Online version at http://www.minervamedica.it

The Quarterly Journal of Nuclear Medicine and Molecular Imaging 2020 September;64(3):278-90
DOI: 10.23736/S1824-4785.20.03263-X

REVIEW

NUCLEAR MEDICINE IN PRECISION ONCOLOGY

Nuclear medicine radiomics in precision medicine: why we can't do without artificial intelligence

Wyanne A. NOORTMAN [1, 2] *, Dennis VRIENS [1], Willem GROOTJANS [3], Qian TAO [4], Lioe-Fee de GEUS-OEI [1, 2], Floris H. VAN VELDEN [1]

[1]Section of Nuclear Medicine, Department of Radiology, Leiden University Medical Center (LUMC), Leiden, the Netherlands; [2]Biomedical Photonic Imaging Group, University of Twente, Enschede, the Netherlands; [3]Department of Radiology, Leiden University Medical Center (LUMC), Leiden, the Netherlands; [4]Division of Image Processing (LKEB), Department of Radiology, Leiden University Medical Center (LUMC), Leiden, the Netherlands

*Corresponding author: Wyanne A. Noortman, Section of Nuclear Medicine, Department of Radiology, Leiden University Medical Center (LUMC), Leiden, the Netherlands. E-mail: w.a.noortman@lumc.nl

ABSTRACT

In recent years, radiomics, defined as the extraction of large amounts of quantitative features from medical images, has gained emerging interest. Radiomics consists of the extraction of handcrafted features combined with sophisticated statistical methods or machine learning algorithms for modelling, or deep learning algorithms that both learn features from raw data and perform modelling. These features have the potential to serve as non-invasive biomarkers for tumor characterization, prognostic stratification and response prediction, thereby contributing to precision medicine. However, especially in nuclear medicine, variable results are obtained when using radiomics for these purposes. Individual studies show promising results, but due to small numbers of patients per study and little standardization, it is difficult to compare and validate results on other datasets. This review describes the radiomic pipeline, its applications and the increasing role of artificial intelligence within the field. Furthermore, the challenges that need to be overcome to achieve clinical translation are discussed, so that, eventually, radiomics, combined with clinical data and other biomarkers, can contribute to precision medicine, by providing the right treatment to the right patient, with the right dose, at the right time.

(*Cite this article as:* Noortman WA, Vriens D, Grootjans W, Tao Q, de Geus-Oei LF, Van Velden FH. Nuclear medicine radiomics in precision medicine: why we can't do without artificial intelligence. Q J Nucl Med Mol Imaging 2020;64:278-90. DOI: 10.23736/S1824-4785.20.03263-X)

KEY WORDS: Nuclear medicine; Diagnostic imaging; Radiography; Artificial intelligence.

In precision medicine, medical decisions and interventions are tailored to the needs of the individual patient and based on their susceptibility to a particular disease or predicted response to therapy.[1] In oncology, the choice of therapy for a patient is mostly based upon clinical parameters and molecular characterization of the tumor tissue, with biopsy as the gold standard.[2] However, biopsy comes with the risk of a sampling error, which may be caused by sampling only a small fraction of a heterogeneous tumor, missing the tumor entirely or heterogeneity between multiple lesions.[3] This might lead to misinterpretations and consequently suboptimal clinical management of these patients. The problems related to biopsies might be addressed through the use of the less invasive medical imaging, particularly given that medical imaging is used in routine clinical practice for diagnosis and staging of oncological patients. Due to its less-invasive character imaging can be repeated over time during the development of the disease and during treatment.

Unlike biopsies, medical imaging can provide information about the entire tumor or disease phenotype, including intra- and interlesional heterogeneity.[4] Traditionally, visual interpretation of medical images consists of (qualitative) assessment of size, shape, patterns, signal intensity

(*e.g*. Hounsfield unit for computed tomography [CT] or standardized uptake value [SUV] for positron emission tomography [PET]) or intravenous contrast enhancement of tumors. However, medical images contain much more information hidden in the millions of voxels of both tumors and healthy tissue[5] that cannot be assessed visually by a human observer. Recent developments in computer science have introduced computational methods that can capture this concealed information in the interest of lesion or disease detection, classification and diagnosis, segmentation and quantification.[6]

The extraction of a variety of quantitative features from standard medical imaging is studied in the field of radiomics. Radiomics consists of: 1) the extraction of predefined, handcrafted features that can be combined with sophisticated statistical methods or machine learning algorithms for modelling; 2) deep learning algorithms that both learn features from raw data and perform modelling.[7] Radiomic analysis aims to find stable and clinically relevant image-derived biomarkers, also known as radiomic features, that provide a non-invasive way of quantifying and monitoring tumor characteristics in clinical practice.[8] Radiomics hypothesizes that certain features may reflect biological heterogeneity of the microenvironment in a tumor and consequently provides information about specific tissue characteristics.[9]

Although many quantitative imaging features have been developed in different fields, feature extraction from medical imaging (*i.e*. radiomics) was introduced in 2012 and expectations have been high ever since.[10] Radiomics is believed to have the potential to improve knowledge of tumor biology and, combined with clinical data and other biomarkers, guide clinical management decisions.[5] Imaging features would serve as non-invasive biomarkers for tumor characterization, prognostic stratification and response prediction, thereby contributing to precision medicine. However, especially in nuclear medicine, conflicting results are obtained in different studies and association of different radiomic features with patient outcome is still rather preliminary. Difficulties that researchers and clinicians are facing when developing and validating these potential image-derived biomarkers are mainly related to the small numbers of patients per study, ultimately limiting the possibility to develop and explore robust imaging features that are unequivocally related to relevant tumor characteristics and patient outcome. While individual studies show promising results, it is still challenging to combine, compare and reproduce results of different studies due to little to no standardization in the different steps of the radiomic pipeline. This review investigates the promises of radiomics in precision medicine, the rise of artificial intelligence (AI) techniques in this field, harmonization strategies, challenges that need to be overcome to achieve clinical translation and the problems that have to be solved to be able to move towards validated and clinically accepted imaging biomarkers.

Radiomics

Rationale

Before it was found to be useful to support clinical management decisions in precision medicine, automated image analysis found its way in radiology and nuclear medicine to deal with capacity problems. In the last two decades, the number of diagnostic imaging procedures has increased rapidly.[11, 12] This is caused by improvements of technology (faster hardware and state-of-the-art scanning protocols and post-processing), resulting in lower radiation burden; a wider availability of scanners, lowering costs; and an increased demand of scans by referring physicians. However, the number of trained readers has not increased proportionately. This leads to excessive workloads, which increases the risk of errors.[13] Automated image analysis and other components using AI have been investigated to reduce the workload and resulting errors, and may facilitate in tasks like lesion or disease detection, classification and diagnosis, segmentation, quantification, patient planning and physician order entry.[6, 12] Within the field of nuclear medicine, compared to other fields of medical imaging, the number of procedures has increased less rapidly.[14] Nevertheless, also in this field, workloads have been increasing due to a demand for more advanced and time-consuming image analyses, like PET response criteria in solid tumors (PERCIST)[15] and personalized dosimetry. Additionally, the introduction of total-body PET is, as a result of a lower radiation dose and a faster acquisition time, expected to generate higher throughput and eventually an acceptable cost-to-benefit ratio, probably leading to an increase in PET imaging,[16] thereby increasing the need for automated image analysis.

Qualitative assessment of a medical image by a radiologist or nuclear medicine physician is based on his or her training and experience and is therefore rather subjective and sensitive to recent experiences. When assessing a medical image, clinicians need to interpret many independent variables at the same time, causing high conceptual complexity. Interestingly, Halford *et al.* found that prob-

lems defined with four variables are the limit of the human information processing capacity.[17] The subjectivity is further increased by inter- and intra-observer variability[18] and sensitivity to mental and physical fatigue.[19] AI algorithms can consider a large number of complex quantitative variables together, while being consistent, fast, tireless and efficient.[6] Thereby, AI bears the potential to reduce variation in clinical practice, improve efficiency and prevent avoidable medical errors.[20]

The automated analysis of medical images already started in the 1970s with computer-aided detection or diagnosis (CAD). These algorithms used a limited set of image features and machine learning classifiers for localization of lesions and distinction between benign and malignant lesions. CAD systems have, for instance, been adopted for the detection and diagnosis of lung nodules on chest radiographs and CT,[21] and for the detection of breast cancer on mammograms.[22] Radiomics differs from CAD on two facets: the number of image features has increased from 8-20 in CAD to a few hundred or thousand for radiomics and while CAD mostly focuses on the detection of lesions, radiomics assesses the association of radiomic features with biological and clinical endpoints, resulting in prognostic and predictive models.[23] At the moment, the field of radiomics can be divided into two areas: handcrafted radiomics and deep learning radiomics.[24] Both principles are illustrated in Figure 1.

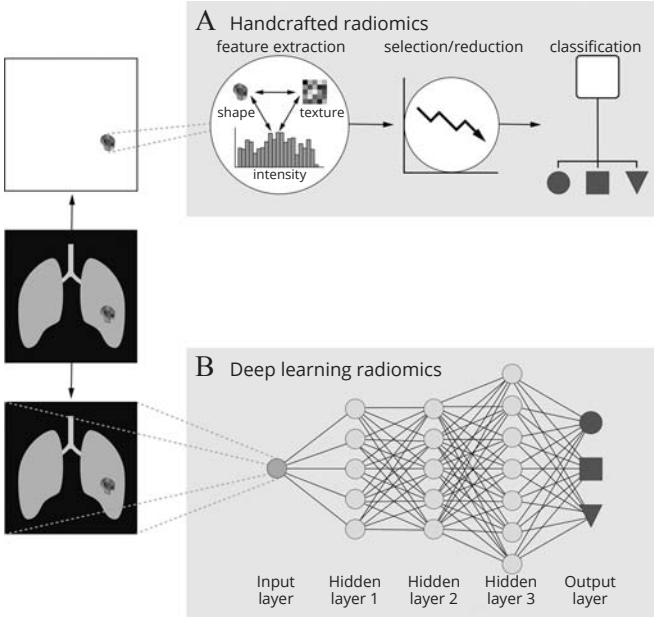

Figure 1.—Handcrafted and deep learning radiomic pipeline. A) In the handcrafted pipeline predefined features are extracted from a manually or (semi-) automatically defined volume of interest (VOI). Feature selection or dimension reduction is performed and these features are consecutively introduced in a statistical or machine learning model. B) Deep learning radiomics does not require VOI delineation, but processes the images in their raw form. The deep learning architecture consists of several hidden layers including convolutional and pooling layers, that extract increasingly complex features and perform feature selection and classification (courtesy of Gerrit Kracht).

Handcrafted radiomics

The traditional, handcrafted, approach consists of image acquisition and reconstruction, possible image post-processing, volume of interest (VOI) segmentation, predefined feature extraction, signature development and validation on one or several datasets.[25] Therefore, large numbers of predefined handcrafted or engineered features are extracted from the VOI of the medical images.[26] These features or selections of features are analyzed in statistical analysis or machine learning models that are trained for patient classification.

Three types of radiomic features, quantitatively describing shape, intensity and texture, are extracted. Shape features, also known as geometric or morphological features, describe the size and outline of a lesion (*i.e.* VOI), for example whether it is a perfect sphere or it is flattened. Intensity features describe the amount of tracer uptake: this class includes the commonly used SUV features such as maximum, peak or mean SUV, but also metrics like the range and skewness of the voxel values determined from histogram statistics. While intensity features express the amount of tracer uptake, they do not capture spatial tracer uptake heterogeneity. Texture features interpret relations in voxel values of neighboring voxels, thereby illustrating spatial (uptake) heterogeneity. Examples of texture features classes are grey level co-occurrence matrix features, expressing how combinations of voxel values of neighboring voxels are distributed, and grey level run length matrix features, expressing the length of a consecutive sequence of voxels with the same grey level.[27] Since there is no absolute definition of heterogeneity, neither from a physical, nor from a biological perspective, many features are extracted, all describing a specific form of heterogeneity.

Harmonization of radiomic features

Especially in the early years of radiomics, promises were high. The realization that medical images contained much more information than could be assessed visually was revolutionary and radiomics was expected to provide insights in disease processes and contribute to medical decision making on a large scale.[10, 28] However, it turned out that radiomic features were sensitive to all kinds of technical

factors. These variations should be reduced in order to attribute differences in feature values to tumor biology instead of technical causes.[8] To date, limited reproducibility (external validation) is still one of the major limitations of radiomics, hampering its clinical translation.

Image acquisition and reconstruction are the first steps in the radiomic pipeline. The factors that could affect these steps encompass, among others, patient preparation (including the amount of injected activity, the time between activity injection and acquisition and radiopharmaceutical-specific factors such as fasting glucose and insulin levels for the most commonly used tracer 2-[^{18}F]fluoro-2-deoxy-D-glucose — [^{18}F]FDG), acquisition settings (*e.g.* scan duration per bed position), equipment characteristics (*e.g.* effective sensitivity of the scanner, and cross calibration between activity calibrator and scanner), reconstruction settings (*e.g.* attenuation and scatter correction algorithms, reconstruction algorithm, number of iterations and subsets, post-reconstruction filter, partial volume correction/resolution modelling and voxel size). All these factors should be performed under the same or at least highly similar conditions, *i.e.* standardized, so that differences in radiomic feature values can be attributed to differences in tumor biology instead of technical variation. In practice, inhomogeneity of imaging data is a problem, since radiomic analysis often consists of retrospective analysis of standard-of-care images and reanalysis of previously published cohorts, where scanners and scan protocols may vary widely between different manufacturers and medical centers. This inhomogeneity complicates absolute quantification of tracer uptake. For traditional quantification metrics used in PET imaging, i.e. the mean, peak and maximum SUV, the European Association of Nuclear Medicine (EANM) has established guidelines for acquisition and reconstruction,[29] which are shaped by an accreditation program set up by the EANM Research Limited (EARL). These guidelines are also used for radiomic analysis, and leads to a larger number of reliable, repeatable, and reproducible features when the most recent EARL accreditation protocol is followed (EARL 2.0).[30]

Another harmonization strategy was proposed by Orlhac *et al.* to perform post-reconstruction harmonization for multicenter radiomic studies.[31] This method intends to standardize features obtained using different image protocols by means of preserving patient-specific effects, while removing center effects (batch effects). This approach, also known as ComBat harmonization, was first described in 2007 to address the batch effect in the field of genomics and was found suitable for radiomics as well.

Image processing factors also influence the reproducibility of image biomarkers. These factors include SUV normalization, image interpolation method, voxel harmonization, image discretization method, image discretization levels and added noise sensitivity,[32] but are beyond the scope of this review.

Image segmentation or volume of interest (VOI) delineation, including manual and (semi)automatic approaches, is also an important source of variation. The choice of segmentation method is often a compromise between accuracy and reproducibility.[33] Manual delineation encounters large inter- and intraobserver variability. Semiautomatic approaches are preferred, since they are more reproducible. Still, the influence of delineation methods on the predictive value of the radiomic features remains unclear.[33]

Furthermore, the *feature extraction* itself was a major impediment for the reproducibility, validation and clinical translation of radiomics as a result of a lack of standardization of (both implementation and extraction of) radiomic features. In the early years of radiomic research, large variation in feature values and feature definitions was seen between different radiomic software packages.[34] In response, the Image Biomarker Standardisation Initiative (IBSI) provided common nomenclature and (mathematical) definitions for image biomarkers, benchmarks for image processing and feature extraction, and reporting guidelines.[27] Iteratively, nineteen software implementations that started with large initial differences, managed to converge to common reference values for radiomic features, first using a simulated phantom (initial strong or better consensus: 7% of features, final strong or better consensus: 95%), later on CT images (initial strong or better consensus: 3%, final strong or better consensus: 91%).[34]

In a search for robust radiomic features that show minimal variations at different time points, under different conditions and with different feature definitions, repeatability (same subject, same conditions) and reproducibility (same subject, different scanners) studies have been performed to assess the effect of all above-mentioned technical factors on the predictive value of the radiomic features.[26, 35-37] In general, intensity features (especially entropy) are assumed more reproducible than shape and texture features.[36]

Feature selection

As the -omics suffix, also used in terms as genomics, metabolomics and proteomics, suggests, radiomics comprises large amounts of features. In most studies, several hundreds of radiomic features are exploratively extracted from medical images. The term explorative indicates that,

for most features, there is no biological rationale that the feature represents certain tissue characteristics. Therefore, many features are investigated, assuming that some features show association with underlying biology. This explorative nature makes radiomics data-driven research instead of hypothesis-driven research, making it difficult to maintain statistical power of individual studies. In this regard, challenges in radiomics differ from hypothesis-driven biomarker research.[38]

The number of investigated features is generally large compared to the number of patients, scans or events in an average study. This introduces the curse of dimensionality, a problem that arises when data points (patients or scans) are analyzed in high-dimensional spaces (*i.e*. the hundreds of radiomic features per scan).[4] The data space increases exponentially with the number of radiomic features, while the number of data points stays the same. Therefore, overfitting occurs and the generalization performance of the statistical model is negatively impacted.[39] In this case, the model is specifically tuned to the training dataset and merely reflects its noise and random fluctuations, whereby it cannot be applied to other datasets. In standard biomarker research, only a few parameters would be tested for their association with outcome measures, while in radiomics, several tens or hundreds of features are tested simultaneously. This leads to multiple testing and increases the false discovery rate, since it could be expected that some of these features can show and association with clinical or biological parameters only based on chance.[40]

To reduce the problem of data dimensionality and multiple testing effects, a smaller number of features should be brought into a model. Therefore, a selection of features has to be made before they are used to create a model predicting specific patient or disease characteristics. Supervised and unsupervised feature selection methods are applied.[41, 42] Supervised approaches use data labelled with outcome and a selection of features is based on the discriminative value of outcomes. Only the features with the best association with outcome are selected and introduced in the regression model. Supervised feature selection is prone to overfitting and ignores the interaction of features among themselves (multicollinearity),[42] while, especially in radiomics, multicollinearity is an important challenge due to its explorative nature: many radiomic features show high correlations with other features. The intensity feature class contains many features based on the SUV: maximum SUV, peak SUV, mean SUV, etcetera. In this example, there is a strong correlation between the different SUV-based features, given that a high mean SUV results generally in high values for the median SUV. Also, between different feature classes, multicollinearity might occur, which is less obvious than for features from the same class, *e.g.* Tixier *et al.* found that grey level run length matrix and grey level size zone matrix features were strongly correlated ($r>0.9$).[43] Multicollinearity is a problem when fitting a regression model and can be seen in many early radiomic studies. Unsupervised feature selection, also known as dimensionality reduction, restrains multicollinearity. Unsupervised methods use unlabeled data (no outcome) to create clusters of features showing similar patterns, for instance by calculating the distance between them in high dimensional space, thereby maintaining interactions between features.[42] Clustering bears the potential to identify previously unknown associations between features and thereby provide new insights in diseases. An example of unsupervised dimension reduction is redundancy filtering, removing highly correlating features based on a threshold using a correlation matrix.[44]

Model building and validation

Subsequently, the (selected) features serve as an input for a predictive model. There are two categories of predictive models: classification models that are able to predict discrete or categorical variables (*e.g*. benign *vs*. malignant) and regression, where a model is fitted to continuous variables (*e.g*. survival time). Modelling is a supervised learning task where a hypothesis (function) is fitted to the input data, and compared with the desired output value (clinical characteristics, tumor phenotype, etc.).[45] A predictive model could be created using AI that allows computers the ability to learn without being explicitly programmed. Examples of these so-called machine learning algorithms are logistic regression, support vector machines, decision trees and neural networks (deep learning).[45] Currently, random forest is a popular supervised learning algorithm in medicine, since it is based on decision trees and hence resembles human "if-then" reasoning.[23]

Although there are many machine learning methods for classification and regression, there is no consensus yet on which one to prefer for radiomic analysis. Parmar *et al.* tested performance variability for several classification methods, feature selection methods and different numbers of features on two datasets of non-small cell lung carcinoma, demonstrating that the choice of classification method is the most dominant source of performance variability and, by that, being a crucial step towards stable radiomic biomarkers.[4] They showed that random forest, bootstrap aggregating and Naïve Bayes algorithms showed relative-

ly high performances. Deist *et al.* investigated the performance of six classifiers on twelve clinical (non-radiomic) datasets, where random forest showed the best discrimination in six of 12 datasets and elastic net logistic regression the best in four of 12 datasets, but there was no single best classifier for all datasets.[46] These studies indicate that there is no optimal algorithm for radiomic classification. The performance of a model might improve when different algorithms are combined, *e.g*. by using different machine learning algorithms for different random subsamples of the data, that are ultimately integrated to an overall decision (bagging). Model improvement could also be achieved by feeding subsequent algorithms data that previous algorithms found difficult to classify (boosting).[45]

After the model is optimized to the training dataset, it can be utilized to predict outcomes for unseen observations (new patients). The performance of the model should be validated on a new, independent, dataset to assess whether the model is predictive for the target population as a whole or only for a specific subset of patients, *e.g.* patients with a specific demographic profile or patients who underwent a scan on a specific scanner. Validation could be performed using a dataset from the same institute, or preferably using one or more datasets from different institutes (external validation).[47]

Deep learning radiomics

More recently, deep learning methods in medical imaging gained interest. Deep learning is a form of machine learning (which is a form of AI), defined for the first time in 1943.[48] It is based on a statistical neural network structure, inspired by the human brain.[24] Deep learning has the ability to learn non linear discriminative features directly from raw data.[49] A convolutional neural network (CNN) is the most common deep learning architecture used in medical imaging. The CNN architecture comprises many convolutional and pooling layers with nonlinear activation functions that map image inputs to outcome measures. Abstract imaging features are learned during the optimization (training) process.[49] The first successful and commercialized application of deep learning was already established in 1998, when LeCun *et al.* developed a CNN for handwritten character recognition,[50] but only in 2012 the use of CNNs took a leap.[51] That year, at the ImageNet competition a CNN (AlexNet, developed by Krizhevsky *et al.*)[52] outperformed all machine learning approaches with half the error rate in an image recognition contest consisting of 1000 object classes.[49] This turn of events was enforced by the development of new techniques for efficient training, availability of more training data and an increase in computing power with graphic processing units (GPU).[51] Deep learning techniques also quickly gained interest in medical imaging for diagnosis support, image quality optimization, image reconstruction, segmentation, data visualization, response assessment and report generation, and in healthcare in general for drug discovery, remote patient monitoring, medical diagnostics, risk management, wearables and hospital management.[24, 53-55] In recent years, the number of publications of deep learning in medical imaging has increased exponentially.[56]

Deep learning radiomic methods differ from the handcrafted pipeline in the way that it is not necessary to define the VOI and to extract handcrafted features. Deep learning processes data in their natural raw form, contrary to machine learning approaches (like in handcrafted radiomics), where careful engineering and considerable domain expertise were needed to build suitable features for the model.[49] The data-driven approach of deep learning is favored over the feature-driven approach, since the handcrafted features set is finite and not necessarily represents the optimal quantification approach for the discrimination task.[24] Deep learning networks, on the other hand, provide more abstract and unconstrained feature definitions driven by data, resulting in a more informative and generalizable model.[24] In addition, the feature extraction and classification are partially or completely connected through the global optimization process. However, a CNN usually needs a large labelled dataset for (supervised) classification.[57] It is previously reported that CNNs do not necessarily result in significantly higher accuracies than handcrafted radiomics in combination with classification machine learning, but they are more user-friendly, since they require less data-handling and are less labor intensive than VOI delineation.[58]

While the diagnostic performance of deep learning models is reported to be similar to that of handcrafted radiomic models[58] and human observers,[59] deep learning methods currently face some difficulties. Similar to handcrafted radiomics, deep learning radiomics suffers from a lack of standardization. There are no reporting standards, which limits reliable interpretation and replication of the studies.[59] Also, (external) validation is limited and not all studies used the same samples to compare performance of the algorithm and the human observers.[59]

A disadvantage of radiomics in general, but of the deep learning approach in particular, is that clinicians struggle with inadequate comprehension of the capabilities of algorithms. The large number of parameters (often millions)

in a deep learning model are not directly interpretable. In contrast to qualitative assessments by physicians, quantitative handcrafted or deep learning (radiomic) features lack an easy and intuitive interpretation.[60] In this sense, algorithms can be regarded as black-boxes; even if we understand the mathematical definitions of the features, these features do often not represent knowledge we can interpret in the domain of biology or radiology.[61] Radiogenomic approaches investigate relationships between imaging phenotypes and genomics,[62] thereby contributing to the biological or clinical context of imaging features.

Also, comprehensive solutions, focusing on linking the features to underlying biological mechanisms, contribute to the confidence and interpretability of the algorithms. Therefore, especially in medicine, explainable AI (XAI) is gaining terrain to interpret the data in the context of a specific application and to retrace the results on demand.[61] Being able to reproduce and comprehend the knowledge extraction process is crucial, since causality is often necessary for clinical decision making and might provide new insight in disease processes.[63] This is an interesting trade-off, since the best models in terms of accuracy might be the least transparent and vice versa.[64] XAI remains a technically challenging field and complete solutions have not been found yet.[61] XAI can be divided into post-hoc approaches, explaining predictions in terms of what is readily interpretable, and ante-hoc approaches, that represent explainability by design.[63] Paul *et al.* presented an example of a *post-hoc* XAI approach, where some deep learning features were successfully related to semantic features, i.e. features that are commonly used by a radiologist, in patients with lung cancer.[65] An example of an ante-hoc, or intelligible, approach was introduced by Caruana *et al.* and presented an accurate model for the prediction of hospital readmission based on 4,000 variables, while still being able to explain predictions for individual patients.[66]

Applications

Automated image analysis using radiomics is performed on CT, magnetic resonance imaging (MRI) and PET images and is, more recently, also investigated for single-photon emission tomography (SPECT) and ultrasound (US).[67, 68] Whereas CT, US and, to some extent, MRI, are morphology-based imaging modalities, PET and SPECT images represent biological and molecular characteristics using radiopharmaceuticals that target specific receptors or pathways. In these molecular imaging techniques, biological processes are quantitatively expressed by the (spatial distribution of) radiotracer uptake and it is hypothesized that these can be quantified using radiomics.[69] In PET imaging, quantitative analysis of tracer uptake has been performed for several decades, for instance for response monitoring, by means of SUV, which is the radiopharmaceutical activity concentration (Bq/mL) normalized for injected activity per unit measure of distribution volume of the radiopharmaceutical in a patient (*e.g.* body mass). For mass-based measures of volume of distribution the unit of the SUV therefore is g/mL.[70] However, the commonly-used SUV metrics do only partially express heterogeneity in tracer uptake.[71] Radiomic shape, intensity and texture features are computed to capture tracer uptake heterogeneity and/or other tumor characteristics.

Between 2012 and the beginning of 2019, an exponential growth of articles with the term radiomics was observed with a total of about 1000 papers, whereof approximately 27% concerned PET or PET/CT.[72] CT remains the most frequently used modality for radiomic analysis, since it has a high spatial resolution and it is relatively easy to acquire large datasets, as a result of the number of patients for whom CT is indicated and the speed of imaging. [^{18}F]FDG is the most widely applied radiopharmaceutical in clinical practice, since it is an important tool in detection and staging of cancer and active inflammation, and also constitutes the majority of PET radiomic studies. Several studies showed the prognostic or predictive abilities of handcrafted radiomic features derived from [^{18}F]FDG-PET in different tumor types.[73-78] These studies illustrate the discriminating capabilities of [^{18}F]FDG-PET radiomic features for the stratification of histology, tumor grades or stages and clinical outcome. As an example, Li *et al.* created a radiomic signature, an outcome-specific combination of a few radiomic features, that could discriminate between mutant and wild-type epidermal growth factor receptor non-small cell lung carcinoma (accuracy: 80.8%, sensitivity: 82.6%, specificity: 78.3%).[78] When combined with clinical characteristics, the predictive performance of the radiomic model even increased (accuracy: 82.7%, sensitivity: 82.1%, specificity: 82.3%).

The field of handcrafted radiomics is evolving and new applications are being developed constantly. As an example, delta radiomics, the monitoring of radiomic features over two or more time points, was identified as indicator for treatment response.[79, 80] Furthermore, more recent radiomic studies extract features from the PET as well as from the low-dose CT component of the hybrid images[81] in order to gain the most information from integrated PET/CT studies. Also, radiomic features derived from other tissues than the primary tumor like peritumoral radiomics in CT,[82]

as well as radiomics extracted from the bone marrow[83] and metastatic lymph nodes[84] in PET, were promising predictors of clinical outcome. Additionally, radiomic features derived from parametric PET images were assessed for additional information compared to radiomic features derived from static images, but were found to be strongly correlated between both imaging datasets.[85] Lastly, handcrafted radiomics has, to some extent, also been evaluated for tracers other than [^{18}F]FDG. Radiomic analysis has been applied to [^{18}F]fluorothymidine ([^{18}F]FLT) PET[86] and [^{68}Ga] prostate-specific membrane antigen-11 ([^{68}Ga]Ga-PSMA-11) PET.[87] Also, radiomics holds potential for more accurate characterization of different aspects of tumor biology, for example by gaining insight in tumor hypoxia using PET-tracers like [^{18}F]fluoromisonidazole ([^{18}F]FMISO) and [^{18}F]fluoroazomycin arabinoside ([^{18}F]FAZA).[88]

Non-oncology applications of handcrafted radiomics can mostly be found in SPECT. Radiomics was successfully applied in cardiology for the prediction of coronary artery calcification in myocardial perfusion SPECT with [^{99m}Tc]sestamibi.[67] In neurology, radiomics were exploited for outcome prediction in Parkinson's disease on dopamine transporter (DaT) [^{123}I]ioflupane SPECT.[89]

While handcrafted radiomics is extensively studied, the use of deep learning radiomics is limited in nuclear imaging. In [^{18}F]FDG-PET, there are some examples of studies comparing CNNs with handcrafted radiomic analyses. Ypsilantis *et al.* showed how a CNN outperformed the handcrafted radiomic approach for the prediction of response to neoadjuvant chemotherapy in oesophageal cancer.[90] Wang *et al.* found similar performances for the CNN, handcrafted radiomics, and nuclear medicine physicians for the classification of mediastinal lymph nodes in non-small lung carcinoma.[91] In this case a CNN is preferred over the handcrafted radiomic analyses, since it is more convenient and suffers less from selection bias. CNNs were also used for to diagnose thyroid diseases on scintigraphy[92] and for the diagnosis of Parkinson's disease on dopamine transporter (DaT) in SPECT.[93] Compared with nuclear medicine, deep learning radiomics is more established in radiology; there are several examples of successful applications of deep learning radiomics in CT and MRI, demonstrating comparable performance to human readers.[94]

Future perspectives

Hospitals accumulate large amounts of patient data: not only radiological images, but also clinical and demographic characteristics, laboratory results, pathological and immunohistochemical images, genomic data, etc. As a result of technological advancements, such as the development of CNN and natural language processing,[95] these large datasets could be used for extensive predictive modelling, thereby bearing the potential to optimize clinical decision making and advance precision medicine. Several initiatives have been established to combine big data of different sources. Multiomics initiatives combine molecular profiling of tumors using genomics, epigenomics, transcriptomics and proteomics.[96] Radiogenomics focuses on relationship between imaging phenotypes and genomics, hypothesising that imaging features will be more informative when interpreted by taking into account their genetic context.[62] Holomics refers to the combination of all available patient data for predictive modelling,[97] thereby integrating radiomic data with previously mentioned types of data. Integration of all the data in a data warehouse could lower current restraints in individual patient management imposed by trial designs and fragmented evidence.

The effort required for the curation of data is a major roadblock to holomics and predictive modelling and makes it a labor-intensive process, since curated predictor variables should often be manually extracted from the electronic health records (EHR) or imaging archives.[98] The process of data extraction usually focuses on a limited number of variables, as described in medical ethics protocols, whereas the vast majority of the information in the patient record is disregarded.[98] However, ethical use of large amounts of data already stored in data repositories is a complex issue, particularly given that there is usually no standard informed consent of the patient. Furthermore, data are unstructured and suffer from poor standardization, *i.e.* events are not registered in the same way and at the same point in time or not at all; data lack information about treatment outcomes; and data are not centrally collected in the EHR, but are also stored in other programs like picture archiving and communication systems.[60] The latter requires better integration of EHRs and other information systems, and could be established by standardized communication formats like Fast Healthcare Interoperability Resources (FHIR). Rajkomar *et al.* demonstrated that deep learning models using entire raw EHR records of 261,221 adult patients based on FHIR format outperformed traditional, clinically-used predictive models for multiple medical events like in-hospital mortality and prolonged length of admission.[98]

To reach the full potential of our data warehouses, good data management and standardization is of utmost importance to forward discoveries and innovation. We should

start to collect patient data systematically on a dedicated platform. Therefore, guidelines should be established on which information is required for a decision or medical intervention. Also, it should be pointed out when data are missing or of poor quality, so that it could be collected or corrected accordingly. Subsequently, additional data, like radiomic features, could be extracted from existing data, emphasizing the requirement of automated processing pipelines. Data should be collected according to the FAIR Guiding Principles, a framework contributing to scientific data management and stewardship, based upon Findable, Accessible, Interoperable, and Reusable data.[99] However, good data management and stewardship are not intrinsic goals, but rather pre-conditions supporting knowledge discovery and innovation.[99] Starting with this prerequisite, we need well-trained bioinformaticians and computational biologists to analyze the large amounts of data. Only then, the thousands or even millions of data points could be used for predictive modelling and have the potential to provide new knowledge on disease characteristics and patient management.

A major impediment of radiomic as well as holomic research, especially when applying deep learning, is the collection of sufficient amounts of data with sufficient variation to answer research questions on specific diseases in precision medicine. Especially within nuclear medicine, institutes struggle to collect homogeneous patient cohorts for radiomic studies, causing the development of a radiomic signature to be a challenge. Nevertheless, the development of the signature is only the beginning, since many hurdles will follow validating the signature and bringing it in clinical practice as image-derived biomarkers. Validation is currently a large roadblock for the clinical translation of radiomic features. Due to the limited size of their datasets, most radiomic research focuses on the development of radiomic signatures. However, even more patients are needed for internal or (preferably) external validation, which both handcrafted and deep learning radiomic research lack.[59, 88] It should be noted that it is also of great value to perform external validation of a radiomic signature from another center in a new patient population.[47]

In addition, the field of radiomics experiences a publication bias, addressed by Buvat *et al.* as 'the dark side of radiomics'.[100] Their study showed that only 6% of radiomic studies concluded negative results, while these results might even have a larger impact and might challenge existing paradigms. Undoubtedly, positive results are more rewarding, increase journal influence metrics and have a greater clinical impact, but from an ethical perspective, negative results need to be published as well. Buvat *et al.* propose stricter editorial standards, focusing on among other things well-designed (statistical) methodologies and biologically relevant research questions. In this way, not publishing false-positive results caused by overanalysis of the data, could leave room for well-designed radiomic studies with true negative results. Furthermore, results could be published in journal sections dedicated to negative results or public repositories. Altogether, this would prevent unnecessary and expensive repeats of studies that were already conducted in different institutes.

Distributed learning might facilitate the investigation of stable and clinically relevant biomarkers. Individual institutes encounter problems collecting satisfactory datasets in terms of size and diversity to create a model. The vast majority of the algorithms is built on retrospective data and its performance is likely to be worse on data from other institutes with different characteristics of data and patient population. Simply combining data from different institutes is not sufficient to create a generalizable model, since these isolated data collection efforts suffer from inhomogeneity of data caused by technological, human and organizational factors.[101] Also, data sharing is hampered by legal and ethical concerns.[101] To overcome these issues, Deist *et al.* developed a distributed learning algorithm that enables data sharing for machine learning without identifiable patient data leaving an institute, so that the institute remains in control of their data to preserve data privacy.[101] This is executed by a central server, located outside the institutions, containing the master algorithm, and for each institution a local algorithm, situated at their server. The master and local algorithms interact. With every patient, the local model of the institution updates, thereby also updating the master algorithm. The master performs some checks and afterwards updates the local algorithms at individual institutions. In this manner, local algorithms differ slightly from the master algorithm, in such a way that the models suit the patient population of an institution.

The potential of radiomics and AI has not remained unnoticed: the number of start-ups focused on AI in medical imaging is rapidly growing. In 2018, McKinsey identified 32 start-ups in this field,[102] albeit in 2020, the number of companies tripled to 113, with total investments of $1.17 billion.[13] It is noteworthy that the share of nuclear medicine in this matter is less clear, since in 2018 only 3% (1/32) of these start-ups was focused on nuclear medicine.[102]

Despite industry being a booster of AI, to date, AI has not yet been adopted into clinical practice on a large scale.

Physicians are reluctant to embrace AI in clinical workflow. In the early years of AI, this reluctance was mostly caused by fear of the unknown and fear to be replaced by AI. It is unlikely that radiologists and nuclear medicine physicians will be completely replaced by AI, even in the future. AI is not capable of performing advanced analysis in uncommon diseases, since these tasks require human reasoning and the number of scans is too small to fit a model. All the same, a reallocation of tasks from clinicians to AI will definitely take place. Tasks that occur frequently and lead to high workloads or double readings are likely to be the first to be replaced by AI. AI could identify and mark regions on for example chest radiographs, CT or mammographs that require specific attention. This enables the radiologist to focus on more advanced tasks. The efficacy of radiologists might even grow, when they embrace new technologies.[24]

Another roadblock to clinical use is the lack of scientific evidence of the AI algorithms, when developing an algorithm or when implementing an elsewhere developed algorithm. Peer-reviewed randomized controlled trials (RCT) are the gold standard for evidence generation and would contribute to trust and adoption of AI within the medical community.[20] Nevertheless, the number of RCTs validating AI algorithms is limited.[20] Theoretically, a specific algorithm could be compared to the standard of care (*e.g.* a biopsy or mutational status) with a clinical outcome as trial endpoint in a superiority or non-inferiority RCT. However, a RCT might not be the optimal method to prove effectiveness of AI algorithms in precision medicine,[103] since the algorithms are not static entities, but they are constantly evolving, based on new data. Primarily, a RCT takes into account two interventions (discrete) and includes a small number of variables that describe the context. AI algorithms, contrarily, encompass a large dataspace with thousands of variables and result in combinations of different treatments, at different doses (continuous). This would introduce an infinite number of possible RCTs in a specific care path, testing the effect of the individual variables. By conducting one specific RCT, we assume that the relationship between context and actions is only as complex as the factors addressed in the RCT, thereby negating treatment heterogeneity, the fundament of precision medicine.[103] In this sense, all patients are unique and, consequently, large enough homogeneous samples will never be available to come to deterministic decisions using RCTs.[103]

For this reason, we should look for alternative methods to evaluate the effectiveness of AI in precision medicine. Opportunities can be seen in allocation schemes, where probabilities of receiving different treatments are increased, proportional to the expected effectiveness of the treatment, conceptually proposed by Thomson *et al.* in 1933.[104] In this way, learning and earning are balanced, resulting in the best choice for a specific patient given all the available information and knowledge.[103] Further information on these schemes are beyond the scope of this review but can be found in Kaptein's review.[103] While allocation schemes seem promising, it should be noted that scientific evidence is a delicate topic, with the RCT the undisputable best measure of evidence. The healthcare sector is reluctant to accept other forms of evidence, which might hamper the use of AI for precision medicine in clinical practice.

Conclusions

In the next decade, radiomics should redeem its promises in the field of radiology and nuclear medicine. The translational gap should be overcome and it should be demonstrated that radiomic features could safely replace existing biomarkers. Developments are ongoing, including standardization and harmonization initiatives, integration with other datatypes in holomic approaches and initiatives for data or model exchanges. All these improvements together might lead to validated and explainable AI models capable of medical decision making that goes beyond the performance of a physician, thereby improving efficiency, reducing variation in clinical practice and reducing the chance of medical errors. Along these lines, radiomics and holomics can contribute to precision medicine by providing the right treatment to the right patient, with the right dose, at the right time.

References

1. Ashley EA. Towards precision medicine. Nat Rev Genet 2016;17:507–22.

2. Collins LG, Haines C, Perkel R, Enck RE. Lung cancer: diagnosis and management. Am Fam Physician 2007;75:56–63.

3. Travis WD, Brambilla E, Noguchi M, Nicholson AG, Geisinger K, Yatabe Y, *et al.* Diagnosis of lung cancer in small biopsies and cytology: implications of the 2011 International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society classification. Arch Pathol Lab Med 2013;137:668–84.

4. Parmar C, Grossmann P, Bussink J, Lambin P, Aerts HJ. Machine Learning methods for Quantitative Radiomic Biomarkers. Sci Rep 2015;5:13087.

5. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. Radiology 2016;278:563–77.

6. McBee MP, Awan OA, Colucci AT, Ghobadi CW, Kadom N, Kansagra AP, *et al.* Deep Learning in Radiology. Acad Radiol 2018;25:1472–80.

7. Parekh VS, Jacobs MA. Deep learning and radiomics in precision medicine. Expert Rev Precis Med Drug Dev 2019;4:59–72.

8. Limkin EJ, Sun R, Dercle L, Zacharaki EI, Robert C, Reuzé S, *et al.* Promises and challenges for the implementation of computational medical imaging (radiomics) in oncology. Ann Oncol 2017;28:1191–206.

9. Junttila MR, de Sauvage FJ. Influence of tumour micro-environment heterogeneity on therapeutic response. Nature 2013;501:346–54.

10. 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.

11. Smith-Bindman R, Miglioretti DL, Larson EB. Rising use of diagnostic medical imaging in a large integrated health system. Health Aff (Millwood) 2008;27:1491–502.

12. Boland GW, Guimaraes AS, Mueller PR. The radiologist's conundrum: benefits and costs of increasing CT capacity and utilization. Eur Radiol 2009;19:9–11, discussion 12.

13. Alexander A, Jiang A, Ferreira C, Zurkiya D. An Intelligent Future for Medical Imaging: A Market Outlook on Artificial Intelligence for Medical Imaging. J Am Coll Radiol 2020;17(1 Pt B):165–70.

14. Roobol L, van der Reijden A, de Waard-Schalk I, Bijwaard H. Productie en gebruik van medische radio-isotopen in Nederland: Huidige situatie en toekomstverkenning: Rijksinstituut voor Volksgezondheid en Milieu RIVM; 2017 [Internet]. Available from: https://www.rivm.nl/bibliotheek/rapporten/2017-0063.pdf [cited 2020, May 19].

15. Wahl RL, Jacene H, Kasamon Y, Lodge MA. From RECIST to PERCIST: evolving Considerations for PET response criteria in solid tumors. J Nucl Med 2009;50(Suppl 1):122S–50S.

16. Cherry SR, Jones T, Karp JS, Qi J, Moses WW, Badawi RD. Total-Body PET: Maximizing Sensitivity to Create New Opportunities for Clinical Research and Patient Care. J Nucl Med 2018;59:3–12.

17. Halford GS, Baker R, McCredden JE, Bain JD. How many variables can humans process? Psychol Sci 2005;16:70–6.

18. McErlean A, Panicek DM, Zabor EC, Moskowitz CS, Bitar R, Motzer RJ, *et al.* Intra- and interobserver variability in CT measurements in oncology. Radiology 2013;269:451–9.

19. Elliott J, Williamson K. The radiology impact of healthcare errors during shift work. Radiography (Lond) 2019;S1078-8174(19)30272-X. [Epub ahead of print]

20. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. BMC Med 2019;17:195.

21. Abe H, MacMahon H, Engelmann R, Li Q, Shiraishi J, Katsuragawa S, *et al.* Computer-aided diagnosis in chest radiography: results of large-scale observer tests at the 1996-2001 RSNA scientific assemblies. Radiographics 2003;23:255–65.

22. Gromet M. Comparison of computer-aided detection to double reading of screening mammograms: review of 231,221 mammograms. AJR Am J Roentgenol 2008;190:854–9.

23. Avanzo M, Stancanello J, El Naqa I. Beyond imaging: the promise of radiomics. Phys Med 2017;38:122–39.

24. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJ. Artificial intelligence in radiology. Nat Rev Cancer 2018;18:500–10.

25. Scrivener M, de Jong EE, van Timmeren JE, Pieters T, Ghaye B, Geets X. Radiomics applied to lung cancer: a review. Transl Cancer Res 2016;5:398–409.

26. van Velden FH, Kramer GM, Frings V, Nissen IA, Mulder ER, de Langen AJ, *et al.* Repeatability of Radiomic Features in Non-Small-Cell Lung Cancer [(18)F]FDG-PET/CT Studies: Impact of Reconstruction and Delineation. Mol Imaging Biol 2016;18:788–95.

27. Zwanenburg A, Leger S, Vallieres M, Lock S. Image biomarker standardisation initiative; 2019 [Internet]. Available from: https://arxiv.org/pdf/1612.07003.pdf [cited 2020, May 19].

28. El Naqa I, Grigsby P, Apte A, Kidd E, Donnelly E, Khullar D, *et al.* Exploring feature-based approaches in PET images for predicting cancer treatment outcomes. Pattern Recognit 2009;42:1162–71.

29. Boellaard R, Delgado-Bolton R, Oyen WJ, Giammarile F, Tatsch K, Eschner W, *et al.*; European Association of Nuclear Medicine (EANM). FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. Eur J Nucl Med Mol Imaging 2015;42:328–54.

30. Pfaehler E, van Sluis J, Merema BB, van Ooijen P, Berendsen RC, van Velden FH, *et al.* Experimental Multicenter and Multivendor Evaluation of the Performance of PET Radiomic Features Using 3-Dimensionally Printed Phantom Inserts. J Nucl Med 2020;61:469–76.

31. Orlhac F, Boughdad S, Philippe C, Stalla-Bourdillon H, Nioche C, Champion L, *et al.* A Postreconstruction Harmonization Method for Multicenter Radiomic Studies in PET. J Nucl Med 2018;59:1321–8.

32. Zwanenburg A. Radiomics in nuclear medicine: robustness, reproducibility, standardization, and how to avoid data analysis traps and replication crisis. Eur J Nucl Med Mol Imaging 2019;46:2638–55.

33. Cook GJ, Azad G, Owczarczyk K, Siddique M, Goh V. Challenges and Promises of PET Radiomics. Int J Radiat Oncol Biol Phys 2018;102:1083–9.

34. Zwanenburg A, Vallières M, Abdalah MA, Aerts HJ, Andrearczyk V, Apte A, *et al.* The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping. Radiology 2020;295:328–38.

35. Frings V, van Velden FH, Velasquez LM, Hayes W, van de Ven PM, Hoekstra OS, *et al.* Repeatability of metabolically active tumor volume measurements with FDG PET/CT in advanced gastrointestinal malignancies: a multicenter study. Radiology 2014;273:539–48.

36. Altazi BA, Zhang GG, Fernandez DC, Montejo ME, Hunt D, Werner J, *et al.* Reproducibility of F18-FDG PET radiomic features for different cervical tumor segmentation methods, gray-level discretization, and reconstruction algorithms. J Appl Clin Med Phys 2017;18:32–48.

37. Traverso A, Wee L, Dekker A, Gillies R. Repeatability and Reproducibility of Radiomic Features: A Systematic Review. Int J Radiat Oncol Biol Phys 2018;102:1143–58.

38. O'Connor JP, Aboagye EO, Adams JE, Aerts HJ, Barrington SF, Beer AJ, *et al.* Imaging biomarker roadmap for cancer studies. Nat Rev Clin Oncol 2017;14:169–86.

39. Clarke R, Ressom HW, Wang A, Xuan J, Liu MC, Gehan EA, *et al.* The properties of high-dimensional data spaces: implications for exploring gene and protein expression data. Nat Rev Cancer 2008;8:37–49.

40. Chalkidou A, O'Doherty MJ, Marsden PK. False Discovery Rates in PET and CT Studies with Texture Features: A Systematic Review. PLoS One 2015;10:e0124165.

41. Nie F, Xu D, Tsang IW, Zhang C. Flexible manifold embedding: a framework for semi-supervised and unsupervised dimension reduction. IEEE Trans Image Process 2010;19:1921–32.

42. Zhang Y, Oikonomou A, Wong A, Haider MA, Khalvati F. Radiomics-based Prognosis Analysis for Non-Small Cell Lung Cancer. Sci Rep 2017;7:46349.

43. Tixier F, Le Rest CC, Hatt M, Albarghach N, Pradier O, Metges JP, *et al.* Intratumor heterogeneity characterized by textural features on baseline 18F-FDG PET images predicts response to concomitant radiochemotherapy in esophageal cancer. J Nucl Med 2011;52:369–78.

44. Rizzo S, Botta F, Raimondi S, Origgi D, Fanciullo C, Morganti AG, *et al.* Radiomics: the facts and the challenges of image analysis. Eur Radiol Exp 2018;2:36.

45. Forghani R, Savadjiev P, Chatterjee A, Muthukrishnan N, Reinhold C, Forghani B. Radiomics and Artificial Intelligence for Biomarker and Prediction Model Development in Oncology. Comput Struct Biotechnol J 2019;17:995–1008.

46. Deist TM, Dankers FJ, Valdes G, Wijsman R, Hsu IC, Oberije C, *et al.* Machine learning algorithms for outcome prediction in (chemo)radiotherapy: an empirical comparison of classifiers. Med Phys 2018;45:3449–59.

47. Lambin P, Leijenaar RT, Deist TM, Peerlings J, de Jong EE, van Timmeren J, *et al.* Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol 2017;14:749–62.

48. Chakraverty S, Sahoo DM, Mahato NR. McCulloch–Pitts Neural Network Model. Concepts of Soft Computing: Fuzzy and ANN with Programming. Singapore: Springer Singapore; 2019. p. 167–73.

49. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521:436–44.

50. LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. Proc IEEE 1998;86:2278–324.

51. van Ginneken B. Fifty years of computer analysis in chest imaging: rule-based, machine learning, deep learning. Radiological Phys Technol 2017;10:23–32.

52. Krizhevsky A, Sutskever I, Hinton GE. ImageNet Classification with Deep Convolutional Neural Networks; 2017 [Internet]. Available from: https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf [cited 2020, May 19].

53. Visvikis D, Cheze Le Rest C, Jaouen V, Hatt M. Artificial intelligence, machine (deep) learning and radio(geno)mics: definitions and nuclear medicine imaging applications. Eur J Nucl Med Mol Imaging 2019;46:2630–7.

54. Shen D, Wu G, Suk HI. Deep Learning in Medical Image Analysis. Annu Rev Biomed Eng 2017;19:221–48.

55. 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.

56. Bernal J, Kushibar K, Asfaw DS, Valverde S, Oliver A, Martí R, *et al.* Deep convolutional neural networks for brain image analysis on magnetic resonance imaging: a review. Artif Intell Med 2019;95:64–81.

57. Razzak MI, Naz S, Zaib A. Deep Learning for Medical Image Processing: Overview, Challenges and the Future. In: Dey N, Ashour AS, Borra S, editors. Classification in BioApps: Automation of Decision Making. Cham: Springer International Publishing; 2018. p. 323–50.

58. Afshar P, Mohammadi A, Plataniotis KN, Oikonomou A, Benali H. From Handcrafted to Deep-Learning-Based Cancer Radiomics: challenges and Opportunities. IEEE Signal Process Mag 2019;36:132–60.

59. Liu X, Faes L, Kale AU, Wagner SK, Fu DJ, Bruynseels A, *et al.* A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. The Lancet Digital Health. 2019;1:e271–97.

60. Morin O, Vallières M, Jochems A, Woodruff HC, Valdes G, Braunstein SE, *et al.* A Deep Look Into the Future of Quantitative Imaging in Oncology: A Statement of Working Principles and Proposal for Change. Int J Radiat Oncol Biol Phys 2018;102:1074–82.

61. Holzinger A, Biemann C, Pattichis CS, Kell DB. What do we need to build explainable AI systems for the medical domain? 2017 [Internet]. Available from: https://arxiv.org/pdf/1712.09923.pdf [cited 2020, May 19].

62. Mazurowski MA. Radiogenomics: what it is and why it is important. J Am Coll Radiol 2015;12:862–6.

63. Holzinger A, Langs G, Denk H, Zatloukal K, Müller H. Causability and explainability of artificial intelligence in medicine. Wiley Interdiscip Rev Data Min Knowl Discov 2019;9:e1312.

64. Bologna G, Hayashi Y. Characterization of Symbolic Rules Embedded in Deep DIMLP Networks: A Challenge to Transparency of Deep Learning. J Artif Intell Soft Comput Res 2017;7:265.

65. Paul R, Schabath M, Balagurunathan Y, Liu Y, Li Q, Gillies R, *et al.* Explaining Deep Features Using Radiologist-Defined Semantic Features and Traditional Quantitative Features. Tomography 2019;5:192–200.

66. Caruana R, Lou Y, Gehrke J, Koch P, Sturm M, Elhadad N. Intelligible Models for HealthCare: Predicting Pneumonia Risk and Hospital 30-day Readmission. Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; Sydney, NSW, Australia: Association for Computing Machinery; 2015. p. 1721–30.

67. Ashrafinia S, Dalaie P, Yan R, Ghazi P, Marcus C, Taghipour M, *et al.* Radiomics Analysis of Clinical Myocardial Perfusion SPECT to Predict Coronary Artery Calcification. J Nucl Med 2018;59(Suppl 1):512.

68. Guo Y, Hu Y, Qiao M, Wang Y, Yu J, Li J, *et al.* Radiomics Analysis on Ultrasound for Prediction of Biologic Behavior in Breast Invasive Ductal Carcinoma. Clin Breast Cancer 2018;18:e335–44.

69. Hatt M, Tixier F, Pierce L, Kinahan PE, Le Rest CC, Visvikis D. Characterization of PET/CT images using texture analysis: the past, the present… any future? Eur J Nucl Med Mol Imaging 2017;44:151–65.

70. Vriens D, Visser EP, de Geus-Oei LF, Oyen WJ. Methodological considerations in quantification of oncological FDG PET studies. Eur J Nucl Med Mol Imaging 2010;37:1408–25.

71. van Velden FH, Cheebsumon P, Yaqub M, Smit EF, Hoekstra OS, Lammertsma AA, *et al.* Evaluation of a cumulative SUV-volume histogram method for parameterizing heterogeneous intratumoural FDG uptake in non-small cell lung cancer PET studies. Eur J Nucl Med Mol Imaging 2011;38:1636–47.

72. Hatt M, Le Rest CC, Tixier F, Badic B, Schick U, Visvikis D. Radiomics: Data Are Also Images. J Nucl Med 2019;60(Suppl 2):38S–44S.

73. Cook GJ, Yip C, Siddique M, Goh V, Chicklore S, Roy A, *et al.* Are pretreatment 18F-FDG PET tumor textural features in non-small cell lung cancer associated with response and survival after chemoradiotherapy? J Nucl Med 2013;54:19–26.

74. Cook GJ, O'Brien ME, Siddique M, Chicklore S, Loi HY, Sharma B, *et al.* Non-Small Cell Lung Cancer Treated with Erlotinib: Heterogeneity of (18)F-FDG Uptake at PET-Association with Treatment Response and Prognosis. Radiology 2015;276:883–93.

75. Collarino A, Garganese G, Fragomeni SM, Pereira Arias-Bouda LM, Ieria FP, Boellaard R, *et al.* Radiomics in vulvar cancer: first clinical experience using (18)F-FDG PET/CT images. J Nucl Med 2019;60:199–206.

76. Tixier F, Hatt M, Valla C, Fleury V, Lamour C, Ezzouhri S, *et al.* Visual versus quantitative assessment of intratumor 18F-FDG PET uptake heterogeneity: prognostic value in non-small cell lung cancer. J Nucl Med 2014;55:1235–41.

77. Parmar C, Leijenaar RT, Grossmann P, Rios Velazquez E, Bussink J, Rietveld D, *et al.* Radiomic feature clusters and prognostic signatures specific for Lung and Head & Neck cancer. Sci Rep 2015;5:11044.

78. Li X, Yin G, Zhang Y, Dai D, Liu J, Chen P, *et al.* Predictive Power of a Radiomic Signature Based on 18F-FDG PET/CT Images for EGFR Mutational Status in NSCLC. Front Oncol 2019;9:1062.

79. Fave X, Zhang L, Yang J, Mackin D, Balter P, Gomez D, *et al.* Delta-radiomics features for the prediction of patient outcomes in non-small cell lung cancer. Sci Rep 2017;7:588.

80. Dong X, Sun X, Sun L, Maxim PG, Xing L, Huang Y, *et al.* Early Change in Metabolic Tumor Heterogeneity during Chemoradiotherapy and Its Prognostic Value for Patients with Locally Advanced Non-Small Cell Lung Cancer. PLoS One 2016;11:e0157836.

81. Dissaux G, Visvikis D, Da-Ano R, Pradier O, Chajon E, Barillot I, *et al.* Pre-treatment 18F-FDG PET/CT Radiomics predict local recurrence in patients treated with stereotactic radiotherapy for early-stage non-small cell lung cancer: a multicentric study. J Nucl Med 2019;jnumed.119.228106.

82. Dou TH, Coroller TP, van Griethuysen JJ, Mak RH, Aerts HJ. Peritumoral radiomics features predict distant metastasis in locally advanced NSCLC. PLoS One 2018;13:e0206108.

83. Mattonen SA, Davidzon GA, Benson J, Leung AN, Vasanawala M, Horng G, *et al.* Bone Marrow and Tumor Radiomics at 18F-FDG PET/CT: Impact on Outcome Prediction in Non-Small Cell Lung Cancer. Radiology 2019;293:451–9.

84. Carvalho S, Leijenaar RT, Troost EG, van Timmeren JE, Oberije C, van Elmpt W, *et al.* 18F-fluorodeoxyglucose positron-emission tomography (FDG-PET)-Radiomics of metastatic lymph nodes and primary tumor in non-small cell lung cancer (NSCLC) - A prospective externally validated study. PLoS One 2018;13:e0192859.

85. Tixier F, Vriens D, Cheze-Le Rest C, Hatt M, Disselhorst JA, Oyen WJ, *et al.* Comparison of Tumor Uptake Heterogeneity Characterization Between Static and Parametric 18F-FDG PET Images in Non-Small Cell Lung Cancer. J Nucl Med 2016;57:1033–9.

86. Antunes J, Viswanath S, Rusu M, Valls L, Hoimes C, Avril N, *et*

al. Radiomics Analysis on FLT-PET/MRI for Characterization of Early Treatment Response in Renal Cell Carcinoma: A Proof-of-Concept Study. Transl Oncol 2016;9:155–62.

87. Zamboglou C, Carles M, Fechter T, Kiefer S, Reichel K, Fassbender TF, *et al.* Radiomic features from PSMA PET for non-invasive intraprostatic tumor discrimination and characterization in patients with intermediate- and high-risk prostate cancer - a comparison study with histology reference. Theranostics 2019;9:2595–605.

88. Sanduleanu S, Woodruff HC, de Jong EE, van Timmeren JE, Jochems A, Dubois L, *et al.* Tracking tumor biology with radiomics: A systematic review utilizing a radiomics quality score. Radiother Oncol 2018;127:349–60.

89. Rahmim A, Huang P, Shenkov N, Fotouhi S, Davoodi-Bojd E, Lu L, *et al.* Improved prediction of outcome in Parkinson's disease using radiomics analysis of longitudinal DAT SPECT images. Neuroimage Clin 2017;16:539–44.

90. Ypsilantis PP, Siddique M, Sohn HM, Davies A, Cook G, Goh V, *et al.* Predicting Response to Neoadjuvant Chemotherapy with PET Imaging Using Convolutional Neural Networks. PLoS One 2015;10:e0137036.

91. Wang H, Zhou Z, Li Y, Chen Z, Lu P, Wang W, *et al.* Comparison of machine learning methods for classifying mediastinal lymph node metastasis of non-small cell lung cancer from 18F-FDG PET/CT images. EJNMMI Res 2017;7:11.

92. Ma L, Ma C, Liu Y, Wang X. Thyroid Diagnosis from SPECT Images Using Convolutional Neural Network with Optimization. Comput Intell Neurosci 2019;2019:6212759.

93. Ortiz A, Munilla J, Martínez-Ibañez M, Górriz JM, Ramírez J, Salas-Gonzalez D. Parkinson's Disease Detection Using Isosurfaces-Based Features and Convolutional Neural Networks. Front Neuroinform 2019;13:48.

94. Chartrand G, Cheng PM, Vorontsov E, Drozdzal M, Turcotte S, Pal CJ, *et al.* Deep Learning: A Primer for Radiologists. Radiographics 2017;37:2113–31.

95. Aronson AR. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. Proc AMIA Symp 2001;17–21.

96. Vasaikar SV, Straub P, Wang J, Zhang B. LinkedOmics: analyzing multi-omics data within and across 32 cancer types. Nucleic Acids Res 2018;46(D1):D956–63.

97. Papp L, Spielvogel CP, Rausch I, Hacker M, Beyer T. Personalizing Medicine Through Hybrid Imaging and Medical Big Data Analysis. Front Phys 2018;6.

98. Rajkomar A, Oren E, Chen K, Dai AM, Hajaj N, Hardt M, *et al.* Scalable and accurate deep learning with electronic health records. NPJ Digit Med 2018;1:18.

99. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, *et al.* The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 2016;3:160018.

100. Buvat I, Orlhac F. The dark side of radiomics: on the paramount importance of publishing negative results. J Nucl Med 2019;60:1543–4.

101. Deist TM, Jochems A, van Soest J, Nalbantov G, Oberije C, Walsh S, *et al.* Infrastructure and distributed learning methodology for privacy-preserving multi-centric rapid learning health care: euroCAT. Clin Transl Radiat Oncol 2017;4:24–31.

102. Alexander A, McGill M, Tarasova A, Ferreira C, Zurkiya D. Scanning the Future of Medical Imaging. J Am Coll Radiol 2019;16(4 Pt A):501–7.

103. Kaptein M. Personalization in biomedical-informatics: methodological considerations and recommendations. J Biomed Inform 2019;90:103088.

104. Thompson WR. On the likelihood that one unknown probability exceeds another in the view of the evidence of two samples. Biometrika 1933;25:285–94.

Conflicts of interest.—The authors certify that there is no conflict of interest with any financial organization regarding the material discussed in the manuscript.

Acknowledgements.—The authors would like to thank Gerrit Kracht for providing the figure.

History.—Article first published online: May 12, 2020. - Manuscript accepted: May 12, 2020. - Manuscript received: April 10, 2020.