The Diagnostic Accuracy And Clinical Utility Of Radiomics In Oncology Imaging

Rakan Gzian Alotaibi,Turky Nasser A Bin Ateeq,Dowahim Abdullah Aldosrai,Maged Hamad Ali Hakami,Sami Ayesh Alenazi, Amer Ahmed Mohsen Khormi ,Mohammad Hoaded Al Otabei,Abdullah Ayidh Alotaibi,Alreem Hussain Alanazi,Majid Khalaf Alenazi,Zaid Adeeb Alenazi,Faisal Thari Alshammari,Sultan Faleh Alshammari,Naif Hamoud Aldhubayti,Mohammed Hassan Tawhari

Abstract

Tumour features obtained from computer analysis have been included into computer-aided diagnosis (CAD) methods used in medical imaging for many years. The emergence of radiomics, which is an expansion of computer-aided diagnosis (CAD) that involves the use of high-throughput capabilities computer algorithms for obtaining quantitative information from medical images, has generated significant interest in computer-based assessments of healthy or diseased components and activities. Nevertheless, despite the extensive body of radiomic research, the use of radiomics as a therapeutically valuable tool or its approval by the FDA is limited to a very small number of cases. The limited availability of studies in this area can be attributed to several factors, including the use of different imaging as well as radiomic feature identification procedures in each study, the potential challenges in analyzing radiomic information, and the absence of studies demonstrating the positive impact of using radiomic-based tools on the benefit-risk equilibrium for patients. While there are currently several standards on particular areas of acquiring and analyzing radiomic data, there is a lack of a comprehensive roadmap for the whole process of transforming radiomics into practical instruments for clinical treatment.

Keywords: Radiomics, oncology imaging, FDA, computer-aided diagnosis, diagnostic accuracy.

Introduction

Computer-aided diagnosis (CAD) algorithms have used computer-extracted tumor features for many decades to enhance disease detection, diagnosis, treatment planning, and follow-up. Significant advancements have been achieved in breast and lung cancer screening using CAD technology (1-3). Radiomics is a relatively new field that uses computer analysis to quantitatively characterize healthy or diseased structures and processes obtained by medical imaging. It is considered an extension of CAD (4). Like other 'omics' technologies, extracting a large amount of information from images obtained during standard clinical workflows allows for thorough tumor characterization. This also helps in evaluating the differences within tumors, between tumors, and over time.

Like computer-aided design (CAD), radiomics may aid in clinical decision-making. Radiomic features, which are measurements derived from medical images such as CT, MRI, or digital radiography, are integrated with clinical characteristics and data from other omics analyses. This integration aims to identify diseases, predict the probability of death, disease progression, or recurrence within a specific time frame, assess the response to therapy, and determine the most suitable treatment approach. The primary objective of radiomic analyses should be the creation of a test, as defined by the FDA-NIH Biomarker Working Group, which consists of materials, procedures, and criteria for interpretation. This test should be utilized to inform medical decision-making, such as disease diagnosis and management (5-8).

Although there has been a significant surge in research productivity in the last twenty years, the bulk of radiomic investigations have not yet resulted in tests that are clinically beneficial. Out of all medical indications, there are presently 343 tests that use artificial intelligence and machine learning that have been approved by the FDA. However, only a tiny number of these tests are based on radiomics. The absence of clinical application may be attributed to many variables. Most radiomic studies primarily focus on examining the relationships between specific radiomic features and a particular biological or clinical outcome. As a result, the potential benefits of the radiomic test, such as enhanced clinical performance or decreased invasiveness, are often overlooked. Additionally, the clinical utility of the test, which refers to the favorable balance between benefits and risks for the patient when acting upon the information provided, is also frequently disregarded (9-12).

Furthermore, as stated in the statistical and machine learning literature, the examination of high-throughput data, such as radiomics data, is accompanied by various potential problems. These include limited data for the creation and verification of models, as well as the inappropriate use of statistical techniques for the intended test. In addition, many researches have used significantly diverse techniques for the capture of images and the extraction of features. Multiple studies have demonstrated the impact of variations in data collection, image reconstruction, and image post-processing on subsequent analyses. Different software platforms, or even different versions of the same software, can yield significantly different outcomes in terms of the strength and direction of associations between features and outcomes (13).

The current guidelines for obtaining and examining radiomic data consist of a radiomic quality score, which assesses the thoroughness and suitability of the analysis. Additionally, there are computational methods for commonly employed feature types and protocols for image acquisition, feature extraction, and statistical analysis. Nevertheless, radiomics would also gain advantages from a comprehensive plan that outlines the entire process of converting radiomic data into practical tools for guiding clinical care. This plan should include not only guidelines for image acquisition and processing, feature extraction, and statistical analysis, but also considerations such as test standardization and demonstrating the clinical usefulness of the tools. Currently, there is no published roadmap specifically for radiomics. However, there have been efforts to develop criteria and standards for other omics technologies.

Practical use in a medical setting

Before undertaking any formal development and validation, it is essential to specify the intended clinical use of the radiomic test and identify the specific population it is designed for. It is anticipated that the test will be used in clinical care to provide guidance for illness evaluation and treatment choices. This should result in a favorable balance between the benefits and risks, and the test should give advantages over existing tests that serve the same purpose for the target group. The intended clinical application will have significant consequences for the subsequent phases of development and validation, including the selection of features to extract from the imaging data, the optimal timing for

imaging, and the design of the clinical trial to directly evaluate the performance of the test in its intended function.

Purpose and intended audience

Radiomics is often used for either the purpose of screening or the detection of cancer. For instance, MRI radiomics is valuable in diagnosing breast abnormalities13, whereas CT radiomics is effective in detecting lesions in several organs such as the lungs, brain, and prostate (14). Radiomics is being increasingly studied for its potential to predict the clinical outcomes of patients undergoing standard therapy. For instance, CT-based radiomics could be a valuable approach for predicting the outcomes of patients with head and neck squamous cell carcinomas or non-small-cell lung cancer who are receiving standard-of-care therapies (15-17). Radiomic tests can be utilized to aid in treatment selection by serving as assays that determine the effectiveness, or lack thereof, of a particular type of therapies. For instance, a model has been created that uses tumor size, shape, and entropy features on dynamic contrast-enhanced MRI (DCE-MRI) to assess estrogen receptor expression. This model is used to guide treatment decisions for breast cancer patients. Radiomic tests may also be used to evaluate the effectiveness of therapy and track the current condition of the disease (18,19,20).

A full summary of the potential applications of radiomic testing in various roles has been provided elsewhere (21). Under some circumstances, a single radiomic test may serve several purposes. For instance, the previously stated model for assessing oestrogen receptor expression might also be valuable for making prognostic predictions. Nevertheless, the use of radiomic tests in a manner that deviates from their clinically proven usefulness, known as 'off label' usage, is strongly discouraged. The clinical performance criteria are highly dependent on the intended purpose, as is customary in the regulatory clearance and approval procedures for new medications and medical devices. Diagnostic radiomic tests must possess a sufficient degree of precision in order to diagnose diseases.

Prognostic radiomic tests must possess sufficient predictive capability to determine the likelihood of mortality, illness recurrence, or disease development, depending on the specific purpose of the test. Tests intended for treatment selection should also possess the capability to accurately

forecast outcomes, such as mortality or the advancement of illness, in individuals undergoing the therapy in question. In order to choose the most appropriate treatment option, it is necessary to analyze the results of patients who have undergone each therapy. However, if the aim of prediction is simply to identify patients who are most likely to respond to a specific therapy, then the test should have sufficient capability to predict either a response or the level of expression of a well-established predictive biomarker that indicates a response to the treatment of interest. Consequently, the translation procedure described in this Review should be implemented for every job in which a particular radiomic test is expected to be beneficial.

When specifying the target population, it is important to consider several aspects related to the illness, such as the kinds and grades of primary tumors, disease stage, molecular subtypes, risk groups, and receptor expression status, as well as the treatment history. A radiomic test might potentially be beneficial in several target groups. For instance, the test, which is based on the model proposed by Aerts et al.(16), could be valuable in predicting the outcomes of patients with head and neck cancer or non-small-cell lung cancer who are undergoing standard-of-care treatments. Researchers are advised against making assumptions about the effectiveness of a radiomic test for different populations without proper evidence. This is because the technical performance of the imaging device and feature extraction software, as well as the clinical performance of the test, may not be consistent across various populations.

Evaluation of the test's impact on patient well-being in clinical practice

The utility of using a radiomic test should be explicitly delineated within the framework of the existing treatment options for the specific demographic and the availability of other tests fulfilling comparable functions. A radiomic test may be used to categorize patients, allowing for the optimal selection of therapy for each person, hence avoiding the administration of inefficient or needless therapies. A prognostic test meant to inform treatment selection may distinguish between individuals who are expected to have clinical benefits, such as a longer median progression-free survival (PFS) or overall survival length, from a certain therapy or group of treatments, and those who will not.

A prognostic test has the potential to identify patients who are likely to have unfavorable outcomes with standard-of-care therapy and may benefit from a more intense treatment. However, the usefulness of such a test depends on the availability of a suitable alternative treatment. Furthermore, the utilization of a radiomic test could assist in guiding clinical decision-making by minimizing treatment-related toxicities, such as financial burdens. Additionally, prognostic tests can identify patients who have excellent outcomes with standard, well-tolerated treatment regimens, thereby eliminating the need for additional aggressive or toxic treatments or allowing for treatment de-escalation (22-24).

A persuasive rationale is necessary to justify the choice of a radiomic test over other tests that address the same clinical issue. The radiomic test may exhibit higher clinical efficacy compared to a conventional test fulfilling the same function. Radiomic tests have the potential to identify underlying characteristics that are not easily detectable through other methods. For instance, assessing the heterogeneity of oestrogen receptor expression within and between tumors may be less challenging with radiomic tests compared to immunohistochemistry assays. On the other hand, the radiomic test could offer comparable clinical performance while being less invasive (such as avoiding the need for a biopsy), causing less financial strain, providing greater convenience, or reducing one or more associated risks (such as potential harm, discomfort, or exposure during the testing process).

Image processing and the extraction of features

It is necessary to establish standard operating procedures for imaging, which include protocols for administering contrast or imaging agents, specifications for image acquisition, procedures for image processing, and the timing of the scans. Additionally, procedures for feature extraction should be in place, which involve creating a list of quantities to compute from the imaging data, using segmentation algorithms, and employing computational algorithms and software to calculate these quantities (25-28).

The resultant feature measurements should have shown sufficient technical validity. Typically, this would include each feature demonstrating high levels of repeatability and reproducibility, or alternatively, demonstrating great

agreement with a standardized reference measurement of the underlying characteristic, if possible. These artefacts include elements like imaging center, device, operator, and device-calibration parameters, which might impact the distribution of the feature measurements (29-32).

Development and validation of a model

To gather patient-level data, such as images, outcomes, clinical variables, in vitro biomarker measurements, and other relevant information, it is necessary to obtain this data from the target population. This data can be collected either prospectively or retrospectively from completed studies, imaging repositories, or health-care databases. To create a radiomic model, it is necessary to use suitable statistical or machine learning methods while implementing measures to prevent overfitting. In order to demonstrate the reliability of a model in accurately forecasting a certain outcome, it is necessary to use appropriate model validation methodologies (32-34).

Every potential result of the test is thereafter associated with a clear and definite explanation in relation to medical treatment, and it is necessary to demonstrate that the consistency of these results is adequately robust. It is important to establish processes to address drift in the performance of the radiomic test. Drift refers to changes that occur over time due to factors such as changes in image acquisition and processing protocols, feature extraction procedures, software upgrades, obsolescence, and replacement of devices with newer models. These processes should include monitoring and performing technical validation and model adjustment as needed (35-37).

Conclusion

The statistical aspects related to the creation and validation of models, as well as the design of studies to determine clinical usefulness, has many similarities with issues for in vitro tests. Our suggestions draw heavily from these sources. Nevertheless, it is crucial to take into account several significant and far-reaching distinctions that are unique to radiomics. Radiomic approaches are increasingly employing a variety of machine learning and deep learning techniques. This introduces new challenges in terms of establishing standard procedures for extracting features, ensuring test reliability, and interpreting machine learning results, understanding

correlations with biology, addressing regulatory concerns, and assessing the accuracy of the analysis. These criteria are expected to continue developing in the future as researchers gain knowledge of other concerns and as more radiomic models are established, verified, and assessed for their therapeutic usefulness.

It is important to note that these suggestions only apply to the methods and analysis of radiomic studies. They are not meant to serve as reporting criteria for radiomic and CAD investigations, similar to REMARK guidelines for tumor prognostic studies or other reporting guidelines listed by the EQUATOR project. Nevertheless, a few of these suggestions are anticipated to form the foundation of radiomic-specific reporting rules.

Radiomics is becoming more inclined towards using comprehensive machine learning-based image analysis techniques, such as deep learning-based features or the direct application of artificial intelligence and machine learning algorithms to voxel-level data. This transformation is anticipated to reduce the amount of variability caused by human error and enhance the performance of the model in various situations. However, it will also be advantageous to combine it with clinical information in order to customize the test result for each patient. For instance, this kind of test might be used to identify cancer as well as to do so when there are other coexisting medical conditions (such as checking a kidney abnormality in the context of diabetes mellitus, chronic inflammatory processes, and/or hypertension). The enhanced accessibility of diverse data sets should enable these sorts of enhancements.

References

Gillies, R. J., Kinahan, P. E. & Hricak, H. Radiomics: images are more than pictures, they are data. Radiology **278**, 563–577 (2016).

Giger, M. L. Update on the potential of computer-aided diagnosis for breast cancer. Fut. Oncol. **6**, 1–4 (2010).

Doi, K. Computer-aided diagnosis in medical imaging: historical review, current status, and future potential. Comput. Med. Imaging Graph. **31**, 198–211 (2007).

Lambin, P. et al. Radiomics: extracting more information from medical images using advanced feature analysis. Eur. J. Cancer **48**, 441–446 (2012).

FDA-NIH Biomarker Working Group. BEST (Biomarkers, EndpointS, and other Tools) Resource (Food and Drug Administration and National Institutes of Health, 2016).

FDA. Artificial Intelligence and Machine Learning (AI/ML)-Enabled Devices https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices. (2022).

Fornacon-Wood, I. M. et al. Reliability and prognostic value of radiomic features are highly dependent on choice of feature extraction platform. Eur. Radiol. **30**, 6241–6250 (2020).

Radiomics. Radiomics Quality Score – RQS 2.0 https://www.radiomics.world/rqs2 (2022).

Zwanenburg, A. et al. The image biomarker standardization initiative: standardized quantitative radiomics for high throughput image-based phenotyping. Radiology **295**, 328–338 (2020).

Kumar, V. et al. Radiomics: the process and the challenges. Magn. Reson. Imaging **30**, 1234–1248 (2012).

Fournier, L. et al. Incorporating radiomics into clinical trials: expert consensus endorsed by the European society of radiology on considerations for data-driven compared to biologically driven quantitative biomarkers. Eur. Radiol. **31**, 6001–6012 (2021).

McShane, L. M. et al. Criteria for the use of omics-based predictors in clinical trials: explanation and elaboration. BMC Med. **11**, 220 (2013).

Jiang, Y., Edwards, A. V. & Newstead, G. M. Artificial intelligence applied to breast MRI for improved diagnosis. Radiology **298**, 39–46 (2021).

Data Science Institute, American College of Radiology. FDA Cleared AI Algorithms https://www.acrdsi.org/DSI-Services/FDA-Cleared-AI-Algorithms, (2022).

Clark, G. M. Prognostic factors versus predictive factors: examples from a clinical trial of erlotinib. Mol. Oncol. **1**, 406–412 (2008).

Aerts, H. J. W. L. et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat. Commun. **5**, 4006 (2014).

Li, H. et al. Quantitative MRI radiomics in the prediction of molecular classifications of breast cancer subtypes in the TCGA/TCIA data set. NPJ Breast Cancer **2**, 16012 (2016).

Li, H. et al. MRI radiomics signatures for predicting the risk of breast cancer recurrence as given by research versions of gene assays of MammaPrint, Oncotype DX, and PAM50. Radiology **281**, 382–391 (2016).

Cha, K. H. et al. Bladder cancer treatment response assessment in CT using radiomics with deep learning. Nat. Sci. Rep. **7**, 8738 (2017).

Drukker, K. et al. Most-enhancing tumor volume by mri radiomics predicts recurrence-free survival “Early On” in neoadjuvant treatment of breast cancer. Cancer Imaging **18**, 12 (2018).

Huang, E. P., Lin, F. I. & Shankar, L. K. Beyond correlations, sensitivities, and specificities: a roadmap for demonstrating utility of advanced imaging in oncology treatment and clinical trial design. Acad. Radiol. 24, 1036–1049 (2017).

Subramanian, J. & Simon, R. What should physicians look for in evaluating prognostic gene-expression signatures? Nat. Rev. Clin. Oncol. 7, 327–334 (2010).

Shafiq-Ul-Hassan, M. et al. Intrinsic dependencies of CT radiomic features on voxel size and number of gray levels. Med. Phys. 44, 1050–1062 (2017).

Berenguer, R. et al. Radiomics of CT features may be nonreproducible and redundant: influence of CT acquisition parameters. Radiology 288, 407–415 (2018).

Luo, J. et al. A comparison of batch effect removal methods for enhancement of prediction performance using MACQ-II microarray gene expression data. Pharmacogenomics J. 10, 278–291 (2010).

Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical bayes methods. Biostatistics 8, 118–127 (2007).

Orlhac, F. et al. A post-reconstruction harmonization method for multicenter radiomic studies in PET. J. Nucl. Med. 59, 1321–1328 (2018).

Parker, H. S. & Leek, J. T. The practical effect of batch on genomic prediction. Stat. Appl. Genet. Mol. Biol. 11, 10 (2012).

Robinson, K., Li, H., Lan, L., Schacht, D. & Giger, M. Radiomics robustness assessment and classification evaluation: a two-

stage method demonstrated on multivendor FFDM. Med. Phys. 46, 2145–2156 (2019).

The Cancer Imaging
Archive http://cancerimagingarchive.net (2020).

Clark, K. et al. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. J. Digital Imaging 26, 1045–1057 (2013).

Zhu, Y. et al. Deciphering genomic underpinnings of quantitative MRI-based radiomic phenotypes of invasive breast carcinoma. Nat. Sci. Rep. 5, 17787 (2015).

Riley, R. D. et al. Minimum sample size for developing a multivariable prediction model: part II — binary and time-to-event outcomes. Stat. Med. 38, 1276–1296 (2018).

Riley, R. D. et al. Minimum sample size for external validation of a clinical prediction model with a binary outcome. Stat. Med. 40, 4230–4251 (2021).

Riley, R. D. et al. Minimum sample size calculations for external validation of a clinical prediction model with a time-to-event outcome. Stat. Med. 41, 1280–1295 (2022).

Cho, J., Lee, K., Shin, E., Choy, G. & Do, S. How much data is needed to train a medical image deep learning system to achieve necessary high accuracy? Preprint at https://doi.org/10.48550/arXiv.1511.06348 (2015).

Whitney, H., Li, H., Ji, Y., Liu, P. & Giger, M. L. Comparison of breast MRI tumor classification using human-engineered radiomics, transfer learning from deep convolutional neural networks, and fusion methods. Proc. IEEE 108, 163–177 (2020).