

80204_Auto_Edited.docx

Name of Journal: *Artificial Intelligence in Medical Imaging*

Manuscript NO: 80204

Manuscript Type: MINIREVIEWS

Radiomics: Status quo and future challenges

Zhi-Yun Jiang, Li-Shuang Qi, Jia-Tong Li, Nan Cui, Wei Li, Wei Liu, Ke-Zheng Wang

Abstract

18

Noninvasive imaging (computed tomography, magnetic resonance imaging, endoscopic ultrasonography, and positron emission tomography) as an important part of the clinical workflow in the clinic, but it still provides limited information for diagnosis, treatment effect evaluation and prognosis prediction. In addition, judgment and diagnoses made by experts are usually based on multiple years of experience and subjective impression which lead to variable results in the same case. With accumulation of medical imaging data, radiomics emerges as a relatively new approach for analysis. Via artificial intelligence techniques, high-throughput quantitative data which is invisible to the naked eyes extracted from original images can be used in the process of patients' management. Several studies have evaluated radiomics combined with clinical factors, pathological, or genetic information would assist in the diagnosis, particularly in the prediction of biological characteristics, risk of recurrence, and survival with encouraging results. In various clinical settings, there are limitations and challenges needing to be overcome before transformation. Therefore, we summarize the concepts and method of radiomics including image acquisition, region of interest segmentation, feature extraction and model development. We also set forth the current applications of radiomics in clinical routine. At last, the limitations and related deficiencies of radiomics are pointed out to direct the future opportunities and development.

1.INTRODUCTION

Radiomics was first proposed by Lambin et al. in 2012, which converts medical images into high-throughput quantitative features^[1]. Radiomic features can capture tissue and lesion properties noninvasively, such as shape and heterogeneity, and radiomics acts as a new approach to extract the information underlying the medical images that fail to be appreciated by naked eyes^[2]. In the meantime, radiomics also possesses several advantages over molecular assays, such as being ⁵ non-tissue-destructive, rapid analysis, easily serialized, fairly inexpensive, and being fully compatible with the existing clinical workflows^[3]. In 2014, Aerts et al. demonstrated the role of radiomics in disease prognostication^[4], promoting the development of radiomic-based signatures. Subsequently, the Pyradiomics framework based on the image biomarker standardization initiative (IBSI) criteria published in 2017 strongly supported the standardized application of radiomics^[5].

Radiomics has evolved tremendously in the last decade, with the objective of precision medicine. However, the interpretability of radiomic-based signatures and the correlation with biology and pathology need to be further discussed. Additional multi-center data and prospective validation are also required for verification, in order to improve the confidence of applications^[6]. There are still several substantial barriers to realize the objective of transforming artificial intelligence (AI) into the real clinical practice.

¹ In the present study, the basic principles and methodologies of radiomics were reviewed and an outline of the representative clinical utilization was provided to highlight the benefits of radiomics in diagnosis, staging, tumor biological features, and prognosis. Additionally, it is essential to explore the deficiencies of radiomics to achieve a balanced interpretation between AI and clinical practice.

2.CONCEPT AND METHODOLOGIES

¹¹ “Radiomics,” a term that describes the “omics” approach for the analysis of imaging data, has emerged as a novel tool for diagnosis and prognosis^[2]. Using advanced

6
computational tools, high-throughput quantitative imaging features beyond inspections of naked human eyes are extracted and the desensitized medical images are transformed into multiple textural features for quantitative assessment^[7-9]. With semantic features, radiomics enables clinicians to make more objective and accurate clinical decisions in diagnosis and prognosis^[10,11]. The workflow of radiomics analysis, consisting of several steps, is illustrated in Figure 1.

2.1 IMAGE ACQUISITION

Image acquisition is approved by the ethics committee and informed consent form is signed by participants or their close relatives. The right to know patients is protected by relevant regulations. As the research of radiomics concentrated on human participants, it complies with the basic principles of 1964 《Helsinki Manifesto》 and its later revisions. Sensitive information is erased from medical imaging data exported from imaging databases, including but not limited to organization name, organization address, physician's name, patient's name, patient's birthday, etc. Besides, personal data are kept confidential, such as ID number, home address, contact information, medical insurance information, etc. Acquisition, transmission, and use of data should meet relevant legal requirements.

In addition, medical imaging data, which are consistent with standard imaging protocols, are the foundation of radiomics^[12,13]. It can be single- or multi-center, and retrospective or prospective. Although there are various types of imaging examinations, including 15
computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET), ultrasound (US), etc.^[11,14-16] for different research purposes, the dominant examination methods or sequences are more recommended. Hence, more eligible cases are included to find out common features, which may contribute to the stability of models^[17]. There is no general standard for the medical imaging data from different examination methods using different imaging methods, acquisition methods, imaging parameters, and imaging quality that may affect the subsequent analysis.

Therefore, how to normalize the data and conform to the imaging standard is the focus of radiomics studies at present.

After data collection, the data need to be checked and confirmed, in order to correct or eliminate unqualified data. The specific inspection content includes the validity of the file format, the integrity of the sequence, and the correctness of the image content, in order to exclude unrecognizable images, sequence deletion, and wrong image layers. More detailed image quality specifications can also be formed according to specific research requirements. In the process of image quality control, it is necessary to sort out the imaging problems encountered, so that the data can be traced back when the inclusion and exclusion criteria are defined.

2.2 PREPROCESSING

Because of different scanning parameters, reconstruction procedures (slice thickness, voxel size, and reconstruction algorithm), and inconsistent imaging acquisition of multi-brand manufactories, it has a significant influence on distribution of features^[18,19]. In order to decrease this discrepancy, preprocessing of the collected imaging data is essential. At present, the most common methods include resampling, gray-level discretization, and intensity normalization. Image resampling involves generation of equal-size voxels by applying the linear interpolation algorithm to improve image quality and to eliminate bias introduced by non-uniform imaging resolution^[20]. Gray-level discretization refers to the bundling of pixels based on their density, either by relative discretization (fixed number) or absolute discretization (fixed size)^[21]. Image intensity normalization is used to correct inter-subject intensity variation by transforming all images from original greyscale into a standard greyscale. Furthermore, image enhancement approaches, such as image flipping, image rotation, image distortion, image transformation, and image scaling, can enrich data diversity, improve model generalization ability, and reduce the risk of model overfitting.

In addition to the above-mentioned methods, not only for images, we also need to preprocess clinical data. Deidentification of data is beneficial to protect personal

information and query data among multiple departments. Hospital number is advised to be the unique identification, realizing the mapping of images. In order to effectively eliminate the deficiency of data inconsistency and bias in multi-center studies, it is necessary to conduct data consistency processing, which is advantageous to realize cross-center data modeling and verification. The methods of data consistency processing include: (a) Standardization of data collection: Data are collected according to the unified data acquisition standard in each center; (b) Consistency processing based on extracted features: The method of Z-score can be used to standardize data; (c) Consistency processing based on image domain: According to the annotated information, the size of region of interest (ROI) is kept consistent.

2.3 SEGMENTATION

Segmentation of ROI can be divided into manual and semiautomatic/automatic segmentation, two-dimensional (2D) and three-dimensional (3D) segmentation, and intratumoral and peritumoral segmentation^[22-26]. This process is relatively tedious and requires open-source or dedicated software to support^[12]. The process at least needs one labeling physician and one senior physician. The knowledge of relevant anatomy and imaging should be well known by labeling physicians and they must be familiar with the sketching software. In addition, for manual segmentation, intra-class correlation coefficient and concordance correlation coefficient can be advantageous to reduce the discrepancy of subjective judgement and the intra- and inter-reader variability^[17,27]. Due to the rapid development of computer science, semiautomatic/automatic segmentation has been frequently applied. Automatic segmentation aims to draw ROIs automatically^[28], while semiautomatic segmentation still requires partially manual intervention to mark the center of the lesion before automatic segmentation^[29]. They both decrease instability to a certain extent, however, they are less applied because of technical restriction. At present, automatic segmentation can be summarized into three categories^[30]: (a) algorithms based on intensity thresholds and regions; (b) algorithms

based on statistical approaches and deformable models; (c) algorithms incorporating empirical knowledge into the segmentation process.

2.4 FEATURE EXTRACTION

Features are extracted from ROIs using different software with the similar code, which consist of first-order, second-order, and higher-order features. First-order features describe the geometric attributes and the distribution of voxel intensities of the ROIs, including mean, median, maximum, and minimum values, as well as the skewness, kurtosis, and entropy. Second-order features represent the relationships between adjacent voxels to measure features^[31]. Second-order textural features describe the gray-scale alterations and are extracted by different algorithms. Higher-order features are extracted via wavelet, Laplacian, and Gaussian filters from multiple dimensions^[32]. With the combination of multiple omics, semantic features, which are based on the experience and knowledge of radiologists, pathological features, genetic features, etc., all promote the transformation of radiomics into clinical practice. In recent years, depiction of deep learning (DL)-based features, which are supplementary high-dimensional features, by observers has been reported as a challenge^[33]. Although DL-based features reveal certain advantages in terms of estimating prognosis of malignancies, it is enslaved to be widely used by data size and technological development.

2.5 FEATURE SELECTION

According to the fourth step (feature extraction), the great number of extracted features is achieved, and how to select the most relevant features is the key to establish a robust radiomics model. This process simplifies the mathematical problem by decreasing the number of parameters and also reduces the risk of overfitting. Specific methods include univariate, the least absolute shrinkage and selection operator (LASSO), RELIEF algorithm, redundancy maximum relevance (MRMR), etc.^[34].

2.6 MODELING AND VERIFICATION

The ultimate objective of radiomics is to establish an effective model for classification and prediction. The data should be clustered into training and validation datasets. Different classifiers, including logistics, ¹⁶ support vector machine (SVM), Bayes, k-Nearest Neighbor (KNN) algorithm, Tree and Forest, are used to set up models and to select the most effective model by seed circling for clinical transformation^[35]. Meanwhile, the predictive performance of the final model should be verified on a separate cohort, and an external validation cohort is highly appropriate to confirm its generalization. Owing to the lack of data sharing, obtaining the results of external validation of the model is a challenge at this stage.

3. CLINICAL APPLICATION OF RADIOMICS

3.1 DIAGNOSIS AND STAGING

In previous studies, radiomics has shown a great potential in the diagnosis and staging of different diseases. Although the diagnosis of some lesions is easy according to imaging manifestations, radiomics can improve physicians' diagnostic confidence and patients' examination strategies. In a plain CT study, ¹⁴ 168 patients with hepatocellular carcinoma (HCC) and 117 patients with hepatic hemangioma were analyzed. Textural features were extracted from plain CT images and 13 features were selected from 1,223 candidate features to constitute the radiomics signature, in order to establish a logistic regression model to classify benign and malignant liver tumors. The final ¹⁷ model achieved an average area under the curve (AUC) of 0.87. In spite of the lack of innovation, it helps patients who cannot successfully undergo contrast-enhanced CT (CECT) because of iodine contrast agent allergy for a relatively accurate diagnosis^[36].

In another study, Ding et al. explored the capacity of the combined ¹⁰ model for differentiating HCC from focal nodular hyperplasia (FNH) in non-cirrhotic livers using Gd-DTPA contrast-enhanced MRI. For this purpose, 8 radiomics features were selected for the radiomics model, and 4 clinical factors (age, gender, hepatitis B surface antigen ¹⁰ (HbsAg), and enhancement pattern) were chosen for the clinical model. The combined model was established using the factors from the previous models. The classification

accuracy of the combined model that differentiated HCC from FNH in both the training and validation datasets was 0.956 and 0.941, respectively. The model could support clinicians to make more reliable clinical decisions^[37].

Serous cystadenomas (SCN) are considered as mostly benign cystic neoplasm in the pancreas. Mucinous cystic neoplasm (MCN) is an easily misdiagnosed lesion of SCN, which is associated with the risk of malignant transformation^[38]. Therefore, Xie et al. confirmed the value of CT-based radiomics analysis in preoperatively discriminating pancreatic MSN and SCN. A total of 103 MCN and 113 SCN patients who underwent surgery were retrospectively enrolled. The Rad-score model was proved to be robust and reliable (average AUC, 0.784; sensitivity, 0.847; specificity, 0.745; positive-predictive value (PPV), 0.767; negative-predictive value (NPV), 0.849; accuracy, 0.793), which could serve as a novel tool for guiding clinical decision-making^[39].

In another multi-center study, researchers took advantages of radiomics to develop a nomogram for preoperatively predicting grade 1 and grade 2/3 tumors in patients with pancreatic neuroendocrine tumors (PNETs). Totally, 138 patients from two institutions with pathologically confirmed PNETs were included in that retrospective study. The nomogram integrating an independent risk factor of tumor margin and fusion radiomic signature showed a strong discrimination with an AUC of 0.974 (95% confidence interval (CI): 0.950–0.998) in the training cohort and 0.902 (95% CI: 0.798–1.000) in the validation cohort, with a satisfactory calibration. Decision curve analysis (DCA) verified the clinical applicability of the predictive nomogram^[40].

3.2 EVALUATION OF TUMOR BIOLOGICAL BEHAVIORS

Concurrent advancements in imaging and genomic biomarkers have facilitated identification of noninvasive imaging surrogates of molecular phenotypes. Villanueva et al. investigated the genomic features of HCC and peritumoral tissues that were associated with patients' outcomes, and they explored the relationship between imaging traits and genomic signatures^[41]. Patients who underwent pre-operative CT or MRI and transcriptome profiling were assessed using 11 qualitative and 4 quantitative (size,

enhancement ratio, wash-out ratio, tumor-to-liver contrast ratio) imaging traits. Several imaging traits, including infiltrative pattern and macrovascular invasion were found to be associated with gene signatures of aggressive HCC phenotype, such as proliferative signatures and CK19 signature.

Microvascular invasion (MVI) is one of the strongest predictors of hepatic transplantation or hepatectomy for HCC, which is one of the independent factors for early recurrence and poor prognosis^[42]. MVI could be diagnosed postoperatively and it was defined as the presence of tumor within microscopic vessels of the portal vein, hepatic artery, and lymphatic vessels^[43]. Conventional imaging methods cannot reveal MVI because of the poor resolution before operation. Therefore, it is important to develop a non-invasive tool to detect MVI for clinical decision-making. Zhang et al. proposed a nomogram for the prediction of MVI that included a radiomic score and alpha fetoprotein, tumor type, peritumoral enhancement, arterial rim, and internal arteries. This nomogram was superior to a clinical and radiologic model with an AUC of 0.858 versus 0.729^[44]. In another research, Renzulli et al. demonstrated that non-smooth tumor margins and peritumoral enhancement, combined with the radio-genomic features were independent predictors for MVI with a PPV of 0.95^[45]. In a large-scale study, Xu et al. collected CT scan images from 495 patients and developed a combined model which consisted of semantic features (aspartate aminotransferase, alpha fetoprotein (AFP), non-smooth tumor margin, extrahepatic growth, ill-defined pseudocapsule, and peritumoral arterial enhancement) and radiomic features to predict histological MVI, with an AUC of 0.909 and 0.889 in the training cohort and the test cohort, respectively^[46].

Gao et al. assessed the preoperative prediction of TP53 status based on multiparametric MRI (mp-MRI) radiomic features extracted from 3D images. In total, 57 patients with pancreatic cancer who underwent preoperative MRI were included. The 3D ADC-ap-DWI-T2WI model with 11 selected features yielded the best performance for differentiating TP53 status, with an accuracy of 0.91 and an AUC of 0.96. The model revealed a good calibration, and the DCA proved the clinical value of the model. The radiomics model derived from mp-MRI provided a non-invasive, quantitative method to

predict mutational status of TP53 in patients with pancreatic cancer that might contribute to the precision treatment^[47].

3.3 PROGNOSIS

Current guidelines recommend surgical resection as the first-line therapy for patients with HCC^[48]. However, postoperative recurrence rate remains high and there is no reliable prediction tool. In a multi-center study, the potential of radiomics coupled with machine learning algorithms was assessed to improve the predictive accuracy for HCC recurrence. Using the machine learning framework, they identified a three-feature signature that demonstrated a favorable prediction of HCC recurrence across all datasets, with C-index of 0.633-0.699. AFP, albumin-bilirubin, hepatic cirrhosis, tumor margin, and radiomic signature were selected for developing a preoperative model; the postoperative model incorporated satellite nodules into the above-mentioned predictors. The two models showed a superior prognostic performance, with C-index of 0.733-0.801 and integrated Brier score of 0.147-0.165, compared with rival models without radiomics, and are widely used in staging systems. Combined with clinical data, a three-feature fusion signature generated by aggregated ML-based framework could accurately predict individual recurrence risk, enabling appropriate management and surveillance of HCC (49). In another study, CECT with measurement of Gabor and Wavelet radiomics features in patients with a single HCC tumor treated by hepatectomy revealed that several features were associated with both overall survival (OS) and disease-free survival (DFS) (P-values < 0.05) (50). Similar results were reported by a separate study that risk scores developed from radiomics nomograms obtained from CECT textural data overmatched traditional clinical staging systems in both the training and validation cohorts for both tumor recurrence and OS^[51].

Patients with pancreatic cancer have a poor prognosis, therefore, it is necessary to identify tumor characteristics associated with prognosis. Toyama et al. enrolled 161 patients with pancreatic cancer who underwent FDG-PET/CT before treatment. The area of the primary tumor was semi-automatically contoured with a threshold of 40% of the

maximum standardized uptake value, and 42 PET-based features were extracted. Among the PET parameters, 10 features showed statistical significance for predicting OS. Multivariate Cox regression analysis revealed gray-level zone length matrix (GLZLM)-gray-level non-uniformity (GLNU) as the only PET parameter showing statistical significance. In the random forest (RF) model, GLZLM-GLNU was the most relevant factor for predicting 1-year survival, followed by total lesion glycolysis (TLG). Radiomics with machine learning using fluorodeoxyglucose (FDG)-PET in patients with pancreatic cancer provided valuable prognostic information^[16].

4. DISCUSSION

There is no doubt that radiomics as a newly emerged quantitative technique is burgeoning in disease management. Nevertheless, the majority of the research of radiomics encountered common problems, and whether the radiomic-based signatures can be used in clinical practice needs to be discussed.

Reproducibility is one of the primary challenges that radiomic techniques must overcome for clinical application. At present, imaging protocols are not standardized worldwide, and hence, variability in image acquisition and reconstruction parameters is inevitable in clinical practice. A recent study demonstrated that the quantitative values of radiomic features varied according to imaging protocols^[52]. In addition, although IBSI seeks standardization for radiomic extraction, the differences in techniques or platforms adopted in different centers may lead to differences in feature values^[5], propagating to the radiomic signatures. Most radiomic signatures have a sharp drop in performance from training cohort to validation cohort. Researchers have adopted data normalization methods to correct for multicenter effects, such as ComBat harmonization^[53]. However, whether the radiomic-based signature developed by normalized radiomic features is appropriate for clinical practice has not yet been studied. It is urgent to develop a reproducible radiomic signature that could overcome inherent multicenter effects, which is the basis for clinical individualized application.

Data sharing for independent validation is a challenge for radiomic signatures. To date, studies have mainly developed and validated the radiomic signatures using imaging data derived from their own center or multiple centers according to the same imaging protocols^[54]. However, whether the signatures would be effective in completely independent centers needs further validation. Although images are more readily available than tissue molecular assays, the current open radiomic datasets are not enough for the independent validation. To eliminate this deficiency, data sharing among institutes and hospitals around the country or even around the world is important for radiomics, although it presents complex logistical problems. The Cancer Imaging Archive (TCIA) provides a good example of data sharing with a large portion of clinical data^[55], and it is still growing with contribution from different institutes and hospitals. A previous study indicated that signatures should be validated using an open dataset that could become the standard to demonstrate their effectiveness^[9].

Biological interpretability of radiomic signatures would accelerate their clinical application. Clinical experts mainly assume the radiomic model as a black box that can provide promising prediction results for clinical outcomes, which may make radiomics as a less accepted approach. The problem is further aggravated in the context of deconvolutional neural or DL networks, which even lack the observable model that solely concentrates on maximizing performance. A great number of these so-called 'black-box' approaches may be perfectly viable in the diagnostic setting; however, when it comes to radiomic signatures for optimizing treatment, the question of interpretability becomes more paramount because a biomarker-driven treatment decision needs an explanation rooted in pathophysiology^[56]. The emergence of radio-genomics provides a bridge for linking the radiomics to the underlying biological progression. The biological interpretability may provide biological evidence for the predictive ability of the radiomic signatures.

Clinical operability is the key in the clinical adoption of prognostic and predictive radiomic tools. To date, radiomic-based studies have mainly concentrated on developing robust signatures, and their application details in clinical practice are lack. Therefore,

translating the computer language into a simple software or system may be an effective method to promote clinical application of radiomics.

5.CONCLUSION

In conclusion, the current researches have achieved encouraging results of radiomics and revealed potential of clinical applications, while poor standardization and generalization of radiomics limit the further translation of this method into clinical routine. How to make reproducibility of data, multi-center data sharing, biological interpretability of radiomic signatures and clinical operability come true, will become the crucial issue for development of radiomics. Only then will radiomics be more comparable and increase reliability to get clinician's approval. In foreseeable future, the development of radiomics will occupy a significant position in personalization and precision medicine. At present, it is more important to make clinical participants be conscious of benefits and limitations of radiomics in order to obtain reasonable decision towards clinical practice.

27%

SIMILARITY INDEX

PRIMARY SOURCES

1	link.springer.com Internet	146 words — 4%
2	pubmed.ncbi.nlm.nih.gov Internet	126 words — 3%
3	doctorpenguin.com Internet	95 words — 2%
4	www.frontiersin.org Internet	93 words — 2%
5	www.nature.com Internet	88 words — 2%
6	Jingwei Wei, Hanyu Jiang, Dongsheng Gu, Meng Niu, Fangfang Fu, Yuqi Han, Bin Song, Jie Tian. "Radiomics in Liver Diseases: Current Progress and Future Opportunities", Liver International, 2020 Crossref	76 words — 2%
7	www.thno.org Internet	76 words — 2%
8	www.e-cmh.org Internet	74 words — 2%
9	www.researchgate.net Internet	

74 words — 2%

10 lib.smu.edu.cn
Internet

63 words — 2%

11 Emily Harding - Theobald, Jeremy Louissaint, Bharat Maraj, Edward Cuaresma et al. "Systematic review: Radiomics for the diagnosis and prognosis of hepatocellular carcinoma", Alimentary Pharmacology & Therapeutics, 2021
Crossref

60 words — 2%

12 Gu-Wei Ji, Fei-Peng Zhu, Qing Xu, Ke Wang, Ming-Yu Wu, Wei-Wei Tang, Xiang-Cheng Li, Xue-Hao Wang. "Machine-learning analysis of contrast-enhanced CT radiomics predicts recurrence of hepatocellular carcinoma after resection: A multi-institutional study", EBioMedicine, 2019
Crossref

25 words — 1%

13 Jingwei Wei, Hanyu Jiang, Dongsheng Gu, Meng Niu, Fangfang Fu, Yuqi Han, Bin Song, Jie Tian. "Radiomics in liver diseases: Current progress and future opportunities", Liver International, 2020
Crossref

20 words — 1%

14 content.iospress.com
Internet

17 words — < 1%

15 bmccancer.biomedcentral.com
Internet

13 words — < 1%

16 journal.ipb.ac.id
Internet

13 words — < 1%

17 Qi Lang, Chongli Zhong, Zhiyun Liang, Yizhou Zhang, Baokang Wu, Feng Xu, Ling Cong, Shudong Wu, Yu Tian. "Six application scenarios of artificial

12 words — < 1%

intelligence in the precise diagnosis and treatment of liver cancer", Artificial Intelligence Review, 2021

Crossref

18 tarj.in
Internet

12 words — < 1%

EXCLUDE QUOTES ON
EXCLUDE BIBLIOGRAPHY ON

EXCLUDE SOURCES OFF
EXCLUDE MATCHES < 12 WORDS