S, et al. Radiomics in Oncology: A Practical Guide. *RadioGraphics* 2021;41(6):1717–1732.
- Mu W, Schabath MB, Gillies RJ. Images Are Data: Challenges and Opportunities in the Clinical Translation of Radiomics. *Cancer Res* 2022;82(11):2066–2068.
- Stanzione A, Cuocolo R, Ugga L, et al. Oncologic Imaging and Radiomics: A Walkthrough Review of Methodological Challenges. *Cancers (Basel)* 2022;14(19):4871.
- Henderson M. Radiology Facing a Global Shortage: Specialty affected by COVID-19, aging population and demand for imaging. *Radiological Society of North America*. <https://www.rsna.org/news/2022/may/global-radiologist-shortage>. Published 2022. Accessed September 26, 2023.
- Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* 2021;71(3):209–249.
- Gillies RJ, Schabath MB. Radiomics Improves Cancer Screening and Early Detection. *Cancer Epidemiol Biomarkers Prev* 2020;29(12):2556–2567.
- Bhandari A, Ibrahim M, Sharma C, Liang R, Gustafson S, Prior M. CT-based radiomics for differentiating renal tumours: a systematic review. *Abdom Radiol (NY)* 2021;46(5):2052–2063.
- Jia LL, Zhao JX, Zhao LP, Tian JH, Huang G. Current status and quality of radiomic studies for predicting KRAS mutations in colorectal cancer patients: A systematic review and metaanalysis. *Eur J Radiol* 2023;158:110640.
- Gao Y, Cheng S, Zhu L, et al. A systematic review of prognosis predictive role of radiomics in pancreatic cancer: heterogeneity markers or statistical tricks? *Eur Radiol* 2022;32(12):8443–8452.
- Chiacchio G, Castellani D, Nedbal C, et al. Radiomics vs radiologist in prostate cancer. Results from a systematic review. *World J Urol* 2023;41(3):709–724.
- Pesapane F, Agazzi GM, Rotili A, et al. Prediction of the Pathological Response to Neoadjuvant Chemotherapy in Breast Cancer Patients With MRI-Radiomics: A Systematic Review and Meta-analysis. *Curr Probl Cancer* 2022;46(5):100883.
- Carbonara R, Bonomo P, Di Rito A, et al. Investigation of Radiation-Induced Toxicity in Head and Neck Cancer Patients through Radiomics and Machine Learning: A Systematic Review. *J Oncol* 2021;2021:5566508.
- Lee K, Le T, Hau E, et al. A Systematic Review Into the Radiologic Features Predicting Local Recurrence After Stereotactic Ablative Body Radiotherapy (SABR) in Patients With Non-Small Cell Lung Cancer (NSCLC). *Int J Radiat Oncol Biol Phys* 2022;113(1):40–59.
- Staal FCR, van der Reijdt DJ, Taghavi M, Lambregts DMJ, Beets-Tan RGH, Maas M. Radiomics for the Prediction of Treatment Outcome and Survival in Patients With Colorectal Cancer: A Systematic Review. *Clin Colorectal Cancer* 2021;20(1):52–71.
- Yuan E, Chen Y, Song B. Quality of radiomics for predicting microvascular invasion in hepatocellular carcinoma: a systematic review. *Eur Radiol* 2023;33(5):3467–3477.
- Bourbonne V, Fournier G, Vallières M, et al. External Validation of an MRI-Derived Radiomics Model to Predict Biochemical Recurrence after Surgery for High-Risk Prostate Cancer. *Cancers (Basel)* 2020;12(4):814.
- Bulens P, Couwenberg A, Intven M, et al. Predicting the tumor response to chemoradiotherapy for rectal cancer: Model development and external validation using MRI radiomics. *Radiother Oncol* 2020;142:246–252.
- Chang N, Cui L, Luo Y, Chang Z, Yu B, Liu Z. Development and multicenter validation of a CT-based radiomics signature for discriminating histological grades of pancreatic ductal adenocarcinoma. *Quant Imaging Med Surg* 2020;10(3):692–702.
- Healy GM, Salinas-Miranda E, Jain R, et al. Pre-operative radiomics model for prognostication in resectable pancreatic adenocarcinoma with external validation. *Eur Radiol* 2022;32(4):2492–2505.
- Horvat N, Veeraghavan H, Nahas CSR, et al. Combined artificial intelligence and radiologist model for predicting rectal cancer treatment response from magnetic resonance imaging: an external validation study. *Abdom Radiol (NY)* 2022;47(8):2770–2782.
- Ji GW, Zhu FP, Xu Q, et al. Machine-learning analysis of contrast-enhanced CT radiomics predicts recurrence of hepatocellular carcinoma after resection: A multi-institutional study. *EBioMedicine* 2019;50:156–165.
- Ji GW, Zhu FP, Xu Q, et al. Radiomic Features at Contrast-enhanced CT Predict Recurrence in Early Stage Hepatocellular Carcinoma: A Multi-Institutional Study. *Radiology* 2020;294(3):568–579.
- Wei W, Liu Z, Rong Y, et al. A computed tomography-based radiomic prognostic marker of advanced high-grade serous ovarian cancer recurrence: A multicenter study. *Front Oncol* 2019;9(APR):255.
- Zaid M, Widmann L, Dai A, et al. Predictive Modeling for Voxel-Based Quantification of Imaging-Based Subtypes of Pancreatic Ductal Adenocarcinoma (PDAC): A Multi-Institutional Study. *Cancers (Basel)* 2020;12(12):3656.
- Zhong X, Long H, Su L, et al. Radiomics models for preoperative prediction of microvascular invasion in hepatocellular carcinoma: a systematic review and meta-analysis. *Abdom Radiol (NY)* 2022;47(6):2071–2088.

26. Eertink JJ, van de Brug T, Wiegers SE, et al. 18F-FDG PET baseline radiomics features improve the prediction of treatment outcome in diffuse large B-cell lymphoma. *Eur J Nucl Med Mol Imaging* 2022;49(3):932–942.
27. Yu Y, He Z, Ouyang J, et al. Magnetic resonance imaging radiomics predicts preoperative axillary lymph node metastasis to support surgical decisions and is associated with tumor microenvironment in invasive breast cancer: A machine learning, multicenter study. *EBioMedicine* 2021;69:103460.
28. Tibermacine H, Rouanet P, Sbarra M, et al. Radiomics modelling in rectal cancer to predict disease-free survival: evaluation of different approaches. *Br J Surg* 2021;108(10):1243–1250.
29. Dercle L, Zhao B, Gönen M, et al. Early Readout on Overall Survival of Patients With Melanoma Treated With Immunotherapy Using a Novel Imaging Analysis. *JAMA Oncol* 2022;8(3):385–392.
30. Gong J, Zhang W, Huang W, et al. CT-based radiomics nomogram may predict local recurrence-free survival in esophageal cancer patients receiving definitive chemoradiation or radiotherapy: A multicenter study. *Radiother Oncol* 2022;174:8–15.
31. Wang XX, Ding Y, Wang SW, et al. Intratumoral and peritumoral radiomics analysis for preoperative Lauren classification in gastric cancer. *Cancer Imaging* 2020;20(1):83.
32. Papp L, Spielvogel CP, Grubmüller B, et al. Supervised machine learning enables non-invasive lesion characterization in primary prostate cancer with [68Ga]Ga-PSMA-11 PET/MRI. *Eur J Nucl Med Mol Imaging* 2021;48(6):1795–1805.
33. Chauvie S, De Maggi A, Baralis I, et al. Artificial intelligence and radiomics enhance the positive predictive value of digital chest tomosynthesis for lung cancer detection within SOS clinical trial. *Eur Radiol* 2020;30(7):4134–4140.
34. Zheng H, Zheng Q, Jiang M, et al. Contrast-enhanced CT based radiomics in the preoperative prediction of perineural invasion for patients with gastric cancer. *Eur J Radiol* 2022;154:110393.
35. Gugliandolo SG, Pepa M, Isaksson LJ, et al. MRI-based radiomics signature for localized prostate cancer: a new clinical tool for cancer aggressiveness prediction? Sub-study of prospective phase II trial on ultra-hypofractionated radiotherapy (AIRC IG-13218). *Eur Radiol* 2021;31(2):716–728.
36. Orton MR, Hann E, Doran SJ, et al. Interpretability of radiomics models is improved when using feature group selection strategies for predicting molecular and clinical targets in clear-cell renal cell carcinoma: insights from the TRACERx Renal study. *Cancer Imaging* 2023;23(1):76.
37. de Jong EEC, Sanders KJC, Deist TM, et al. Can radiomics help to predict skeletal muscle response to chemotherapy in stage IV non-small cell lung cancer? *Eur J Cancer* 2019;120:107–113.
38. Vuong D, Bogowicz M, Denzler S, et al. Comparison of robust to standardized CT radiomics models to predict overall survival for non-small cell lung cancer patients. *Med Phys* 2020;47(9):4045–4053.
39. Zhuang Z, Liu Z, Li J, et al. Radiomic signature of the FOWARC trial predicts pathological response to neoadjuvant treatment in rectal cancer. *J Transl Med* 2021;19(1):256.
40. Hou D, Zheng X, Song W, et al. Association of anaplastic lymphoma kinase variants and alterations with ensartinib response duration in non-small cell lung cancer. *Thorac Cancer* 2021;12(17):2388–2399.
41. Hongwei S, Xinzhong H, Huiqin X, et al. Standard deviation of CT radiomic features among malignancies in each individual: prognostic ability in lung cancer patients. *J Cancer Res Clin Oncol* 2023;149(10):7165–7173.
42. Horvat N, Miranda J, El Homsi M, et al. A primer on texture analysis in abdominal radiology. *Abdom Radiol (NY)* 2022;47(9):2972–2985.
43. 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.
44. 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.
45. Kocak B, Akinci D'Antonoli T, Mercaldo N, et al. METHodological RadiomicCs Score (METRICS): a quality scoring tool for radiomics research endorsed by EuSoMII. *Insights Imaging* 2024;15(1):8.
46. Papanikolaou N, Matos C, Koh DM. How to develop a meaningful radiomic signature for clinical use in oncologic patients. *Cancer Imaging* 2020;20(1):33.
47. Orlhac F, Lecler A, Savatovski J, et al. How can we combat multicenter variability in MR radiomics? Validation of a correction procedure. *Eur Radiol* 2021;31(4):2272–2280.
48. Leithner D, Nevin RB, Gibbs P, et al. ComBat Harmonization for MRI Radiomics: Impact on Nonbinary Tissue Classification by Machine Learning. *Invest Radiol* 2023;58(9):697–701.
49. Owens CA, Peterson CB, Tang C, et al. Lung tumor segmentation methods: Impact on the uncertainty of radiomics features for non-small cell lung cancer. *PLoS One* 2018;13(10):e0205003.
50. Whybra P, Zwanenburg A, Andrearczyk V, et al. The Image Biomarker Standardization Initiative: Standardized Convolutional Filters for Reproducible Radiomics and Enhanced Clinical Insights. *Radiology* 2024;310(2):e231319.
51. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed. Springer, 2009.
52. Kubben P, Dumontier M, Dekker A, eds. *Fundamentals of Clinical Data Science*. Springer, 2019.
53. Han YE, Cho Y, Kim MJ, et al. Hepatocellular carcinoma pathologic grade prediction using radiomics and machine learning models of gadoteric acid-enhanced MRI: a two-center study. *Abdom Radiol (NY)* 2023;48(1):244–256.
54. Wilson L, Lin L, Singh K. *The Patient Perspective: Putting the Patient at the Center of the Translational Innovation Process*. *Clin Pharmacol Ther* 2020;107(1):82–84.
55. Ho MP, Gonzalez JM, Lerner HP, et al. Incorporating patient-preference evidence into regulatory decision making. *Surg Endosc* 2015;29(10):2984–2993.
56. Zhang XJ, Song Y, Jiang T, Ding N, Shi TY. Interventions to reduce burnout of physicians and nurses: An overview of systematic reviews and meta-analyses. *Medicine (Baltimore)* 2020;99(26):e20992.
57. Hall LH, Johnson J, Watt I, Tsipa A, O'Connor DB. Healthcare Staff Wellbeing, Burnout, and Patient Safety: A Systematic Review. *PLoS One* 2016;11(7):e0159015.
58. Tsevat J, Moriates C. Value-Based Health Care Meets Cost-Effectiveness Analysis. *Ann Intern Med* 2018;169(5):329–332.