The Applications of Radiomics in Precision Diagnosis and Treatment of Oncology: Opportunities and Challenges

Zhenyu Liu1,5*, Shuo Wang1,5*, Di Dong1,5*, Jingwei Wei1,5*, Cheng Fang3*, Xuezhi Zhou1,4, Kai Sun1,4, Longfei Li1,6, Bo Li3✉, Meiyun Wang2✉, Jie Tian1,4,7✉

1. CAS Key Laboratory of Molecular Imaging, Institute of Automation, Beijing, 100190, China
2. Department of Radiology, Henan Provincial People’s Hospital & the People’s Hospital of Zhengzhou University, Zhengzhou, Henan, 450003, China
3. Department of Hepatobiliary Surgery, The Affiliated Hospital of Southwest Medical University, Luzhou, Sichuan, 646000, China
4. Engineering Research Center of Molecular and Neuro Imaging of Ministry of Education, School of Life Science and Technology, Xidian University, Xi’an, Shaanxi, 710126, China
5. University of Chinese Academy of Sciences, Beijing, 100080, China
6. Collaborative Innovation Center for Internet Healthcare, Zhengzhou University, Zhengzhou, Henan, 450052, China
7. Beijing Advanced Innovation Center for Big Data-Based Precision Medicine, Beihang University, Beijing, 100191, China

*These authors contributed equally to this work.

✉Corresponding authors: Jie Tian, Address: Institute of Automation, Chinese Academy of Sciences, Beijing, 100190, China. E-mail: jie.tian@ia.ac.cn or Bo Li, Address: Department of Hepatobiliary Surgery, the Affiliated Hospital of Southwest Medical University, Luzhou, Sichuan, 646000, China. E-mail: liboer2002@126.com or Meiyun Wang, Address: Henan Provincial People’s Hospital & the People’s Hospital of Zhengzhou University, Zhengzhou, Henan, 450003, China. E-mail: marian99940163.com

© Ivyspring International Publisher. This is an open access article distributed under the terms of the Creative Commons Attribution (CC BY-NC) license (https://creativecommons.org/licenses/by-nc/4.0/). See http://ivyspring.com/terms for full terms and conditions.

Received: 2018.09.30; Accepted: 2019.01.10; Published: 2019.02.12

Abstract

Medical imaging can assess the tumor and its environment in their entirety, which makes it suitable for monitoring the temporal and spatial characteristics of the tumor. Progress in computational methods, especially in artificial intelligence for medical image process and analysis, has converted these images into quantitative and minable data associated with clinical events in oncology management. This concept was first described as radiomics in 2012. Since then, computer scientists, radiologists, and oncologists have gravitated towards this new tool and exploited advanced methodologies to mine the information behind medical images. On the basis of a great quantity of radiographic images and novel computational technologies, researchers developed and validated radiomic models that may improve the accuracy of diagnoses and therapy response assessments. Here, we review the recent methodological developments in radiomics, including data acquisition, tumor segmentation, feature extraction, and modelling, as well as the rapidly developing deep learning technology. Moreover, we outline the main applications of radiomics in diagnosis, treatment planning and evaluations in the field of oncology with the aim of developing quantitative and personalized medicine. Finally, we discuss the challenges in the field of radiomics and the scope and clinical applicability of these methods.

Key words: radiomics, medical imaging, precision diagnosis and treatment, oncology

Introduction

Multimodality medical images contain a great deal of valuable information reflecting the development and progression of cancer. Advancements in data mining and machine learning make it possible to extract many quantitative features and convert the rapidly increasing number of medical images into minable data. This comprehensive method used to analyze medical images is known as radiomics. The concept of radiomics was first proposed in 2012 [1, 2], and has since attracted the attention of researchers throughout the world (Figure 1).

Initially, radiomics refers to extracting large amounts of high-dimensional quantitative features from multimodality medical images such as those of
computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET) and ultrasonography (US), then the mining of correlations between these features and the diagnosis/prognosis of cancer [3, 4]. These correlations are revealed with imaging analyses, and tumors can then be decoded into different imaging phenotypes [5]. With complimentary information from clinical reports, treatment responses, and genomic/proteomic assays, radiomics may reflect the global outlook of cancer [6]. Compared with previous ways that process medical images as pictures for visual inspection, radiomics introduces a new way to mine the information contained in the medical images. Radiomics has been rapidly developed toward clinical application [7-9] in the hope that it will advance precision diagnostics and cancer treatment.

Radiomics may provide quantitative and objective support for decisions surrounding cancer detection and treatment [10]. This is accomplished with obtaining quantitative information from medical images, combining imaging features with clinical information, genomic information and other information, and mining these data to detect radiomic biomarkers. Radiomics incorporates a series of computational technologies, and the methodologies used in radiomics are usually oriented for clinical problems. Great progress has been made within the field of radiomics regarding technology that meets clinical requirements and benefits diagnostics and cancer treatments.

Here, we will review recent developments in the methodologies in radiomics, mainly involving feature detection and model construction. We will also discuss the scope and the challenges associated with radiomics for precision diagnostics and cancer treatment. Finally, we will offer our opinion on the future steps required for better acceptance of radiomics.

## Development of Methodology and Technology in Radiomics

Radiomics involves several essential steps (see pipeline of radiomics in Figure 2 and Figure 3), including data acquisition and preprocessing, tumor segmentation, feature extraction, knowledge discovery, and modeling. A series of advancements have been made in each of these areas.

### Data acquisition and preprocessing

Radiomics usually begins with data acquisition. It relies on a large volume of medical images and corresponding clinical data to reveal the correlation that exists between them. Radiomics was first proposed using CT images [1] and soon after was applied in the analysis of MR images [2]. There were also studies based on PET [11, 12] and US [13] images utilizing similar strategy. Medical images used for radiomic analysis are collected from different hospitals or data centers; thus, these images are usually obtained using different parameters and protocols and reconstructed with different software. The differences may bring unexpected influences on the radiomic model [14]. The influence of different image acquisition parameters on the reproducibility and quality of the extracted features in different modalities including CT, PET, MRI and US as well as the preprocessing of multi-modality images have been discussed in several recent literatures and a systematic review published by Larue et al. in 2017 [15].

For CT images, differences in CT scanners, voxel size and reconstruction kernels are the main source of influence [16-24]. Shafiq-ul-Hassan et al. [16] suggested that resampling the pixel size could minimize the dependency of radiomic features on voxel size. In addition, applying normalization on the gray level and voxel size in radiomic studies could improve the robustness of features [17]. Zhao et al. [19] found that smooth and sharp reconstruction kernels should not co-occur in a study, as combining data that includes both two kernels could lead to uncontrollable discrepancies. Another recent study by Mackin et al. found that variations in tube current do not significantly affect the radiomic features [23]. A most recent study comprehensively demonstrated that using a controlled protocol can reduce variability in radiomic features, while resampling image thickness cannot reduce the adverse effect derived from thickness variations [24].

There are also several studies focused on the influence of image acquisition on features extraction for PET images. In 2010, Galavis et al. has shown that most of the quantitative textural features presented large variations due to the changes in the grid size, reconstruction algorithm and number of iterations.

![Figure 1. Publication statistics of radiomics since 2012. The number of publications is going straight up. Abbreviations: CT, Computed Tomography; MRI, Magnetic Resonance Imaging; PET, Positron Emission Tomography](http://www.thno.org)
Recent results from Yan et al. [26] and Veld et al. [27] further confirmed this finding. Recently, Gallivan et al. [28] found that reconstruction settings largely affects the stability of radiomic features, and only 26% were stable when different parameters were considered in this anthropomorphic phantom study. Respiratory motion is another factor that should be concerned. Oliver et al. [29] and Grootjans et al. [30] showed that respiratory motion significantly affects the quantitative textural features reflecting tumor heterogeneity with PET. In addition, they found that standard uptake value (SUV) discretization could also impact the results of textural features detecting. Compare to three-dimensional (3D)-PET, respiratory-gated (4D) PET can reduce motion blurring and generate more robust features [29, 31, 32].

The effects of MRI field strength, imaging protocols, and manufacturers on radiomic features have been investigated by several studies using either living samples or physical phantoms [33-38]. For instance, Jirak et al. [34] and Mayerhoefer et al. [35] suggested that some features were sensitive to MR acquisition parameters including TR, TE and bandwidth, which was confirmed by a recent simulation study [39]. Savio et al. [36] investigated the effect of slice thickness on texture-based model in differentiating plaque from normal tissue, and showed that the moderate differences caused by thickness does not affect the classification results. The research conducted by Waugh et al. [40] showed that differences of acquisition parameters may not significantly affect the ability of texture analysis in classification task. Moreover, some mature algorithms and tools could help to remove unwanted low-frequency intensity nonuniformity [41, 42] and perform image intensity normalization [43-45] in order to reduce the difference caused by different acquisition parameters of brain MRI.

US-derived quantitative features have been proved to be strongly correlated with breast biologic characteristics [46], gestational age [47], neonatal respiratory morbidity [48]. They can also be used to detecting malignant tumors from benign tumors in thyroid [49, 50] and breast [51, 52]. In a review in terms of ultrasound image processing methods for carotid plaque morphology analysis [53], Kyriacou et al. demonstrated that a standard acquisition protocol, image normalizing method helps to generate reproducible measurements. So far, there are still few studies concerning on the reproducibility and stability of quantitative features obtained from US in oncology.

Considering the influence of different image acquisition parameters, researchers should pay more attention on the differences of imaging protocols and provide essential parameters to achieve reproducibility and comparability with other radiomic studies [10]. As the importance of using standard imaging protocols should be taken into account [6], researchers should develop specific methods aiming at one disease compared with others.

**Figure 2.** The initially radiomics pipeline with medical images. Reproduced with permission from [4]. (a) Example CT images of patients with lung cancer. (b) Strategy of radiomic analysis.
Tumor segmentation

In radiomic analysis, it is the entire primary tumor that is usually analyzed. Tumor segmentation determines which region will be analyzed further, so this becomes a fundamental step in radiomics. Segmentation includes manual, semiautomatic, and automatic segmentation methods. However, regardless of the method used, there will be certain challenges. First, there is no golden standard for tumor segmentation [2], and the manual segmentation is quite boring and time consuming. Radiologists have to go through every slice of the imaging sequence to delineate the regions of interest, and a mask for a large tumor may contain dozens of slices. Second, there exist numerous morphological variations as tumors are very different from geometric objects, and the variations are usually difficult to model. Third, tumor margins could be blurred by the partial volume effect and not well defined in medical images. Furthermore, segmentation methods especially their reproducibility and reliability are very important. For instance, inter-reader variability of the interpretation for in-plane nodule boundaries could be very large [54]. Several researches have provided that features can be changed with different delineation. For example, van Velden et al. found 25 features were sensitive to the change in delineation using two baseline whole-body PET/CT scans from eleven NSCLC patients [27]. Also Leijenaar et al. found that 71% of radiomic features were stable in a test-retest cohort of 11 NSCLC patients using a semiautomatic segmentation, and 91% of radiomic features were stable in an inter-observer cohort of 23 NSCLC patients with manual segmentation [12]. Since the features are extracted based on the segmented tumors, segmentation methods could ensure that the radiomic features are reproducible and reliable.

Intra-class correlation coefficients (ICC) are usually used to evaluate the inter- and intra-reader agreement. Several studies suggested that ROIs should be detected with an acceptable ICC, so that the further extracted features can be used [7, 8]. Several automatic or semi-automatic segmentation methods have been developed to minimize manual cost and increase reproducibility of tumor segmentation. Parmar et al. [55] compared radiomic features computed by a semiautomatic region growing volumetric segmentation algorithm and manual segmentation with NSCLC tumors in twenty patients; they found that the radiomic features generated with the semiautomatic segmentation had significantly higher reproducibility and more robustness compared with the manual segmentation. Heye et al. [56] compared the inter- and intra-observer variability with manual segmentation versus that with semiautomatic segmentation (“random walker”), preformed with MR Oncotreat software using 15 DCE MR studies. They concluded that the semiautomatic segmentation can provide a significant reduction in inter-observer variability compared with the manual method. These studies demonstrated the potential of semiautomatic segmentation in consistency and repeatability. In addition, Christ et al. [57] proposed a method for medical image segmentation integrating marker controlled watershed segmentation and clustering algorithm, which achieved better segmentation than the conservative watershed algorithm in MR images was achieved. Hatt et al. [58] summarized some PET automatic segmentation algorithms and
concluded that the segmentation with manual and visual verification show higher reproducibility and reliability than simple threshold-based approaches.

Until now, in morphological imaging (CT, US or MRI), most radiomic studies have used manually segmented ROIs for analysis even though it is quite time consuming and entails large inter-reader variability. In PET or SPECT imaging, several automatic or semiautomatic segmentation algorithms were performed. Although segmentation methods for medical images have been a hot spot of concern for a long time, there is still a long way to detect a fully automated application [59-61].

**Feature extraction**

Radiomics was initially defined as the extraction of high-throughput features from images. Quantitative imaging features are an important characteristic in radiomics because they bridge medical images and the clinical end point. Typically, there are two main categories in imaging features: manually defined features (including semantics and non-semantics) and deep learning features (Figure 4).

**Manually defined Features**

Semantic features could be defined as the qualitatively described empirical features proposed by radiologists. These features cannot be described with efficient mathematical expression, but they are useful in clinical settings and in radiomic analysis. For example, Liu et al. suggested that semantic features of CT images were correlated with epidermal-growth-factor-receptor (EGFR) status in patients with lung cancer [62].

---

**Figure 4.** The main features we used for radiomic analysis could be divided into three parts: Empirical features, Statistical features and Deep learning features. All these features could be visualized and interpreted with physical meanings. However, what we should do further is to unravel their physiological significance.
Non-semantic features could be defined as quantitatively described imaging features with mathematical expressions. For the current radiomic studies, the mostly used non-semantic features are usually obtained from predefined ROIs. First group contains the shape features describing the tumors, including volume, surface area, compactness, and etc. Second group contains a large number of first-, second-, and high-order features constructed from the intensity of each voxel in the ROIs [6]. First-order features describe the distribution of the intensity within the ROI, including maximum, median, minimum, and entropy. Second-order features describe the statistical relationships between voxels, such as textural features related with the representative of tumor heterogeneity. High-order features usually referred to statistical features calculated on matrices that consider relationships between three or more pixels. In addition, wavelet features, referred to as multiresolution image scaling features, and fractals, referred to as model-based features, were also used a lot in radiomic studies [63-65]. Other computational imaging features such as local binary patterns (LBP) [66] and scale-invariant feature transform (SIFT) [67] can also be used in radiomic analysis.

Deep Learning Features

Deep learning represents a class of algorithms that use the stacked neural network structure [68]. A deep learning model generally includes the following computational components: convolution, pooling, activation, full connection and batch normalization. All the computational components are defined as layers, and they are stacked layer-by-layer. The most important components of the deep-learning model are the convolutional layer and fully connected layer that can adaptively learn features from data. Therefore, the outputs of the convolutional layer or the fully connected layer can be defined as deep-learning features.

The typical manner of extracting deep-learning features is to train a deep learning model using the clinical outcomes, and then extract the outputs of several layers as deep learning features [69]. For example, after training a convolutional neural network (CNN) for EGFR mutation prediction in lung cancer, the results of the final fully connected layer was extracted to obtain deep learning features for EGFR status prediction [70]. Another method of extracting deep-learning features is to use an unsupervised auto-encoder. When there is not enough data to train a supervised deep learning model, an auto-encoder provides an alternative method to extract deep learning features [71-73]. A typical auto-encoder includes an encoder network and a decoder network. The result of the final convolutional layer is defined as a deep learning feature.

When compared with manually defined features, deep learning features are more specific to clinical outcomes and data. Most manually defined features describe general information about a tumor (e.g., shape, intensity, and texture), but may lack specificity regarding clinical outcomes. For example, when designing manually defined features, clinical outcomes are not considered in most situations, therefore, the manually defined features may not be suitable for clinical labels. In contrast, deep learning features are learned from data directly, and therefore, the deep learning features can be adapted to specific datasets and related to clinical outcomes more naturally. Generally, deep learning features are extracted from convolutional layers in the deep learning model. Each convolutional layer includes hundreds of convolutional filters that are defined as deep learning features. The deep learning features can describe multi-level tumor information from low-level visual characteristics to high-level abstract features. For example, in recurrence analysis of high-grade ovarian cancer, a convolutional auto encoder was designed to extract deep learning features from CT image. As illustrated in Fig. 3, features in the first convolutional layer describe simple tumor intensity information, and features in the second layer describe tumor edge information. Features in a deeper layer (the 3rd layer) describe complex tumor shape information. As the convolutional layer goes deeper, high-level abstract features are extracted, which are not interpretable visually, but tend to be associated with the recurrence status of the tumor. Without the need to explicitly specify feature formulas, deep learning features naturally include much of the same information as in manually defined features such as the intensity information (the 1st layer), edge information (the 2nd feature) and shape information (the 3rd). Since deep learning features are learned from data directly, they have a relationship with the recurrence status of the tumor (Fig. 3b). In addition, deep learning features can be designed flexibly since different CNN structures provide different deep learning features, and there is no need carefully consider design feature formulas.

Knowledge discovery and modeling

Generally, thousands of quantitative imaging features could be obtained for the tumors. However, too many features may include redundant information and cause overfitting. Thus, feature selection should be performed to preserve the
relevant features. A prediction model can then be constructed with the selected key features according to the clinical endpoint.

Feature Selection

The most commonly used feature selection methods in radiomic studies can be divided into three broad categories: filter, wrapper, and embedding [74].

Filter methods usually evaluate the features without involving the model; therefore, these are also called independent methods. There are two main kinds of filter methods: univariate and multivariate methods. Univariate filters mostly rank features based on quality (i.e., on the Chi-squared test or Mann-Whitney U test). Multivariate filters are composed of rankers and subset selectors like those used in correlation-based feature selection. With wrapper methods, features are generated first and then evaluated using the model. This means all wrapper methods are considered multivariate and subset-selector methods. With embedded methods, a feature subset is proposed and evaluated during the construction of the model.

Most radiomic studies use filters for coarse selection. For example, several studies used ICCs to detect stable features [4, 73, 75], and others also ranked the features using Student’s t-test [8] or the Mann-Whitney U test [76]. Another commonly used feature-selection method in radiomic studies is the least absolute shrinkage and selection operator (LASSO), which is an embedded method that can produce selected features and a prediction model together [77-81].

Modelling

The goal of radiomics is to construct a prediction model for clinical outcomes with selected features. Machine learning provides several modelling methods to achieve this goal. Typically, supervised, semi-supervised, and unsupervised learning are fundamental strategies used according to the different levels of available clinical outcomes.

Supervised learning is a learning structure that requires clinical labels to train the model. A supervised learning model includes two steps: training and testing. In the process of model training, training samples with corresponding clinical labels are paired to train the model. Through a pre-defined loss function, the model learns the relationship between the feature and the clinical labels. In the testing phase, the well-trained model is used directly to test its predictive performance. There are many supervised learning methods that have shown great performance in radiomic analysis, such as the support vector machine [8], LASSO-logistic regression [7] and random forest [73]. Despite the good performance of many supervised-learning methods, they usually require large amounts of training samples to avoid overfitting.

In many situations, clinical labels may not be enough to train a supervised model, and unsupervised learning can be an alternative. Unsupervised learning methods group data into several clusters according to the similarity between samples. In this process, the clinical label is not necessary for model training. The unsupervised model utilizes a distance measurement to calculate the similarity between samples. The similar training samples are grouped together, while the dissimilar samples are assigned into different groups. These models include many clustering algorithms such as k-means clustering [82], fuzzy clustering [83] and consensus clustering [84]. For example, a consensus clustering model was used in a previous study to detect glioblastoma phenotypic subtypes [85], and patients were classified into three groups, each with different molecular pathway activities and a different prognosis.

Generally, supervised learning shows stronger performance than unsupervised learning since the clinical label is considered during the supervised model training. However, insufficiently labelled data limits the performance of supervised learning. As a trade-off between better performance and enough training data, semi-supervised learning could be a good choice. Semi-supervised learning uses a large amount of unlabeled data to mine tumor information and utilizes small amounts of labelled data to build the relationship between features and clinical labels. For predicting the recurrence for high-grade serous ovarian cancer, a semi-supervised deep learning framework was proposed [86]. The semi-supervised framework can be divided into an unsupervised feature learning phase and a supervised model training phase. In the unsupervised learning phase, a convolutional auto encoder was built to compress tumor images into low-dimensional features. In this process, the auto encoder tried to extract the intrinsic characteristics of the tumor without a clinical label. Afterwards, a supervised Cox model was built to model the relationships between the features and clinical labels.

Validation

A radiomic model must be validated to show its potential value for clinical application. An independent, external, validated model is considered to be more credible than an internally validated model, as results from independently obtained data are usually more robust. Furthermore, prospectively validated models have the most credibility for
radiomic studies. There are dozens of tools to measure the performance of radiomic models. For discrimination analysis, the receiver operating characteristic (ROC) curve is the mostly used method to report the performance of proposed models. Moreover, the area under the curve (AUC), sensitivity, and specificity of the model can also be used to evaluate if the model can predictive the clinical outcome. For survival analysis, the concordance index (C-index) and time-dependent ROC curve are usually used for validation. In addition, calibration is a useful tool for both discrimination analysis and survival analysis, as it can obtain the agreement between the observed clinical outcomes and model predictions [87].

Applications of Radiomics in Diagnosis and Treatment of Oncology

Since radiomics was proposed, rapidly increasing radiomic studies have been published to improve the diagnosis and treatment of cancer. More and more studies demonstrate the value of radiomic features as a complimentary tool to decision making of clinicians. Here, we will provide a brief introduction to radiomic studies, focusing on the clinical problems during the whole process of diagnosis and treatment in several typical cancers including brain tumors, head-and-neck cancer, breast cancer, lung cancer, prostate cancer, colorectal cancer, liver cancer and gastric cancer (Figure 5).

Brain Tumors

Malignant brain tumors can be divided into primary tumors which started within the brain (mainly referred to as gliomas, meningiomas, and chordomas), and secondary tumors that spread from elsewhere, known as brain-metastasis tumors [88]. Current brain-tumor imaging protocols incorporate CT, multi-parametric MRI, and sometimes PET. As use of radiomics could extract large amounts of quantitative imaging features and capture intratumoral and intertumoral heterogeneity, it makes a radiomic analysis could assess imaging phenotypes that may influence the diagnosis and treatment evaluation of brain tumors.

Diagnosis

Although standard brain imaging can provide results enough for tumor grading, improve was needed for newly proposed imaging protocols. Bai et al. first utilized diffusion kurtosis MRI for the grading of gliomas and detected good performance in the grading of gliomas [89]. Moreover, radiomics also attracted attention for the prediction of molecular subtypes of brain tumors, as the precision diagnosis of gene-expression patterns could potentially enhance decision making for targeted therapies [90]. Isocitrate dehydrogenase (IDH) [91], O6-methylguanine-DNA methyltransferase (MGMT) [92, 93], 1p/19q co-deletion [94], EGFR expression level [95], Ki-67 expression level [96], p53 status [97], and ATRX mutation [98] have been the main focus in prediction studies of molecular subtype in brain tumors. Radiomic analysis based on multi-parametric MRI aided the prediction of molecular characteristics. More importantly, imaging phenotypes were shown to be associated with molecular pathway activities that may determine the type of targeted therapy [85].

Treatment evaluation and Prognosis

There has been considerable interest in treatment evaluation and prognosis for brain tumors to identify imaging phenotypes that may predict the treatment response in patients with glioblastoma. Two recent studies investigating the responses to bevacizumab treatment in recurrent glioblastoma patients suggested the potential of radiomics to predict different response to the treatment [99, 100]. These studies both detected the potential of radiomics to aid in cancer treatment decision-making at a low cost. However, a bevacizumab-naïve control group was needed for these studies to confirm the predictive value. In addition, radiomics provided a new option for determining prognosis for brain tumors. Recent studies have suggested that features detected from MRI and PET were significantly associated with the
survival of patients with gliomas [101, 102]. In the future, a model could be constructed using radiomics to improve both treatment planning and prognosis.

**Head- and-neck cancer**

Head-and-neck cancer is one of the cancers that radiomics has also been widely applied [103].

**Diagnosis**

Ren et al. built a MRI based radiomic signature to predict the stage of head and neck cancer preoperatively [104]. They found that the radiomic signature from contrast-enhanced T1-weighted MR images and T2-weighted MR images had good performance in discriminating different stages. Leijenaar et al found that CT radiomic signature could predict HPV (p16) status in oropharyngeal squamous cell carcinoma [105]. Zhou et al proposed a 3-dimensional deep learning model to predict lymph node metastasis in nasopharyngeal cancer (NPC) [106]. Chen et al. evaluated the association between Tumor PD-1 expression and Immunohistochemical biomarkers or radiomic features from PET imaging in NPC [107]. Crispin-Ortuzar et al predicted hypoxia status using PET/CT radiomics in NPC patients [108].

**Treatment evaluation and Prognosis**

Zhang et al. proposed that multi-parametric MRI based radiomics could be a novel prognostic factor in advanced NPC [109]. They collected 118 advanced NPC patients and found that radiomic signature achieved significantly improved performance evaluating PFS compared with the TNM staging system. Furthermore, Zhang et al. evaluated 6 different feature selection methods and 9 different classifiers for prediction of local failure and distant failure in advanced NPC [110]. They found that Random Forest performed best among the 9 methods. Wang et al. [111] investigated the value of radiomic signatures in prediction of early response to induction chemotherapy in NPC patients. They found that radiomic signature had a good performance in predicting early response to induction chemotherapy. Lu and Lv et al. evaluated the robustness of radiomic features obtained from different PET images in NPC patients [112, 113]. Wu et al. performed survival prediction in high-grade Osteosarcoma using radiomics of diagnostic CT [114]. They found that radiomic nomogram, which combines a radiomic signature and clinical factors, showed better calibration and classification capacity than a model with only clinical factors. Elhalawani et al investigated radiomic signatures for local recurrence in oropharyngeal cancer [115]. A radiomic signature consisting of 2 radiomic features showed robust discrimination ability of recurrence. Vallières et al used radiomics for assessing tumor failure in patients with NPC from 4 centers [116]. Gabryś et al predicted the xerostomia in NPC patients after radiotherapy [117]. They compared the performance of 6 feature selection, 7 classification, and 10 data balancing methods.

**Breast cancer**

Breast cancer is the cancer with the highest incidence in women worldwide. As it is a known heterogeneous disease, the precise diagnosis and early prediction of response to the treatment are hotspots of the clinical practice and research in breast cancer. Radiomics, combining multi-modality imaging data and clinical information, is now used widely in breast cancer research.

**Diagnosis**

There are several diagnostic imaging modalities for breast cancer, including US, PET/CT, mammography, and MRI. Thus, radiomic researches in breast cancer also covered all these modalities. Several previous studies utilized radiomics for the prediction of breast cancer subtype or the status of ER, PR, Ki67 and HER2 with US [46], mammography [118], PET/CT [119, 120], and MRI [121, 122]. Specifically, Antropova and his colleague proposed a feature fusion method that could combine US, MRI and mammography together for better diagnosis of breast cancer [118]. In addition of the characterization of breast cancer, radiomics could also provide a non-invasive approach for prediction of sentinel lymph node metastasis [123].

**Treatment evaluation and Prognosis**

Most radiomic studies in breast cancer have focused on the evaluation of response to therapy. Chan et al. [124] developed an automated method to predict treatment failure in early breast cancer patients with a pretreatment MRI scan. Most other studies attempted to obtain a radiomic biomarker for pathological complete response (pCR) to neoadjuvant chemotherapy, which is a hot point in breast cancer research. Braman et al found that intra- and peri-tumoral features of DCE-MRI could contribute to pretreatment prediction of pCR [125]. Other studies also showed that T1WI, T2WI and DWI could also help to detect pCR [76, 126]. While several studies approved the potential of MRI as a biomarker for detecting pCR, Tran et al pointed out that diffuse optical spectroscopic could also be a useful tool for pCR prediction [127].

There are also radiomic studies focusing on the prognosis of breast cancer. Park et al [128] have developed a radiomic signature combining MRI features and clinical information for individualized estimation of disease free survival in breast cancer.
patients.

**Lung cancer**

Lung cancer is the most harmful cancer, and its prevalence continues to increase worldwide [129]. Radiomics has been widely used in the diagnosis, treatment evaluation, and prognosis in lung cancer [130, 131].

**Diagnosis**

One of the most important diagnostic applications of radiomics is screening of lung cancer. Kumar et al performed a classification between malignant and benign lesions with diagnostic data from the LIDC-IDRI dataset [132]. The proposed method achieved a sensitivity of 79.06% and a specificity of 76.11%. Shen et al. proposed a deep learning model based on CT images and achieved better prediction results for malignant lung nodule compared with previous methods [133]. Maldonado et al. developed a noninvasive radiomic method to discriminate solid and ground glass opacities using high resolution CT [134]. Besides diagnostic CT, researches on low-dose CT were also performed for screening malignant lung nodules. Carter et al. [135] performed a screening study on patients with low-dose CT screening detected lung cancer in National Lung Cancer Screening Trial (NLST) dataset. Accuracies of 80% and 79% were found for predicting nodules that will develop cancers in one or two years, respectively. Liu et al. also utilized a low-dose CT cohort from the United States and radiological image traits of lung nodule to predict malignancy and achieved impressing results [136].

Accurate preoperative TNM staging is also important for treatment decisions of lung cancer. Aerts et al. extracted 440 CT radiomic features from 1,019 patients and found that many of the radiomic features were correlated with the overall stage, T stage, N stage, and M stage of lung cancer and head-and-neck cancer [4, 137]. Coroller et al. used CT radiomic signature to predict distant metastasis (M staging) on a United States cohort of 182 pathologically confirmed lung adenocarcinoma [138] and Zhou et al. combined CT radiomic features and clinical risk factors to identify the distant metastasis based on a Chinese cohort including 348 lung cancer patients [139]. The results from all of these studies showed good performance of radiomics on preoperative M staging. Wu et al. used PET/CT radiomic features to predict distant metastasis with 101 early-stage NSCLC patients and they found additional value of PET in M staging [140].

Besides screening and staging, radiomics has been applied to predict the gene mutation or pathological type in lung cancer. Liu et al. and Zhang et al. all found that EGFR mutation was associated with CT radiomic features [141-143]. Rios et al. developed CT radiomic signatures to distinguish somatic mutations on a dataset of 763 lung adenocarcinoma patients [73]. They found that a CT radiomic signature that is related with radiographic heterogeneity can successfully predict EGFR status. Wu et al. used radiomics to predict the histologic type of lung cancer, adenocarcinoma or squamous carcinoma, on two NSCLC cohorts from Netherlands [144]. Zhu et al. performed a similar study on a Chinese cohort and also found that radiomic signature could serve as a diagnostic factor for histologic subtype classification of NSCLC [145]. Fan et al. [146] developed a radiomic signature for the preoperative discrimination of lung invasive adenocarcinoma manifesting as a ground-glass nodule.

**Treatment evaluation and Prognosis**

Treatment response evaluation is essential in lung cancer due to its value in treatment decision. Mattonen et al. found that radiomic signatures could predict the recurrence after Stereotactic Ablative Radiation Therapy (SART) in Lung Cancer [147, 148]. Fave et al. used delta-radiomic features to predict outcomes in Stage III NSCLC patients during the radiation therapy [149]. Their results suggest the change of radiomic features due to radiation therapy would be indicators of tumor response. Coroller et al. found that pretreatment CT radiomic features could predict pathological response after neoadjuvant chemoradiation in patients with advanced NSCLC [150]. Aerts et al. found that CT radiomic feature before treatment is able to predict EGFR mutation status in NSCLC and associated with gefitinib response [151]. Cook et al. used heterogeneity features in PET images to evaluate the treatment response of Erlotinib in NSCLC [152]. Song et al. found that CT radiomic signature could predict progression-free survival after TKI therapy [153].

Besides treatment response, many radiomic studies have focused on prognosis. Aerts et al. found the clusters of radiomic features were associated with the prognosis of lung cancer [4]. Song et al. also demonstrated a relationship between CT radiomic features and overall survival in patients with NSCLC [154]. Balagurunathan et al. tested the reproducibility and prognosis of quantitative radiomic features from CT images, and many features were associated with the prognosis of lung cancer [75]. Coroller et al. found that radiomic features are prognostic for both distant metastasis and survival [138]. Huang et al. found that the radiomic signature was associated with DFS [155].
Colorectal cancer

Colorectal cancer is a common and fatal type of cancer. The incidence and deaths of colorectal cancer are increasing every year, and many questions remain regarding the three different stages of diagnosis, treatment evaluation, and prognosis.

Diagnosis

Accurate identification of the extent of lymph-node metastasis in colorectal cancer patients is critical for the determination of treatment strategies. If it is confirmed that the patient has no lymph-node metastasis before surgery, there is no need to perform lymph-node dissection during surgery [156-158]. Huang et al. performed radiomic analysis on 526 patients with colorectal cancer [7]. Based on preoperative CT images of the portal vein, the radiomic features were extracted, and a nomogram was established. The C-index was 0.736 on the training set and 0.778 on the validation set. Moreover, the authors proposed an easy-to-use nomogram based on the radiomic model to help its application (Figure 6). Another relevant application of radiomics in colorectal cancer is the prediction of KRAS/NRAS/BRAF mutation. The National Comprehensive Cancer Network (NCCN) guidelines for colorectal cancer recommend that patients should screen for KRAS/NRAS/BRAF mutations, because these mutations are indicative of poor response to cetuximab and panitumumab [159-162]. Therefore, determination of the KRAS/NRAS/BRAF mutation status before or during treatment is essential to predict treatment efficacy and achieve a personalized diagnosis. Yang et al.[163] extracted 346 radiomic features from the preoperative CT images of 117 patients with colorectal cancer for a radiomic analysis, achieving an AUC of 0.869.

Treatment evaluation and Prognosis

Patients with rectal cancer usually undergo neoadjuvant chemoradiotherapy before surgery [164, 165], but it is not known whether pCR is achieved before neoadjuvant chemoradiotherapy. This information is only obtained after surgery. Approximately 15-27% of locally advanced patients will achieve pCR [166, 167], and studies have shown that these patients do not need surgery [168-170]. Based on information from 48 patients, Ke et al. [171] constructed an artificial neural network based model, and the AUC was 0.71-0.79. Liu et al. applied radiomics to identify patients with locally advanced rectal cancer who achieved PCR before surgery [8]. The study enrolled 222 patients with locally advanced rectal cancer. T2WI and DWI were used to extract 2252 radiomic features. The AUC was 0.976 (95% confidence interval [CI], 0.9185-0.9711), and the predicted PCR accuracy was 94.3% (95% CI, 91.9%-97.1%). Based on 114 patients (18 PCR) Natally et al. extracted radiomic features on T2WI and made predictions using random forest classification; the AUC was 0.93, sensitivity 100%, and specificity 91% [172].

Although treatment may improve survival, it has not proven to improve the DFS of patients with locally advanced rectal cancer [13, 173, 174]. Distant metastasis is the main cause of patient treatment failure. For high-risk patients, increased systemic therapy reduces the risk of metastasis and improves survival. Therefore, pre-operative risk stratification may help to select individualized treatment strategies and improve the prognosis of patients. Based on 108 patients with locally advanced rectal cancer who underwent total mesorectal excision [175], Meng et al. extracted 485 radiomic features; the C-index of the radiomic signature was 0.767 (95% CI, 0.72-0.86), and the 3-year time-dependent AUC was 0.827. Combined
with clinical features, the C-index and AUC of the Cox model were improved: 0.788 (95% CI, 0.72-0.86) and 0.837, respectively.

Identification of patients who need more aggressive treatment and follow-up is critical. Lovinfosse et al. explored the significance of PET/CT-based radiometric features for prognosis. The study extracted the SUV\text{max}, SUV\text{mean}, metabolic tumor volume, total lesion glycolysis, and texture features, established univariate and multivariate Cox models, and analyzed the prognostic performance of the features [176].

**Prostate cancer**

Prostate cancer (PCa) is the most common cancer affecting men in Western societies [177]. Multi-parametric MRI (mp-MRI) plays an important role in diagnosis and treatment of PCa because of the advantage of high soft tissue contrast.

**Diagnosis**

The non-PCa condition such as prostatitis and benign prostatic hyperplasia is easily confused with PCa. Ginsburg SB et al. utilized mp-MRI based radiomic model for differentiation between PCa and non-PCa [178]. The result showed that the validation model achieved relatively good performance with high AUC using T2 WI, ADC, and T2 WI and ADC features for PCa detection. PCa exhibits an extraordinary variable biological behavior. According to NCCN Prostate Cancer, version 3, the identification of aggressive PCa is vital for clinical decision. Some radiomics models based on mp-MRI were developed and validated to identify PCa aggressiveness in patients [179, 180]. The results showed that radiomic models could be helpful for identifying the presence and absence of PCa aggressiveness compared with PIRADS v2.0 assessment, an mp-MRI based subjective assessment method. The pathological Gleason score (GS) of PCa has clinical implications for the treatment and prognosis. Chaddad et al. achieved the best performance for GS = 3 + 4, and 64.76% for GS ≥ 4 + 3 to predict the GS of PCa patients via radiomic features based on mp-MRI [181].

**Treatment evaluation and Prognosis**

In addition, radiomics is also used for treatment evaluation of PCa. Many researchers successfully utilized radiomic features derived from pretreatment mp-MRI to predict biochemical recurrence of PCa after radical prostatectomy, radical radiotherapy or external beam radiotherapy [182, 183].

**Liver and Gastric cancer**

As medical imaging is playing a more and more important role in the management of liver cancer and gastric cancer, there are more opportunities to utilize quantitative imaging analysis in the clinical setting for diagnosis and treatment evaluation.

**Diagnosis**

Radiomics is mainly used in cases of liver cancer for noninvasive prediction of microvascular invasion (MiVI). Two research teams have verified the validity of radiomics for preoperative MiVI prediction [184, 185]. Bakr et al. extracted quantitative radiomic and semantic imaging features from the triphasic CT scans of 28 treatment-naive hepatocellular carcinoma (HCC) patients [184]. Delta radiomic features were designed as the absolute difference and ratio computed from all pairs of the three imaging phases. The proposed imaging features from arterial and venous phases were regarded as potential biomarkers for MiVI prediction in HCC. Jie et al. further used a cohort of 304 patients with hepatitis B virus-related (HBV) HCCs to validate the usefulness of imaging biomarkers for MiVI prediction [185]. The radiomics-based nomogram showed satisfactory predictive accuracy for MiVI in HBV-related HCC.

For gastric cancer (GC), Ma et al. evaluated the value of CT-based radiomic signature to discriminate Borrmann type IV GC from primary gastric lymphoma (PGL) [186]. Ba-Ssalamah et al. used CT texture features to discriminate GC, lymphoma, and gastrointestinal stromal tumors [187]. Liu et al. explored the application of radiomics in predicting pathologic differentiation for the type of GC [188]. Liu et al. used radiomic analysis to predict T stage, N stage, overall stage, and perineural invasion in patients with GC [189-191].

**Treatment evaluation and Prognosis**

Radiomics-based methods for treatment evaluation and prognostic prediction in liver cancer rely on two aspects: overall survival (OS) and recurrence predictions. Zhou et al. proved the prognostic power of radiomic signatures for early recurrence of HCC with partial liver resection [192]. The LASSO was used to build the radiomic signature. The AUC of the combined model was 0.836, which shows it performed better than clinical predictors. Akai et al. investigated the use of random forest along with radiomic analysis to predict DFS and OS for HCCs with resection as treatment [193]. Furthermore, Cozzi et al. proved the radiomic signature could be used as a robust biomarker for local control and survival prediction [194]. Compacity and Barcelona Clinic Liver Cancer (BCLC) stage were finally chosen as effective covariates for prognosis prediction. All these studies indicated that radiomic signatures are potential biomarkers for prognostic prediction in liver
cancer research.

Radiomics is also a useful tool for treatment evaluation and prognosis in GC. Giganti et al. found that radiomic texture features had predictive ability for treatment response for neoadjuvant therapy in patients with GC [195]. Hyun et al. found that textural features on CT images were associated with better prognosis in patients with HER2-positive gastric cancer who received target therapy [196]. Giganti et al. investigated the association between CT based radiomic features and overall survival in patients with GC [197]. Texture features were significantly associated with the prognosis of GC.

Challenges and Prospects for Radiomics

In recent years, we have witnessed the progress of radiomics in methodologies and clinical applications. Its potential has been revealed in helping clinical experts to uncover cancer characteristics that fail to be appreciated by naked eyes. In particular, radiomics not only helps radiologists to make precision diagnoses, but also provides oncologists a useful tool for treatment planning and response evaluation [131, 198]. Based on the progress outlined here, further improvements in radiomics will advance clinical care for cancer patients. Next, we will discuss the challenges and opportunities for further radiomics studies.

**Big clinical data and data sharing**

A reliable conclusion is based on enough data, as is radiomics. Big and standardized clinical data will make radiomics clinically applicable. However, the radiomic researchers need better access to big data, as medical images are dispersed in different hospitals or data centers, and the size of the patient population is the basis of large amounts of medical images. Another challenge is that the collection of medical images is time consuming. Thus, data sharing among institutes and hospitals 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 information [199], and it is still growing with contributions from different institutes and hospitals. Researchers could develop new methods and tools to validate their hypotheses with this large dataset [59, 79]. The Quantitative Imaging Network (QIN) is another useful project for data integration that may promote radiomic studies [200]. However, high quality datasets are still needed for radiomic researches [201].

**Reproducibility and quality control**

Reproducibility and quality control are involved in the entire radiomic process, including data acquisition, feature extraction and selection, modeling, and validation. Every decision in this process affects the result, its reproducibility, and the quality of the research. Moreover, the inclusion and exclusion criteria of patients in radiomic studies are other important factors related that affect reproducibility. Different criteria may result in the availability of different cohorts for analysis, especially for small cohort studies. Therefore, developing unique inclusion and exclusion criteria for similar clinical applications of radiomics should be considered seriously. Disclosure of the details in every step is important; with the development of data sharing and unique criteria of patients’ inclusion, models may be validated using an open dataset that could become the standard for comparison of future studies. Recently, the image biomarker standardization initiative (IBSI) towards seeking standardization for improved reproducibility of high-throughput imaging analyses mainly focused on radiomics was proposed [69]. This was a highly

Figure 7. Flowchart depicting the workflow of radiomics and the application of the RQS. Reproduced with permission from [10]. Abbreviations: RQS, radiomics quality score.
valuable step towards the improvement of radiomic studies. In addition, the radiomics quality score (RQS) (Figure 7) was proposed to aid in the assessment of radiomics studies [10]. Improving research quality will increase the influence of radiomics.

**Interpretability of radiomic features and models**

Clinical experts often think of the radiomic model as a black box that can provide good prediction results for particular clinical outcomes [131, 198], which tends to make radiomics a less accepted approach. Improvement of the interpretability of radiomic features and models is urgently needed. For manually defined imaging features, it is easier to correlate them using pathophysiology; for deep learning, it is a little more difficult. Visualization of deep learning features and prediction models could potentially help to solve this problem. Fortunately, there are many researchers investigating methods to accomplish this goal and several tools have been proposed to diminish the black box perception [202].

**Conclusions**

Radiomics is still a newly proposed and rapidly progressing field that integrates radiology, oncology, and machine learning. With the growth of clinical data and advanced machine-learning methodologies, it is playing an increasingly important role in precision diagnostics and oncology. Further research should focus on the reproducibility and interpretability to further make radiomics more acceptable in the field.

**Table. Specifications of radiomic studies in different cancers**

| Studies | Study design | No. of patients (training + testing) | No. and type of radiomic features | Statistical analysis | Image Modality | Clinical Characteristics |
|---------|--------------|------------------------------------|----------------------------------|----------------------|-----------------|--------------------------|
| **Brain tumor** | | | | | | |
| Kickingeder et al [86] | Retrospective, single center study | 79 + 40 | 6095 (first-order, volume and shape, texture) | Cox regression analysis | MRI | Prognosis |
| Xi et al [90] | Retrospective, single center study | 98 + 20 | 1665 (first-order, size and shape, texture, wavelet) | LASSO, SVM | MRI | Diagnosis |
| Kickingeder et al [91] | Retrospective, single center study | 121 + 60 | 1043 (first-order, size and shape, texture) | Chi-square and Wilcoxon test, LASSO, Cox regression | MRI | Prognosis |
| Han et al [92] | Retrospective, single center study | 184 + 93 | 79 (size and shape, intensity, textural, wavelet) | random forest, U test | MRI | Prognosis |
| Li et al [93] | Retrospective, single center study | 200 + 70 | 431 (first-order, size and shape, texture, wavelet) | logistic regression, t test, Chi-square test | MRI | Diagnosis |
| Li et al [95] | Retrospective, single center study | 180 + 92 | 431 (first-order, size and shape, texture, wavelet) | t test, Chi-square test, LASSO, SVM | MRI | Diagnosis |
| Kickingeder et al [97] | Retrospective, single center study | 112 + 60 | 4842 (first-order, size and shape, texture, wavelet) | PCA, Cox regression model | MRI | Treatment evaluation |
| Grossmann et al [98] | Retrospective, single center study | 57 + 56 | 65 (first order statistics, size and shape, texture) | PCA, Spearman rank correlation, Wilcoxon test, Cox | MRI | Treatment evaluation and Prognosis |
| Papp et al [99] | Retrospective, single center study | 70 | 48 (Histogram, shape, texture) | Geometrical Probability Covering Algorithms | PET | Prognosis |
| **Head and neck cancer** | | | | | | |
| Ren et al [102] | Retrospective, single center study | 85 + 42 | 970 (shape, intensity, textural and wavelet) | LASSO | MRI | Diagnosis |
| Liemenaar et al [103] | Retrospective, multi-center study | 628 + 150 | 902 (intensity, shape, texture, Wavelet and Laplacian) | LASSO and multivariable logistic regression | CT | Diagnosis |
| Zhou et al [104] | Retrospective, single center study | 122 + 39 | 257 (intensity, texture, and geometric) | 3D CNN | PET/CT | Diagnosis |
| Shen et al [105] | Retrospective, single center study | 53 | 41 (textural features or histograms) | logistic regression analysis | PET/CT | Diagnosis |
| Zhang et al [107] | Retrospective, single center study | 88 + 30 | 970 (shape, intensity, textural and wavelet) | LASSO | MRI | Prognosis |
| Wang et al [109] | Retrospective, single center study | 120 | 591 (texture) | LASSO | MRI | Treatment evaluation |
| Wu et al [111] | Retrospective, single center study | 102 + 48 | 474 (shape, intensity, textural and wavelet) | LASSO | CT | Prognosis |
| Elhalawani et al [113] | Retrospective, single center study | 420 + 45 | 134 (intensity direct, histogram, shape and texture) | multivariate Cox | CT | Prognosis |
| Vallières et al [114] | Retrospective, multi-center study | 194 + 106 | 1615 (intensity, shape, texture) | random forest | PET/CT | Prognosis |
| **Breast cancer** | | | | | | |
| Guo et al [46] | Retrospective, single center study | 215 | 463 (morphology, intensity, texture, wavelet) | t test, ICC, LASSO, and SVM | US | Diagnosis |
| Antropova et al [116] | Retrospective, single center study | 2060 | Deep learning features | CNN, SVM | FFDM, US, MRI | Diagnosis |
| Antunovic et al | Retrospective, single | 43 | 20 (size and shape, first-order) | Univariate analysis, | PET/CT | Diagnosis |
| Studies     | Study design               | No. of patients (training + testing) | No. and type of radiomic features                        | Statistical analysis                                                                 | Image Modality | Clinical Characteristics |
|------------|---------------------------|--------------------------------------|----------------------------------------------------------|--------------------------------------------------------------------------------------|----------------|--------------------------|
| Ha et al [118] | Retrospective, single center study | 73                                  | 109 (texture)                                             | hierarchical clustering, exact Fisher’s test                                        | PET            | Diagnosis and Prognosis  |
| Saha et al [120] | Retrospective, single center study | 461 + 461                          | 529 (first order, size and shape, texture)                | Hierarchical clustering, logistic regression, Cox correlation, RF, logistic regression | MRI            | Diagnosis                |
| Chan et al [122] | Retrospective, single center study | 563                                  | 8192 (washin and washout intensity values)                | PCA, LASSO, logistic regression, K-M survival evaluation                            | MRI            | Treatment Evaluation    |
| Braman et al [123] | Retrospective, single center study | 117                                  | 99 (texture)                                              | mRMR, consensus clustering, LDA, DDLA, QDA, naive Bayes, SVM                        | MRI            | Treatment Evaluation    |
| Chamming et al [76] | Retrospective, single center study | 85                                  | texture                                                   | Mann-Whitney U test, logistic regression                                            | MRI            | Treatment Evaluation    |
| Partridge et al [124] | Prospective, multi-center study | 272                                | ΔADC, ΔFTV                                                | Z-test, logistic regression                                                         | MRI            | Treatment Evaluation    |
| Tran et al [125] | Retrospective, single center study | 37                                  | 40 (texture)                                              | Logistic regression, naive Bayes, k-NN                                               | DOG            | Treatment Evaluation    |
| Park et al [126] | Retrospective, single center study | 194*100                             | 156 (texture)                                             | Elastic net, Cox Model, K-M analysis                                                | MRI            | Prognosis                |
| Lung cancer |                           |                                      |                                                          |                                                                                     |                |                          |
| Kumar et al [130] | Retrospective, multi-center study | 38106 + 4234                        | 500 (Deep learning features)                             | 3D CNN                                                                               | CT             | Diagnosis                |
| Shen et al [131] | Retrospective, multi-center study | 825 + 275                          | Deep learning features + 319 (first order, shape and size, textural, and wavelet) | mrmr, SVM, MC-CNN                                                                    | CT             | Diagnosis                |
| Aerts et al [4] | Retrospective, multi-center study | 474 + 575                          | 440 (intensity, shape, texture and wavelet)              | Cluster and Cox                                                                      | CT             | Prognosis and Diagnosis |
| Parmar et al [135] | Retrospective, multi-center study | 558 + 320                          | 440 (intensity, shape, texture and wavelet)              | Cluster and Cox                                                                      | CT             | Prognosis                |
| Coroller et al [136] | Retrospective, single center study | 98 + 84                            | 635 (intensity, shape, texture, LoG and Wavelet)         | mrmr and Cox                                                                         | CT             | Prognosis                |
| Zhou et al [137] | Retrospective, multi-center study | 241 + 107                          | 485 (intensity, shape, texture, and wavelet)            | concave minimization and SVM                                                         | CT             | Prognosis                |
| Jia et al [138] | Retrospective, single center study | 70 + 31                            | 70 (statistical, histogram, morphologic, and texture)   | c-means, ICC, linear correlation coefficient, LASSO and Cox                         | PET            | Prognosis                |
| Liu et al [139] | Retrospective, single center study | 298                                  | 219 (size, shape, location, air space, histogram, laws texture and wavelet) | Multiple logistic regression analysis, and SVM                                      | CT             | Diagnosis                |
| Zhang et al [141] | Retrospective, single center study | 140 + 40                            | 485 (shape, intensity, textural and wavelet)            | LASSO                                                                                | CT             | Diagnosis                |
| Rios et al [56] | Retrospective, multi-center study | 353 + 352                          | 440 (intensity, shape and texture)                       | MRRMR, Random forest classifier                                                     | CT             | Diagnosis                |
| Wu et al [142] | Retrospective, multi-center study | 198 + 152                          | 440 (intensity, shape and texture)                       | 24 Feature Selection Methods, 3 Classifiers                                          | CT             | Diagnosis                |
| Zhu et al [143] | Retrospective, single center study | 81 + 48                            | 485 (shape, intensity, textural and wavelet)            | LASSO                                                                                | CT             | Diagnosis                |
| Fan et al [144] | Retrospective, wavelet        | 160 + 235                          | 355 (shape, intensity, textural and wavelet)            | LASSO                                                                                | CT             | Diagnosis                |
| Fave et al [147] | Retrospective, single center study | 107                                  | 212 (texture, shape)                                    | Cox, AIC                                                                            | CT             | Prognosis                |
| Coroller et al [148] | Retrospective, single center study | 101 + 26                           | 440 (intensity, shape and texture)                       | PCA                                                                                 | CT             | Treatment evaluation    |
| Aerts et al [149] | Retrospective, multi-center study | 47                                  | 183 (shape, intensity, textural and wavelet)            | Coefficient of Variation                                                             | CT             | Treatment evaluation    |
| Song et al [151] | Retrospective, single center study | 117 + 197                          | 1032 (shape, intensity, textural and wavelet)          | LASSO and COX                                                                        | CT             | Treatment evaluation    |
| Song et al [152] | Retrospective, single center study | 661 + 61                           | 592 (shape, intensity, textural and wavelet)           | SVM and COX                                                                          | CT             | Prognosis                |
| Balagurunathan et al [58] | Retrospective, single center study | 32 + 59                            | 329 (shape, size, and texture)                          | concordance correlation coefficient, and K-M                                        | CT             | Prognosis                |
| Huang et al [155] | Retrospective, single center study | 141 + 141                          | 132 (textural)                                          | LASSO and COX                                                                        | CT             | Prognosis                |
| Colorectal cancer |                           |                                      |                                                          |                                                                                     |                |                          |
| Huang et al [7] | Retrospective, single center study | 326 + 200                          | 150 (histogram and GLCM)                                | LASSO                                                                                | CT             | Diagnosis                |
| Yang et al [161] | Retrospective, single center study | 61 + 56                            | 346 (histogram, shape, and texture)                      | ICC, SVM                                                                             | CT             | Diagnosis                |
| Ke et al [169] | Retrospective, single center study | 48                                  | 103 (histogram, GLCM, shape)                             | Mann-Whitney test, T test, ANN, T test, LASSO                                         | MRI            | Treatment evaluation    |
| Liu et al [8] | Retrospective, single center study | 152 + 70                            | 2252 (intensity, shape, wavelet features)               |                                                                                     | MRI            | Treatment evaluation    |

Modality: CT, MRI, PET, PET-MRI, LDA, DDLA, QDA.

Statistical methods: LASSO, mRMR, consensus clustering, LDA, DDLA, QDA, naive Bayes, SVM, PCA, LASSO, logistic regression, Cox correlation, RF, logistic regression, mRMR, consensus clustering, LDA, DDLA, QDA, naive Bayes, SVM, 3D CNN, Logistic regression, naive Bayes, k-NN.

Clinical characteristics: Diagnosis, Prognosis, Treatment evaluation.
### Abbreviations

ADC: apparent diffusion coefficient; AUC: area under receiver operating characteristic curve; C-index: concordance index; CNN: convolutional neural network; CT: Computed Tomography; DFS: Disease-free survival; DWI: diffusion weighted imaging; EGFR: epidermal growth factor receptor; HCC: hepatocellular carcinoma; ICC: Intra-class correlation coefficients; IDH: isocitrate dehydrogenase; KRAS: Kirsten rat sarcoma viral oncogene homolog; LASSO: least absolute shrinkage and selection operator; LBP: local binary pattern; MGMT: O-6-methylguanine-DNA methyltransferase; MiVI: microvascular invasion; MRI: Magnetic Resonance Imaging; NSCLC: non-small cell lung cancer; OS: Overall survival; PET: Positron Emission Tomography; PFS: Progression-free survival; ROC: receiver operating characteristic; ROI: region of interest; RQS: radiomics quality score; SIFT: scale-invariant feature transform; SUV: standard uptake value; TKI: tyrosine kinase inhibitor; US: ultrasound.

### Acknowledgements

This paper is supported by the Beijing Natural Science Foundation under Grant No. 7182109, the National Natural Science Foundation of China (Grant Nos. 81772012 and 81227901), the National Key Research and Development Plan of China under Grant Nos. 2017YFA0205200, Chinese Academy of Sciences under Grant No. GJSTDD2017004. The authors would like to acknowledge the instrumental and technical support of Multi-modal biomedical imaging experimental platform, Institute of Automation, Chinese Academy of Sciences.

### Competing Interests

The authors have declared that no competing interest exists.

### References

1. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, et al. Radiomics: extracting more information from medical images using advanced feature analysis. Eur J Cancer. 2012; 48: 441-6.
2. Kumar V, Gu Y, Bau S, Berglund A, Eschrich SA, Schabath MB, et al. Radiomics: the process and the challenges. Magn Reson Imaging. 2012; 30: 1234-48.
3. Gatenby RA, Cobleigh MJ, Gillies RJ, Quantitative Imaging in Cancer Evolution and Ecology. Radiology. 2013; 269: 8-15.
4. Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun. 2014; 5: 4091.
5. Aerts HJ. The potential of radiomic-based phenotyping in precision medicine: a review. JAMA Oncol. 2016; 2: 1636-42.
Arthur D, Vassilvitskii S. k-means++: The Advantages of Careful Initialization. 2016; 43: 3303-1.

Wang S, Zhou Z, Liu Z, Gu D, Zhang YL, et al. Central focused convolutional neural networks: Developing a data-driven model for lung nodule segmentation. Med Image Anal. 2017; 40: 172-80.

Liu Y, Kim J, Qiu F, Liu S, Wang H, Balagurunathan Y, et al. CT Features Associated with Epidermal Growth Factor Receptor Mutation Status in Patients with Lung Adenocarcinoma. Radiology. 2016; 280: 271-80.

Chickoree S, Goh V, Siddique M, Rivas A, Marsden PK, Cook C. Quantifying tumour heterogeneity in 18F-FDG PET/CT by texture analysis. Eur J Nucl Med Mol Imaging. 2013; 43: 130-43.

Parekh V, Jacobs MA. Radiomics: a new application from established inflammation. Eur Radiol. 2018; 28: 3872-81.

Scaio E, Rizzio G. Texture analysis of medical images for radiotherapy applications. Br J Radiol. 2017; 90: 20160472.

Ojala T, Pietikainen M, Maenpaa T. Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns. IEEE Trans Pattern Anal. 2000; 22: 971-87.

Ambrosini RD, Wang P, O'Dell WG. Computer-Aided Detection of Metastatic Brain Tumors Using Automated Three-Dimensional Template Matching. Journal of Magnetic Resonance Imaging. 2010; 31: 85-93.

Litjens G, Kooi T, Bejnordi BE, Setio AA, Ciompi F, Ghafoorian M, et al. A Survey on Deep Learning in Medical Image Analysis. Med Image Anal. 2017; 42: 60-88.

Rios VE, Parmar C, Liu Y, Dutt I, Seaforth C, O'Donohue G, et al. Somatic Mutations Drive Distinct Imaging Phenotypes in Lung Cancer. Cancer research. 2017; 77: 3922-30.

Yu DD, Zhou M, Yang P, Dong D, Gevaert O, Liu ZY, et al. Convolutional Neural Network for Predicting Molecular Profiles of Non-Small Cell Lung Cancer. 2017 13e International Symposium on Biomedical Imaging (Ibisi 2017). 2017: 569-72.

Chaudhary K, Poirion OB, Lu L, Garmire LX. Deep Learning-Based Multi-Omics Integration Robustly Predicts Survival in Liver Cancer. Clin Cancer Res. 2018; 24: 1248-59.

Litjens G, Kooi T, Bejnordi BE, Setio AA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. Med Image Anal. 2017; 42: 60-88.

Rios VE, Parmar C, Liu Y, Coroller TP, Cruz G, Stringfield O, et al. Somatic Mutations Drive Distinct Imaging Phenotypes in Lung Cancer. Cancer research. 2017; 77: 3922-30.

Guo J, Liu Z, Chen C, Li Z, Yan F, Tian J, et al. MR-based radiomics signature in differentiating ocular adenocarcinoma from idiopathic orbital pseudotumor. Spine (Phila Pa 1976). 2018; 43: 1052-8.

Tang Z, Liu Z, Li R, Yang X, Cui X, Wang S, et al. Identifying the white matter changes in brain using fMRI analysis. Brain Imaging and Behavior. 2018; 12: 1715-27.

Hatt M, Lee JA, Schmidtlein CR, Naqa IE, Caldwell C, De Bernardi E, et al. A Consensus-Based Workflow for Assessing the Quality of Multiparametric MRI as Novel Prognostic Factors in Advanced Nasopharyngeal Carcinoma. Clinical Cancer Research An Official Journal of the American Association for Cancer Research. 2017; 23: 4259.
