European Radiology (2019) 29:5431–5440
https://doi.org/10.1007/s00330-019-06169-w

ONCOLOGY

Image-based biomarkers for solid tumor quantification

Peter Savadjiev[1] · Jaron Chong[2] · Anthony Dohan[2,3] · Vincent Agnus[4] · Reza Forghani[2,5] · Caroline Reinhold[2] · Benoit Gallix[2,4]

Received: 26 November 2018 / Revised: 25 February 2019 / Accepted: 14 March 2019 / Published online: 8 April 2019

Abstract

The last few decades have witnessed tremendous technological developments in image-based biomarkers for tumor quantification and characterization. Initially limited to manual one- and two-dimensional size measurements, image biomarkers have evolved to harness developments not only in image acquisition technology but also in image processing and analysis algorithms. At the same time, clinical validation remains a major challenge for the vast majority of these novel techniques, and there is still a major gap between the latest technological developments and image biomarkers used in everyday clinical practice. Currently, the imaging biomarker field is attracting increasing attention not only because of the tremendous interest in cutting-edge therapeutic developments and personalized medicine but also because of the recent progress in the application of artificial intelligence (AI) algorithms to large-scale datasets. Thus, the goal of the present article is to review the current state of the art for image biomarkers and their use for characterization and predictive quantification of solid tumors. Beginning with an overview of validated imaging biomarkers in current clinical practice, we proceed to a review of AI-based methods for tumor characterization, such as radiomics-based approaches and deep learning.

Key Points

- *Recent years have seen tremendous technological developments in image-based biomarkers for tumor quantification and characterization.*
- *Image-based biomarkers can be used on an ongoing basis, in a non-invasive (or mildly invasive) way, to monitor the development and progression of the disease or its response to therapy.*
- *We review the current state of the art for image biomarkers, as well as the recent developments in artificial intelligence (AI) algorithms for image processing and analysis.*

Keywords Diagnostic imaging · Biomarkers · Artificial intelligence (AI) · Computer-assisted image processing · Computer-assisted image interpretation

✉ Benoit Gallix
benoit.gallix@mcgill.ca

[1] Department of Diagnostic Radiology, McGill University, Montreal, QC, Canada

[2] Department of Diagnostic Radiology, McGill University Health Centre, McGill University, 1001 Décarie Boulevard, Montreal, QC H4A 3J1, Canada

[3] Department of Body and Interventional Imaging, Hôpital Lariboisière-AP-HP, Université Diderot-Paris 7 and INSERM U965, 2 rue Ambroise Paré, 75475 Paris Cedex 10, France

[4] Institut de chirurgie guidée par l'image IHU Strasbourg, 1, place de l'Hôpital, 67091 Strasbourg Cedex, France

[5] Department of Radiology, Jewish General Hospital, 3755 Chemin de la Côte-Sainte-Catherine, Montreal, QC H3T 1E2, Canada

Abbreviations

^{18}F-FDG PET	^{18}F-fluorodeoxyglucose positron emission tomography
AI	Artificial intelligence
CNN	Convolutional neural network
EASL	European Association for the Study of the Liver
mRECIST	Modified Response Evaluation Criteria in Solid Tumors
PERCIST	Positron Emission Tomography Response Criteria in Solid Tumors
RECIST	Response Evaluation Criteria in Solid Tumors
WHO	World Health Organization

Introduction

An important objective in oncology is the creation of a standardized set of criteria to predict and monitor tumor response to treatment and for outcome prognosis based on objectively measured biomarkers. This is important for optimal clinical decision-making in routine clinical practice, as well as in clinical trials.

According to the FDA-NIH definition [1], a biomarker is "a defined characteristic that is measured as an indicator of normal biological processes, pathogenic processes, or responses to an exposure or intervention, including therapeutic interventions." Image-based biomarkers are particularly attractive, because they can provide a comprehensive view of the entire extent of the tumor and can capture regional tumor heterogeneity [2–4]. They can be used on an ongoing basis, in a non-invasive (or mildly invasive) way, to reveal tumor phenotype associated with prognosis and to monitor the development and progression of the disease or its response to therapy. In particular, they have the potential to observe changes early on in the course of treatment, thus providing an opportunity to tailor treatment based on the observed response [4, 5].

For more than 20 years, significant effort has been deployed to derive novel imaging biomarkers, resulting in a large amount of published literature. However, for the vast majority of image biomarkers, clinical validation has been a major hurdle. For this reason, manually measured biomarkers of tumor size are still the most commonly used in everyday clinical practice.

Outside of validated clinical use, however, the imaging biomarker field has been very active, especially with the recently reported success of artificial intelligence (AI)–based algorithms on large-scale datasets. Thus, the goal of the present article is to review the current state of the art for image biomarkers and their use for prognostic and predictive quantification of solid tumors. After a brief discussion of validated imaging biomarkers in current clinical practice, we will focus on AI-based methods for tumor characterization, such as traditional radiomics and radiomics based on deep learning, which we refer to as "deep" radiomics. We aim to adopt a critical approach in reviewing the strengths of each method but also the associated challenges in applying them to radiological image analysis. As such, the present paper is a continuation of an earlier review article on AI-based methods in radiology [6].

Validated imaging biomarkers for oncology in current clinical use

Size- and volume-based imaging biomarkers

In current clinical practice, solid tumor response to therapy is typically measured using 1D or 2D descriptors of tumor size. This standard was adopted in the early 1980s (WHO criteria) [7] and has evolved to become the Response Evaluation Criteria in Solid Tumors (RECIST) [8, 9]. Even though the RECIST criteria are very basic, only measuring tumor size, they remain the only tool currently in widespread use in both clinical practice and clinical trials for evaluating cancer therapeutics in solid tumors. The simplicity of the measurement makes it a very general and robust one and allows it to be used with a large variety of tumor types and treatment approaches in widely different settings and practices.

Such 1D and 2D size criteria have been compared to volumetric tumor measurements, either the full 3D tumor volume or the approximations carried out by measuring three orthogonal tumor axes (e.g., [10]). The comparison results have been somewhat inconclusive, with some studies finding that 3D measures are better than 1D measures [11, 12], some finding a similar performance for both methods [10, 13], and some finding no correlation between any of these methods and survival [14].

Size- and volume-based measurements of tumor response to treatment suffer from important limitations [4, 15]. The manual measurement of tumor size is subject to inter-observer variability, especially in the presence of irregular and/or infiltrating lesions. In addition, measurable changes in tumor size may be significantly delayed with respect to the underlying physiological changes that may occur much earlier. Furthermore, several new treatments may produce changes in shape, morphology, vascularization, or cell metabolism much earlier or in some instance independent of lesion size reduction [15, 16].

Biomarkers combining tumor size and appearance

In an attempt to address some of these issues, proposals have been made to take into account not only tumor size but also qualitative descriptions of intra-tumor changes in appearance before and after treatment. Examples include measuring only the part of the tumor that remains vascularized, as opposed to measuring the entire tumor (EASL [17] or mRECIST [18]), or the association of size with other variables such as tumor density as measured by CT (e.g., Choi criteria [19]). However, such biomarkers are limited to simple empiric measurements of morphological change, and they are usually applicable only to very specific types of tumors and treatments. Because of this, their use remains limited in clinical practice.

Automated size and volume measurements using segmentation

The measurement of size and volume of a particular structure in an image can be automated by first computing the structure's boundaries, a process known as "segmentation." For the last three decades, automatic or semi-automatic image

segmentation has been one of the major applications of AI-based research in medical image analysis (e.g., [6]). Given the very large number of segmentation algorithms that exist today, several competitions have been created to compare, evaluate, and validate different algorithms for different types of imaging. A few examples of such competitions are listed in Table 1. However, even though such international competitions have become the de facto standard for evaluating medical image analysis algorithms, the conception, design, and implementation of such competitions can oftentimes be inadequate or deficient [20]. Thus, the proper evaluation of such algorithms remains a work in progress.

In oncology, the automatic segmentation of tumors can be used to alleviate some of the challenges associated with manual measurement of tumor size and volume, by making the process faster, more cost-effective, and more reproducible. Nevertheless, despite years of progress, fully automatic segmentation remains difficult to achieve in practice, and manual adjustments of the automatic results are almost always needed. In addition, automatic or semi-automatic segmentation algorithms are sensitive to acquisition protocol parameters, which makes it difficult to evaluate the reproducibility of the segmentation results in real clinical practice. On the other hand, promising results have recently been achieved with deep learning methods (see the section "Deep radiomics: fully automated quantification") that could have an impact on tumor assessment in the clinical workflow.

Table 2 A summary of the typical features used in a traditional radiomic workflow

Category	Examples
Size	• Volume • Maximum 3D diameter • Surface area
Shape	• Sphericity • Elongation • Flatness
1st-order statistical analysis—intensity histogram features	• Mean • Variance • Kurtosis
2nd-order statistical analysis—texture features	Features derived from texture matrices, such as • Gray-level co-occurrence matrix • Gray-level run-length matrix • Gray-level size-zone matrix
Higher-order statistical analysis	• Fractals • Wavelets

The first-order statistical descriptors reflect the global distribution of pixel values within a tumor region and are among the most popular in the literature. The second-order statistical determinants take into account the spatial relationship between pixel values. Such features are better known as "texture" features, as they capture patterns with different textural appearance properties. Shape-specific features may be defined to capture morphological tumor properties. Other more advanced features can capture various fractal properties of the tumor appearance (see, e.g., [29, 30] for further details)

Beyond size and volume: strategies for image biomarker development

Recognizing the need to go beyond tumor size and volume, two broad strategies for image biomarker development have become popular in the literature. The first strategy consists in deriving biomarkers from the signal of image acquisitions designed to be specific to particular aspects of tumor chemistry and biology. Novel structural, functional, and metabolic imaging methods have recently been developed to detect small-scale changes in the tumor cytoarchitecture, chemical composition, and metabolic processes [4, 16]. Examples include diffusion-weighted MRI (DW-MRI), dynamic contrast-enhanced MRI (DCE-MRI), magnetic resonance spectroscopy (MRS), positron emission tomography (PET) with specific molecular targets, and others. Among these, ^{18}F-fluorodeoxyglucose positron emission tomography

Table 1 Examples of tumor segmentation competitions

Competition	Image modalities	URL
Multimodal Brain Tumor Segmentation Challenge (BRaTS)	MRI: •T1 • T1Gd • T2 • FLAIR	http://braintumorsegmentation.org/
Liver Tumor Segmentation Challenge (LiTS)	CT	http://lits-challenge.com
QIN Lung CT Segmentation Challenge	CT	https://wiki.cancerimagingarchive.net/display/Public/QIN+Lung+CT+Segmentation+Challenge
NCI-ISBI Challenge on Automated Segmentation of Prostate Structures	MRI: • T1 • T2	https://wiki.cancerimagingarchive.net/display/Public/NCI-ISBI+2013+Challenge+-+Automated+Segmentation+of+Prostate+Structures

(^{18}F-FDG PET) has gained popularity as an imaging modality sensitive to tumor metabolic activity [21]. A set of criteria for structured quantitative image analysis and clinical reporting based on ^{18}F-FDG PET has been proposed as the PERCIST 1.0 criteria [22].

However, while these novel imaging methods have shown great promise in a research setting in academic centers, their generalization to mainstream clinical practice or to large multi-center clinical trials has been difficult. This is because such techniques are difficult to standardize and reproduce, since usually they are specific to a particular vendor's equipment and sensitive to acquisition parameters, contrast injection, and patient physiology.

Because of this, a second strategy is based on using more standard-of-care imaging techniques that are already in wide clinical use, such as conventional CT and MRI sequences. The goal is to better leverage the signal measured within the tumor with standard imaging techniques, by defining biomarkers as a function of a large number of image descriptors (features) extracted from the image (e.g., [2]). These image features are then analyzed with advanced statistical or machine learning techniques, in order to discover the most meaningful/discriminative subset of features that achieves the best performance at tasks such as outcome prediction. The rest of this paper is focused on methods in this category.

Semantic features: quantification through the radiologist's eye

The feature-based characterization of tumors was initially carried out using so-called "semantic" features, i.e., qualitative or semi-quantitative features that are part of radiologist's lexicon and are determined from the image by a trained radiologist. An early example of this approach is found in [23], where 28 radiologist-scored features are found to predict gene expression patterns in hepatocellular carcinoma. Examples of features include the presence of arteries within the tumor, the presence of a hypodense halo around the tumor, apparent tumor heterogeneity, etc. In 2009, similar semantic features were introduced for the evaluation of the response of colorectal liver metastasis treated by bevacizumab [24]. More recently, there has been an increase in radiogenomic studies involving semantic features, used either alone or in conjunction with quantitative features [25]. Unfortunately, the process of visual semantic feature quantification can be subjective. It also does not scale well with increasingly large datasets, in terms of time, workflow integration, and indirect cost, since manual annotation takes considerable amounts of time and may involve large intra- and inter-rater variability.

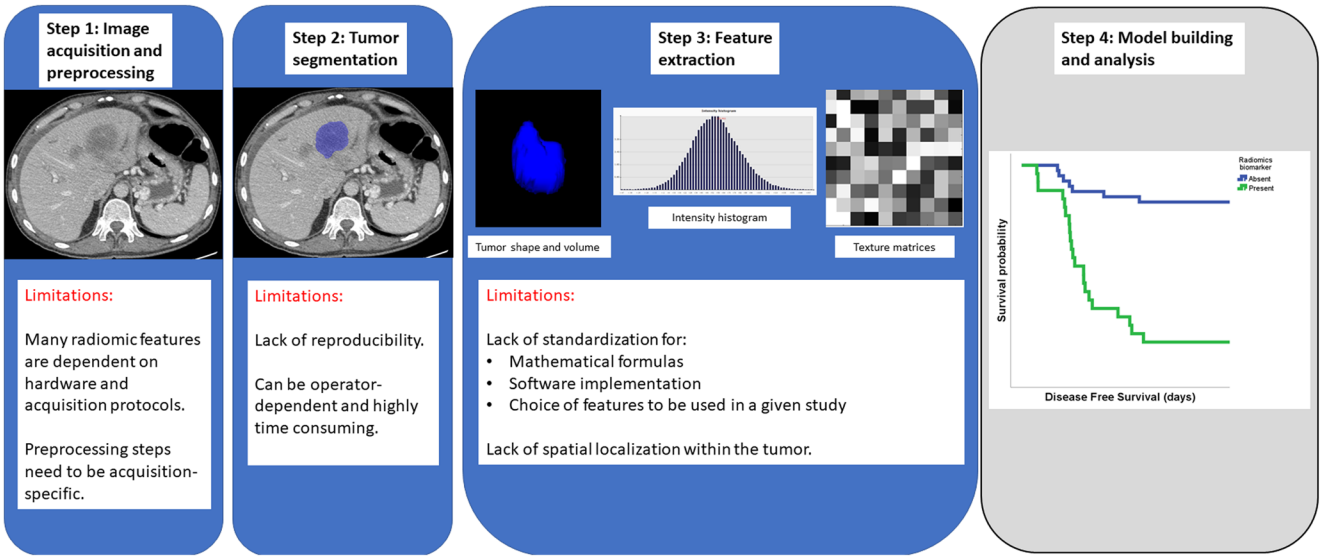

Fig. 1 A summary of the traditional radiomic workflow. The heart of the approach consists in computing features from a segmented region in the image, which defines the spatial extent of the tumor. Many of these features are dependent on image acquisition parameters, such as signal-to-noise ratio, image resolution, slice thickness, and intensity of the enhancement after contrast injection (e.g., [34, 35]). Thus, image preprocessing is a necessary first step in order to normalize the images for subsequent analysis. Unfortunately, preprocessing on its own is rarely sufficient to compensate for all differences in acquisition. Following the feature extraction step, a feature pruning step is required. That is because the computation of a large number of features from the same data can result in redundant and/or highly correlated features, which increases dramatically the dimensionality of the problem without adding useful information. Because of this, reducing the number of features by selecting the most relevant features can significantly increase classification performance [36]. Finally, the last step of the radiomic pipeline is the classification of different types of tissue, based on the features computed in the previous steps. The task of classification is not specific to radiomics. A large part of the machine learning field is concerned with classification and provides a wide variety of tools that can be used to solve the problem [37]

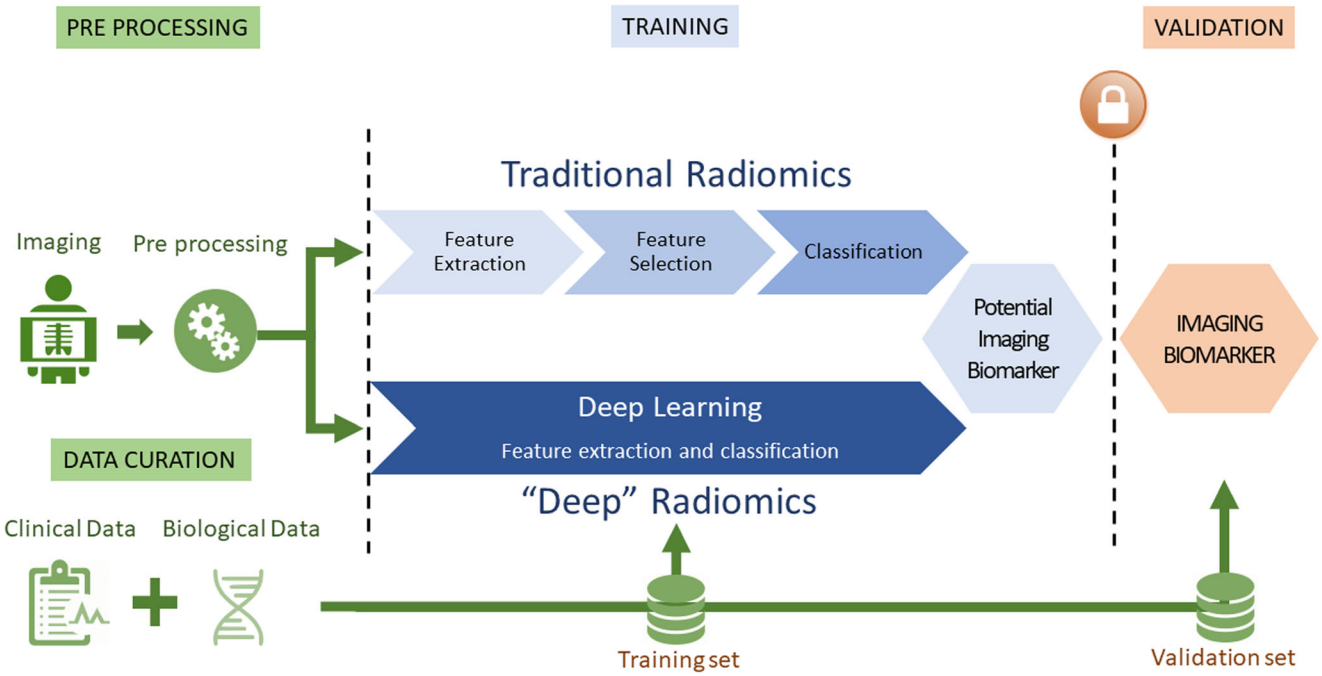

Fig. 2 Traditional and "deep" radiomic processing. The first step consists in gathering the imaging data, as well as the clinical and biological data that determine the *gold standard* biomarkers must be characterized. This gold standard could consist, for instance, in a patient outcome such as overall survival, information about tumor biology (such as EGFR mutation in lung cancer (genotype)), or an expert classification of the images if the goal is to reproduce radiologist analysis. Data curation refers to the organization, integration, annotation, and presentation of the data collected. Data should be divided into a training set and one or multiple validation sets. The goal of medical image preprocessing is to decrease the technical variability between different image batches. Discovering new biomarkers through radiomics can be done using traditional radiomics (section "Traditional radiomics: quantification through pre-determined features") or with deep learning methods (section "Deep radiomics: fully automated quantification"). After a biomarker consisting of a combination of basic image features is identified during the training phase, it is then validated on an independent data cohort

Big-data image analysis for tumor phenotyping

Traditional radiomics: quantification through pre-determined features

To address the barriers associated with manual tumor quantification via semantic features, there has recently been an important effort towards the automation of tumor phenotype quantification. This is done with the computation of *agnostic* features, i.e., mathematically defined quantitative descriptors, most of which are not part of the radiologists' lexicon, and which are computed automatically by image analysis algorithms.

This approach has become known as radiomics [2, 26–28]. It is a hypothesis-free approach based on agnostic features; i.e., no a priori hypothesis is made about the clinical relevance of the features. Rather, the goal here is to automatically discover previously unseen patterns from a very large number of features describing the tumor's appearance and to perform classification based on the most discriminative subset of these features, thus constructing a radiomic *signature*. Being part of the agnostic category, the features used with the traditional radiomic approach are pre-defined by image processing experts. Some of the most commonly used ones are summarized in Table 2. Several works have described extensively this process (e.g., [2, 26–28]), and it has been applied in many research studies in radiology and oncology (e.g., [26, 31–33]). A summary of its main steps, as well as the associated limitations, is provided in Fig. 1.

Deep radiomics: fully automated quantification

The advent of large digital datasets in almost every sphere of technological activity has fostered an increased interest in hypothesis-free, data-driven analysis techniques, which emphasize mining of large datasets to discover new patterns and to help formulate new hypotheses [38]. Radiomics is one example of such an approach, as it allows to formulate new candidate biomarkers as a result of analyzing a large number of image features that may not be accessible to the radiologist's eye. Each such candidate biomarker then becomes a clinical hypothesis that could be tested in a clinical setting. This discovery process can help augment existing human knowledge and is one of the powers of modern AI approaches to medical image analysis.

In contrast to traditional radiomics, where the features are handpicked by a human image processing expert, a class of machine learning methods known as deep learning [39] has

recently gained popularity because they can automatically discover the best features for a given task, without requiring human intervention for feature design [6, 40–42]. See Fig. 2 for a schematic comparison between traditional and deep radiomics. Recent studies have shown an improvement in performance by deep learning methods over traditional radiomics (e.g., [43, 44]).

Solid tumor analysis via deep learning

The models underlying deep learning algorithms are multilayer artificial neural networks, which typically contain millions of parameters [39]. The strength of such complex models is that they can learn highly complex data representations in a manner that can often surpass human ability. The downside is that models with millions of parameters need millions of data examples in order to train, at least for their initial training. In areas where large image datasets with unambiguous labeling of their contents are available, e.g., online databases of photographs of the real world such as images of animals and man-made objects (see Table 3 for examples of datasets), deep learning has been very successful. In radiology, however, large image datasets can be difficult to assemble and curate [6]. Privacy rules and ethical constraints protect patient data and limit accessibility to image data outside approved use, both within and outside the institution that acquired the images, unless rigorous deidentification procedures have been followed [45]. In addition, the low prevalence of certain types of disease presents a practical barrier to gathering images from a sufficiently large number of patients presenting the disease. In turn, this may oftentimes result in unbalanced datasets, where only a small proportion of the images represent a disease state, which poses problems for several machine learning algorithms. Finally, a majority of deep learning algorithms have been designed for labeled data, i.e., data where image content has been pre-identified in order for the algorithm to train with known object examples. Such labeling is easier to achieve in photographs of the real world, where particular expertise is not required to identify, e.g., animals, planes, or cars. In contrast, identifying disease in radiological images requires assessment by a trained radiologist and, oftentimes, a consensus among several radiologists. Despite all these barriers, multi-year and multi-institutional efforts have made some medical image databases publicly available, with examples listed in Table 3.

To circumvent challenges posed by the small size of datasets in application domains such as medical imaging, techniques have been developed that make harnessing the power of deep learning algorithms on small-size datasets possible. One such technique is known as "data augmentation," which consists in artificially augmenting the dataset by generating new artificial images, for instance by deforming the initial data [46]. However, when applied to radiological data, it is important to ensure that the synthetically generated data maintains the same

Table 3 Examples of image repositories, used for developing, testing, and validating image analysis algorithms, both in the radiological domain and in the world of natural images

Database	Number of images	Image types
Example of real-world image datasets		
ImageNet (http://www.image-net.org/)	1.4 million	Tens of thousands object categories, organized according to a lexical database
COCO (http://cocodataset.org)	330,000	91 object categories
CIFAR-100 (https://www.cs.toronto.edu/~kriz/cifar.html)	60,000	100 object categories
Caltech 256 (http://www.vision.caltech.edu/Image_Datasets/Caltech256/)	30,608	256 object categories
Example medical image datasets		
NIHCC ChestXRay (https://nihcc.app.box.com/v/ChestXray-NIHCC)	112,120	X-ray images of 30,805 unique patients with 14 disease image labels
The Cancer Imaging Archive (TCIA) (http://www.cancerimagingarchive.net/)	Variable	Images organized in collections based on disease, image modality, or research focus
Open Access Series of Imaging Studies (OASIS-3) (https://www.oasis-brains.org/)	1098 subjects, 3776 imaging sessions	Longitudinal neuroimaging dataset for normal aging and Alzheimer's disease
Breast Cancer Digital Repository (https://bcdr.eu/)	1734 subjects	Mammography, ultrasound, annotated and BIRADS classified
Grand Challenges in Biomedical Image Analysis (https://grand-challenge.org/)	Variable	Data from more than 100 medical imaging competitions
Alzheimer's Disease Neuroimaging Initiative (ADNI) (http://adni.loni.usc.edu/)	1921 subjects	Brain imaging data from Alzheimer's disease patients, mild cognitive impairment patients, and elderly control subjects
Human Connectome Project (https://www.humanconnectome.org/study/hcp-young-adult/document/1200-subjects-data-release)	1200 subjects	Brain imaging data from young adults

relationship to outcome as the initial data. To do so, only simple geometric transforms (e.g., translation, flip, rotation) can be safely used. Another strategy is to make use of networks already pre-trained on a different dataset of a sufficiently large size, an approach known as transfer learning. Such pre-trained networks can then be fine-tuned with the current dataset at hand [47, 48]. In the long term, as the performance and data efficiency of deep learning algorithms improves, it is conceivable that the data requirements will be reduced and the impact of such methods to healthcare drastically increased [49].

The most common application of deep learning to image analysis lies in the classification of entire images into two or more categories, based, for instance, on whether they show disease or not (e.g., [50–52]). Alternatively the classification can be performed at the level of each image pixel, labeling it as belonging (for instance) to background, organ, or tumor, effectively performing image segmentation [53].

As opposed to the more general problem of image classification/segmentation, tumor characterization represents a more challenging task, since the lesion appearance in 3D, as well as the 3D contextual information, must be taken into account. Traditional deep learning architectures developed for 2D image classification are not well suited for this purpose. Adaptations of 2D deep learning techniques to the full 3D image domain have been carried out, for instance, with so called "multi-stream" or "multi-view" strategies. In this class

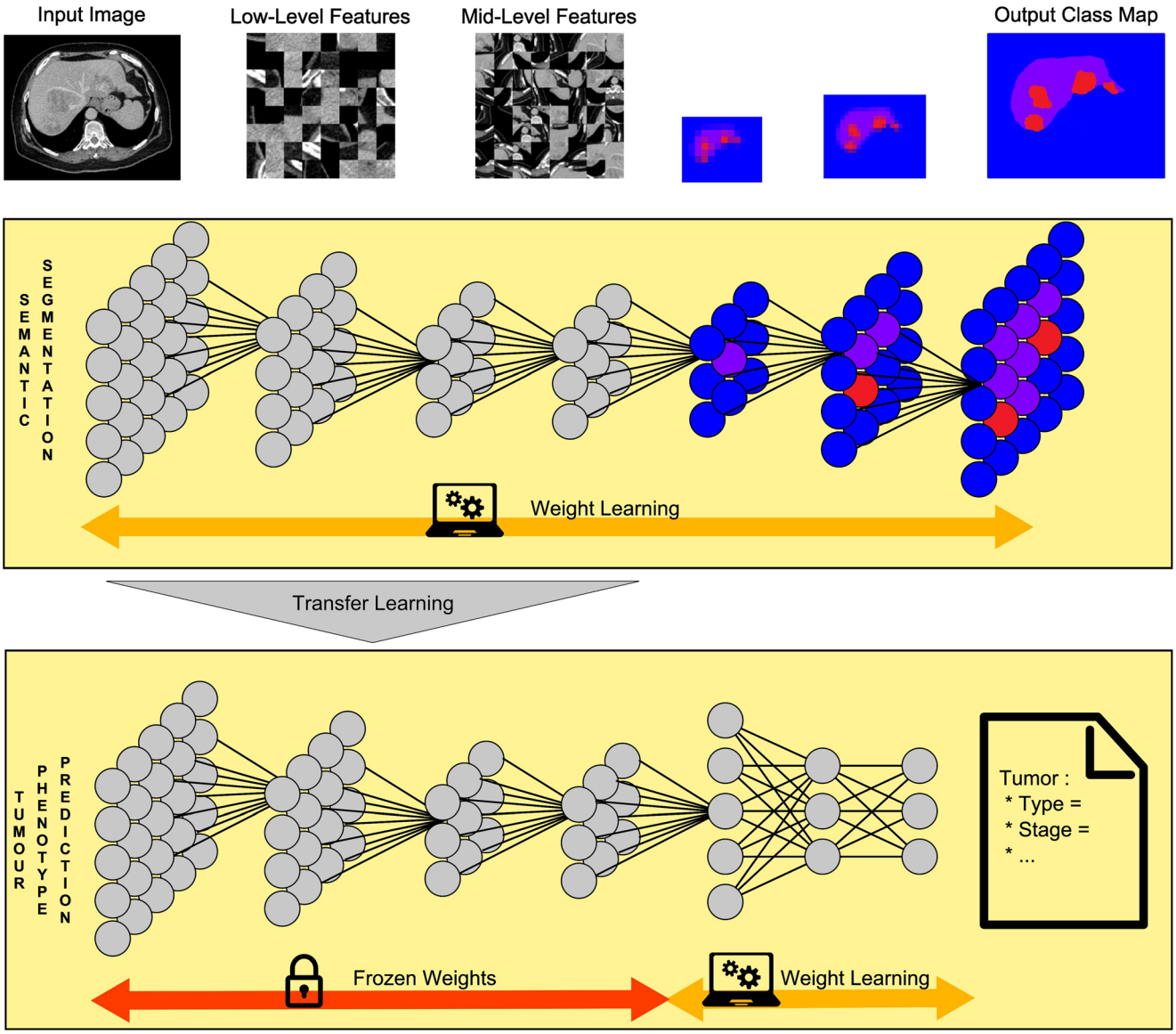

Fig. 3 Tumor characterization using a two-step deep neural network analysis. The first step is devoted to retrieving radiomic features. It focuses on semantic segmentation of the liver and hepatic tumors. The second network uses these features as an input to predict the tumor phenotype. A transfer learning approach was used to import the image encoding from the first network. These weights are locked during the training of the second neural network

of methods, multiple neural networks analyze different 2D aspects of tumor appearance, for instance different 2D image patches through the tumor (e.g., [54]) or different spatial scales [55]. Such approaches to 3D image analysis are often ad hoc, but heuristics are oftentimes needed in order to avoid the dramatic increase of computational requirements associated with truly 3D convolutional neural network (CNN) architectures [56].

Alternative approaches could be constructed with a two-step analysis, as illustrated in Fig. 3. First, a segmentation-specific CNN can be applied to delineate the organ and tumor, thus narrowing the analysis only to the relevant parts of the image. Then, the image encoding in terms of radiomic features can be *locked* and transferred to a second neural network, which can be trained to classify tumor phenotype, given the encoding of tumor information achieved at the first step.

Limitations

Deep radiomics for tumor characterization comes with limitations, such as challenges in reproducibility and interpretability. Unlike traditional radiomics where each feature is described with a mathematical formula, deep radiomic features are not easily interpretable. This is because a deep neural network models a large number of complex but weak regularities in the data via highly complex and non-linear interactions between multiple network nodes and layers [49]. As such, the model is not designed to be interpretable. Furthermore, a multi-layer deep network with possibly millions of parameters is not easily reproducible. For the model to be shared and/or reproduced, e.g., by a different group in a different institution, either the entire trained network must be shared or, alternatively, a network with the same architecture needs to be retrained, with the same training data and the same parameter initialization, which is not a simple task.

Conclusion

Efforts to develop image biomarkers for tumor characterization have been ongoing for decades. However, despite the large number of proposed biomarkers, only a handful has been adopted widely in clinical practice. With the recent advances in data-mining AI algorithms, there has been a surge of interest in a *discovery science* approach to image biomarker development through the radiomic approach, using either human-defined image descriptors or deep learning algorithms where the best features for image characterization are discovered automatically.

This novel data-driven approach offers a tremendous potential for innovation and discovery. However, for AI methods to become practical and useful in the clinic, rigorous translation pipelines as well as workflow integration still need to be established, as discussed elsewhere (e.g., Section 5 in [6, 57]). The AI discovery process can augment human knowledge, but it also needs to be constrained by expert knowledge where appropriate. This will enable focusing the AI analysis and making it more efficient but also will serve to ensure better interpretability, stability, and clinical acceptability of the results.

Funding The authors state that this work has not received any funding.

Compliance with ethical standards

Guarantor The scientific guarantor of this publication is Dr. Benoit Gallix.

Conflict of interest The authors declare that they have no competing interests.

Statistics and biometry No complex statistical methods were necessary for this paper.

Informed consent Written informed consent was not required for this study because this is a review article, and no study was performed.

Ethical approval Institutional review board approval was not required because this is a review article, and no study was performed.

References

1. Food and Drug Administration & National Institutes of Health. BEST (biomarkers, endpoints, and other tools) resource. NCBI http://www.ncbi.nlm.nih.gov/books/NBK326791. Accessed on 10 Jan 2019.
2. Aerts HJ (2016) The potential of radiomic-based phenotyping in precision medicine. A review. JAMA Oncol 2(12):1636–1642
3. Amin S, Bathe OF (2016) Response biomarkers: re-envisioning the approach to tailoring drug therapy for cancer. BMC Cancer 16:850
4. Harry VN, Semple SI, Parkin DE, Gilbert FJ (2010) Use of new imaging techniques to predict tumour response to therapy. Lancet Oncol 11:92–102
5. O'Connor JP, Aboagye EO, Adams JE et al (2017) Imaging biomarker roadmap for cancer studies. Nat Rev Clin Oncol 14(3):169–186
6. Savadjiev P, Chong J, Dohan A, et al (2019) Demystification of AI-driven medical image interpretation: past, present and future. Eur Radiol 29(3):1616–1624
7. World Health Organization. (1979) . WHO handbook for reporting results of cancer treatment. World Health Organization. Geneva, Switzerland https://www.who.int/iris/handle/10665/37200
8. Miller AB, Hoogstraten B, Staquet M, Winkler A (1981) Reporting results of cancer treatment. Cancer 47(1):207–214
9. Therasse P, Arbuck SG, Eisenhauer EA et al (2000) New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada. J Natl Cancer Inst 92(3):205–216
10. Shah GD, Kesari S, Xu R, et al (2006) Comparison of linear and volumetric criteria in assessing tumor response in adult high-grade gliomas. Neuro Oncol 8(1):38–46

11. Dempsey MF, Condon BR, Hadley DM (2005) Measurement of tumor "size" in recurrent malignant glioma: 1D, 2D, or 3D? AJNR Am J Neuroradiol 26(4):770–776
12. Aghighi M, Boe J, Rosenberg J et al (2016) Three-dimensional radiologic assessment of chemotherapy response in Ewing sarcoma can be used to predict clinical outcome. Radiology 280(3):905–915
13. Lubner MG, Stabo N, Lubner SJ, Del Rio AM, Song C, Pickhardt PJ (2017) Volumetric versus unidimensional measures of metastatic colorectal cancer in assessing disease response. Clin Colorectal Cancer 16(4):324–333
14. Galanis E, Buckner JC, Maurer MJ et al (2006) Validation of neuroradiologic response assessment in gliomas: measurement by RECIST, two-dimensional, computerassisted tumor area, and computer-assisted tumor volume methods. Neuro Oncol 8(2):156–165
15. Jaffe CC (2006) Measures of response: RECIST, WHO, and new alternatives. J Clin Oncol 24(20):3245–3251
16. Atri M (2006) New technologies and directed agents for applications of cancer imaging. J Clin Oncol 24(20):3299–3308
17. Bruix J, Sherman M, Llovet JM et al (2001) Clinical management of hepatocellular carcinoma. Conclusions of the Barcelona-2000 EASL conference. European Association for the Study of the Liver. J Hepatol 35(3):421–430
18. Lencioni R, Llovet JM (2010) Modified RECIST (mRECIST) assessment for hepatocellular carcinoma. Semin Liver Dis 30:52–60
19. Choi H, Charnsangavej C, Faria SC et al (2007) Correlation of computed tomography and positron emission tomography in patients with metastatic gastrointestinal stromal tumor treated at a single institution with imatinib mesylate: proposal of new computed tomography response criteria. J Clin Oncol 25:1753–1759
20. Maier-Hein L, Eisenmann M, Reinke A et al (2018) Why rankings of biomedical image analysis competitions should be interpreted with care. Nat Commun 9(1):5217
21. Kelloff GJ, Hoffman JM, Johnson B et al (2005) Progress and promise of FDG-PET imaging for cancer patient management and oncologic drug development. Clin Cancer Res 11(8):2785–2808
22. Wahl RL, Jacene H, Kasamon Y, Lodge MA (2009) From RECIST to PERCIST: evolving considerations for PET response criteria in solid tumors. J Nucl Med 50(Suppl 1):122S–150S
23. Segal E, Sirlin CB, Ooi C et al (2007) Decoding global gene expression programs in liver cancer by noninvasive imaging. Nat Biotechnol 25:675–680
24. Chun YS, Vauthey JN, Boonsirikamchai P et al (2009) Association of computed tomography morphologic criteria with pathologic response and survival in patients treated with bevacizumab for colorectal liver metastases. JAMA 302(21):2338–2344
25. Jansen RW, van Amstel P, Martens RM, Kooi IE, Wesseling P, de Langen AJ (2018) Non-invasive tumor genotyping using radiogenomic biomarkers, a systematic review and oncology-wide pathway analysis. Oncotarget 9(28):20134–20155
26. Aerts HJ, Velazquez ER, Leijenaar RT et al (2014) Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun 5:4006
27. Larue RT, Defraene G, De Ruysscher D, Lambin P, van Elmpt W (2017) Quantitative radiomics studies for tissue characterization: a review of technology and methodological procedures. Br J Radiol 90(1070):20160665
28. O'Connor JP, Rose CJ, Waterton JC, Carano RA, Parker GJ, Jackson A (2015) Imaging intratumor heterogeneity: role in therapy response, resistance, and clinical outcome. Clin Cancer Res 21(2):249–257
29. Vallières M, Freeman CR, Skamene SR, El Naq I (2015) A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities. Phys Med Biol 60:5471–5496
30. van Griethuysen JJM, Fedorov A, Parmar C et al (2017) Computational radiomics system to decode the radiographic phenotype. Cancer Res 77(21):e104–e107
31. Chamming's F, Ueno Y, Ferré R et al (2017) Features from computerized texture analysis of breast cancers at pretreatment MR imaging are associated with response to neoadjuvant chemotherapy. Radiology. https://doi.org/10.1148/radiol.2017170143
32. Ueno Y, Forghani B, Forghani R et al (2017) Endometrial carcinoma: MR imaging-based texture model for preoperative risk stratification—a preliminary analysis. Radiology 284(3):748–757
33. Parmar C, Leijenaar RTH, Grossmann P et al (2015) Radiomic feature clusters and prognostic signatures specific for lung and head & neck cancer. Sci Rep 5:11044
34. Schad L (2004) Problems in texture analysis with magnetic resonance imaging. Dialogues Clin Neurosci 6(2):235–242
35. Schad L, Lundervold A (2006) Influence of resolution and signal to noise ratio on MR image texture. In: Hajek M, Dezortova M, Materka A, Lerski R (eds) Texture analysis for magnetic resonance imaging. HRaNa, Prague, pp 129–149
36. Guyon I, Elisseeff A (2003) An introduction to variable and feature selection. J Mach Learn Res 3:1157–1182
37. Hastie T, Tibshirani R, Friedman JH (2009) The elements of statistical learning: data mining, inference, and prediction, 2nd edn. Springer, New York
38. Kraus WL (2015) Editorial: would you like a hypothesis with those data? Omics and the age of discovery science. Mol Endocrinol 29(11):1531–1534
39. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521:436–444
40. Chartrand G, Cheng PM, Vorontsov E et al (2017) Deep learning: a primer for radiologists. Radiographics 37(7):2113–2131
41. Litjens G, Kooi T, Bejnordi BE et al (2017) A survey on deep learning in medical image analysis. Med Image Anal 42:60–88
42. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL (2018) Artificial intelligence in radiology. Nat Rev Cancer 18(8):500–510
43. Ypsilantis PP, Siddique M, Sohn HM et al (2015) Predicting response to neoadjuvant chemotherapy with PET imaging using convolutional neural networks. PLoS One 10(9):e0137036
44. Li Z, Wang Y, Yu J, Guo Y, Cao W (2017) Deep learning based radiomics (DLR) and its usage in noninvasive IDH1 prediction for low grade glioma. Sci Rep 7(1):5467
45. US Department of Health and Human Services. Guidance regarding methods for de-identification of protected health information in accordance with the Health Insurance Portability and Accountability Act (HIPAA) privacy rule. http://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html. Accessed on 10 Jan 2019
46. Perez L, Wang J (2017) The effectiveness of data augmentation in image classification using deep learning. Retrieved from http://arxiv.org/abs/1712.04621
47. Tajbakhsh N, Shin JY, Gurudu SR et al (2016) Convolutional neural networks for medical image analysis: full training or fine tuning? IEEE Trans Med Imaging 35(5):1299–1312
48. Shin HC, Roth HR, Gao M et al (2016) Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. IEEE Trans Med Imaging 35(5):1285–1298
49. Hinton G (2018) Deep learning—a technology with the potential to transform health care. JAMA 320(11):1101–1102
50. Anthimopoulos M, Christodoulidis S, Ebner L, Christe A, Mougiakakou S (2016) Lung pattern classification for interstitial lung diseases using a deep convolutional neural network. IEEE Trans Med Imaging 35:1207–1216

51. Lakhani P, Sundaram B (2017) Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. Radiology 284(2):574–582
52. Chang P, Grinband J, Weinberg BD et al (2018) Deep-learning convolutional neural networks accurately classify genetic mutations in gliomas. AJNR Am J Neuroradiol 39(7):1201–1207
53. Ronneberger O, Fischer P, Brox T (2015) U-net: convolutional networks for biomedical image segmentation. In: Proceedings of the medical image computing and computer-assisted intervention. In: Lecture notes in computer science, vol 9351, pp 234–241
54. Setio AAA, Ciompi F, Litjens G et al (2016) Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks. IEEE Trans Med Imaging 35:1160–1169
55. Ciompi F, Chung K, van Riel SJ et al (2017) Towards automatic pulmonary nodule management in lung cancer screening with deep learning. Sci Rep 7:46479
56. Lai M (2015) Deep learning for medical image segmentation. Retrieved from https://arxiv.org/abs/1505.02000
57. Bodalal Z, Trebeschi S, Beets-Tan R (2018) Radiomics: a critical step towards integrated healthcare. Insights Imaging 9(6):911–914