Radiomics

Citation for published version (APA):

Document status and date:
Published: 01/01/2020

DOI:
10.1259/bjr.20190948

Document Version:
Publisher's PDF, also known as Version of record

Document license:
Taverne

Please check the document version of this publication:
• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.
• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:
www.umlib.nl/taverne-license

Take down policy
If you believe that this document breaches copyright please contact us at:
repository@maastrichtuniversity.nl
providing details and we will investigate your claim.

Download date: 17 Sep. 2023

BJR 125th ANNIVERSARY: REVIEW ARTICLE

Radiomics: from qualitative to quantitative imaging

William Rogers, MSc, Sithin Thulasi Seetha, MSc, Turkey A. G. Refaee, Relinde I. Y. Lieverse, MD, Renée W. Y. Granzier, MD, Abdalla Ibrahim, MD, Simon A. KEEK, MSc, Sebastian Sanduleanu, MD, Sergey P. Primakov, MSc, Manon P. L. Beuque, Ing., Damienne Marcus, MSc, Alexander M. A. Van Der Wiel, MSc, Fadila Zerka, MSc, Cary J. G. OBERJE, PhD, Janita E Van Timmeren, PhD, Henry C. Woodruff, PhD and Philippe Lambin, MD, PhD

1,2 The D-Lab & The M-Lab, Department of Precision Medicine, GROW - School for Oncology and Developmental Biology, Maastricht University Medical Centre, Maastricht, The Netherlands
2 Department of Thoracic Oncology, IRCCS Foundation National Cancer Institute, Milan, Italy
3 Department of Diagnostic Radiology, Faculty of Applied Medical Sciences, Jazan University, Jazan, Saudi Arabia
4 Department of Radiology and Nuclear Imaging, GROW-School for Oncology and Developmental Biology, Maastricht University Medical Centre, Maastricht, The Netherlands
5 Department of Surgery, Maastricht University Centre, Grow-School for Oncology and Developmental Biology, Maastricht, The Netherlands
6 Department of Nuclear Medicine and Comprehensive diagnostic center Aachen (CDCA), University Hospital RWTH Aachen University, Aachen, Germany
7 Division of Nuclear Medicine and Oncological Imaging, Department of Medical Physics, Hospital Center Universitaire De Liege, Liege, Belgium
8 Division of Radiation Oncology, University Hospital Zürich, Zürich, Switzerland
9 University of Zürich, Zürich, Switzerland

Address correspondence to: Dr Henry C. Woodruff
E-mail: h.woodruff@maastrichtuniversity.nl

William Rogers and Sithin Thulasi Seetha have contributed equally to this study and should be considered as co-first authors.

Henry C. Woodruff and Philippe Lambin have contributed equally to this study and should be considered as senior authors.

ABSTRACT

Historically, medical imaging has been a qualitative or semi-quantitative modality. It is difficult to quantify what can be seen in an image, and to turn it into valuable predictive outcomes. As a result of advances in both computational hardware and machine learning algorithms, computers are making great strides in obtaining quantitative information from imaging and correlating it with outcomes. Radiomics, in its two forms “handcrafted and deep,” is an emerging field that translates medical images into quantitative data to yield biological information and enable radiologic phenotypic profiling for diagnosis, theragnosis, decision support, and monitoring. Handcrafted radiomics is a multistage process in which features based on shape, pixel intensities, and texture are extracted from radiographs. Within this review, we describe the steps: starting with quantitative imaging data, how it can be extracted, how to correlate it with clinical and biological outcomes, resulting in models that can be used to make predictions, such as survival, or for detection and classification used in diagnostics. The application of deep learning, the second arm of radiomics, and its place in the radiomics workflow is discussed, along with its advantages and disadvantages. To better illustrate the technologies being used, we provide real-world clinical applications of radiomics in oncology, showcasing research on the applications of radiomics, as well as covering its limitations and its future direction.

INTRODUCTION

Medical imaging technologies in healthcare have expanded remarkably from the discovery of X-Rays 124 years ago to the use of CT, MRI, and positron emission tomography (PET), among others in modern-day clinical practice (Figure 1). These tools have become an integral part in detection and diagnosis for many diseases due to several factors, including: the minimally invasive nature of imaging, rapid technological developments, lower costs compared...
Medical imaging in its infancy generated analogue images, which underwent subjective interpretation based on visual inspection and verbal communication. By the end of the 20th century, information technology has brought radiology to the digital world, although the interpretation of radiographs remained mostly qualitative. Humans excel at recognising patterns through visual inspection, however, they are often lacking when performing complex quantitative assessments. In the early 1960s, researchers started to focus on computerised quantitative analysis of medical data for aiding clinical diagnosis, what later came to be known as computer-aided diagnosis (CAD) systems. However, these systems were using a classical approach using statistical analysis and probability theories, and the volume of available data was low, so the results were often too inaccurate for clinical use. Later in the 1980s, further advances in theoretical computer science and digital imaging lead to the development of advanced machine learning and pattern recognition algorithms, which when integrated with CAD systems were able to generate clinically reliable results.

In recent decades, simple quantitative image analysis (QIA)(10,14),(987,989) has been adopted by clinicians (e.g. RECIST), and has been primarily focused on assisting qualitative observations. For instance, CAD systems can be found in healthcare worldwide, aiding radiologists and clinicians in making diagnostic and theragnostic decisions. One of the most typical applications of CAD systems is in recognising abnormalities during cancer screening. Notable contributions are in the area of lung and breast cancer research. For example, there are many CAD studies which focus on detecting and diagnosing lung nodules (as benign or malignant) on CT and chest radiographs. Similarly, many such studies have been conducted in breast mammography images for highlighting microcalcifications, architectural distortions, and the prediction of mass type. It is conceivable that the lack of quantitative information leads to increased follow-ups or invasive biopsies that would be deemed unnecessary given the unused information in medical images. Even though there have been various developments in QIA, traditionally radiologists are trained to understand the behaviour of the underlying disease through visual inspection of radiographic images. This partially explains why most of the developments in imaging technology are in optimising the visual representation of the generated images, with vendors competing to generate the highest quality images. With the exception of CT, with its semi-parametric calibrated Hounsfield Units, and some particular MRI sequences, individual voxel values do not correlate with the underlying biology without further calibration and modelling. Furthermore, qualitative analysis is not so dependent on reproducible voxel values, while machines on the other hand only process numerical values and rely on the standardisation of image acquisition and reconstruction to yield reproducible results. The lack of standardisation of medical images has been a major hurdle in the development of QIA in medical imaging. However, in recent years, quantitative imaging is becoming more popular with the advent of, e.g. quantitative fluorodeoxyglucose-PET or quantitative MRI for treatment response assessment.

The ubiquitous computer, vast amounts of data, and advanced algorithms have opened a new era in medical imaging. The high information density of images allows for many quantitative metrics since intricate pixel and voxel relationships can be captured by complex operations. Radiomics involves the process of extraction of quantifiable features from vast amounts of data that might correlate with the underlying biology or clinical outcomes using advanced machine learning analysis techniques. Radiomics has two main arms, based on how imaging information is transformed into mineable data: handcrafted radiomics and deep learning. Handcrafted features are formulas mostly based on intensity histograms, shape attributes, and texture, that can be used to fingerprint phenotypic characteristics of the radiograph while in deep learning a complex network "creates" its own features. Various statistical and machine learning models have been widely researched, and are envisioned to be complementary to best medical practice by aiding in making informed clinical decisions in both oncological and non-oncological diseases.

Since the 1990s predictions were being made that genomics, spearheaded by the Human Genome Project, would completely transform therapeutic medicine, heralding precision medicine. Precision medicine, also termed personalised medicine, originally referred to the view that incorporating genomic information in the clinical workflow will lead to marked improvements in the prediction, diagnosis, and treatment of diseases. Recently, the scope of precision medicine has expanded to incorporate inputs beyond the genome. Radiomics and other "-omic" developments, such as metabolomics and proteomics, are contributing
Radiomics: from qualitative to quantitative imaging

In this review, we provide a broad overview and update on the fast-growing field of quantitative imaging research, focussing on the two arms “handcrafted radiomics and deep learning” describing some of its caveats and giving examples of the budding clinical implementation, the stepping stones towards precision medicine.

**Radiomics: from feature extraction to correlation with outcomes**

Performing feature extraction of textures in medical imaging is nothing new and in fact serious research had begun in the early 1980s at Kurt Rossmann Laboratories for Radiologic Image Research in the Department of Radiology at the University of Chicago to develop CAD systems for the detection of lung nodules as well as detection of clustered microcalcifications in mammograms. The first CAD patent was filed all the way back in 1987 using a method of pixel thresholding and contiguous pixel area thresholding.40

The radiomic workflow begins with the medical image, which can be represented in two, three, or four dimensions. Images contain quantitative data in the form of signals that are captured at different scales and variation across medical machines. Normalisation techniques are used to distribute pixel intensities evenly across a data set evenly and within a standardised range. Next, a region of interest (ROI) is defined so that only information related to the lesion can be extracted, and the useful information that can be extracted are called features. There are competing methods to extract features both in two-dimensional and three-dimensional. One such method is the manual segmentation of the lesion or the creation of a bounding box, as seen in Figure 2. This can also be performed using automated segmentation algorithms. Methods for automated segmentation include deep learning architectures such as U-Net, or semi-automatic methods like click-and-grow algorithms.45,46

Once the ROI is defined, the choice of features to be extracted depend on the information being sought. Shape features such as volume relate only to the definition of the ROI, and if this is manually created, suffer from inter- and intraobserver variability.47 First-order features give insight into the distribution of pixel intensities, e.g., histograms of pixel intensities are quantified by a large number of statistical methods, including variance, skewness, and kurtosis. These features, however, are unable to quantify how pixels are positioned in relation to each other. Second and higher-order features may capture this relationship, with second-order features obtained based on the average relationship between two pixels/voxels, and higher-order features for more than two pixels/voxels. An example of a second-order feature extraction method is the grey level co-occurring matrix (GLCM). GLCMs are co-occurring pixels in each defined direction (Figure 3) and are counted and recorded (Figure 4) into a matrix. Statistical analysis such as contrast, correlation, and homogeneity, as well as tailored formulae can then be applied on the GLCM to extract independent features.48 Features extracted in this manner are considered “handcrafted” features as they are features that are pre-defined by specially designed formulae.
After features have been extracted from all the images in a database, a subset of features needs to be selected that go into the final model. To make a model generalisable, it is important to avoid finding spurious correlations in the data that do not generalise to other similar data sets, an occurrence termed overfitting. If a model has learned to recognise noise, outliers, or other kinds of variance, it is unlikely to perform well when presented with new data. The larger the number of predictors, the larger the chance to find spurious correlations, a major problem in the realm of machine learning. To detect overfitting, ideally, a model’s performance is validated in external data sets with similar population and outcome distributions, but from different centres—if the model performs significantly better on the training set than on the validation set, overfitting is likely. In the absence of an external validation data set, data can be split into different subsets, and the model trained in one group and validated on the other(s) in a process called cross-validation (Figure 5). During this process, the model hyperparameters (settings within the model itself, e.g. degree of polynomial fitting) can be further tuned to increase performance in the training and validation sets.

A method to overcome overfitting is to reduce the number of predictors, in this case, imaging features. Feature selection is the process of reducing the number of predictors while retaining the core important information that correlates with outcomes or the underlying biology. Many feature reduction methods exist, but none are known to work well on all kinds of data sets, and they can be combined in many ways. This remains an active field of research. Similar features can also be grouped to achieve dimensionality reduction, and methods such as principal component analysis and independent component analysis are employed to this end.

Once features are selected, the task is to correlate these features—individually or in groups—to diagnostic and prognostic outcomes or to the underlying biology. There are numerous methods to find and test such models, from simple linear regression and curve-fitting to advanced machine learning methods such as decision trees, support vector machines, random forests, boosted trees, or neural networks. Ensembling is the combination of models that get trained on random samples of data from the training set called bags and then combined as a whole using a voting system. This is the basis for algorithms such as Random Forests, AdaBoost, and Gradient Boosting. An intuitive explanation is that even though the individual models can show a large amount of variance due to being trained on small subsets of the data, their averaging or voting smooths out the variance while improving the ability to better generalise.

Once a generalisable model has been trained and externally validated, it might be desirable to expand the interoperability of the model to all hardware, acquisition, and reconstruction parameters found in general clinical practice. Instead of relying on the standardisation of images, the features themselves can be harmonised to a common frame-of-reference using combined batch methods such as ComBat, originally developed for similar problems encountered in gene sequencing assays.

Figure 4. Calculating a GLCM for horizontal co-occurring pixel intensities. In total, 3 co-occurring pixel intensities of 3 and 2 that are next to each other on a horizontal plane can be totalled and tracked in the corresponding matrix. GLCM, grey level co-occurring matrix.

Figure 5. An example of fivefold cross-validation which can be used to evaluate machine learning models. Cross-validation gives the ability to test the result across the entirety of a data set, giving a better estimation of a model’s overall performance.
Artificial neural networks (ANNs) are a class of machine learning architecture that are loosely based on how biological brains work. With the exception of unsupervised learning (such as autoencoders), deep learning architectures usually rely on information regarding the outcome in order to craft their features, and unlike in handcrafted radiomics, feature extraction and correlation are intertwined. Also, unlike radiomics, there is generally no need for image segmentation, as the whole image can be presented to a deep learning model, both during training and in clinical routine.

An ANN is able to use a collection of neurons and weights, one for each of the inputs preceding the neuron. These weights get continuously updated, or corrected, in steps called epochs that work together to create a very complex function able to make predictions. The weights are inputs for each neuron and are multiplied and averaged, resulting in a transfer function, which is converted to an output via a function called an activation function. These activation functions are often a sigmoidal function such as a hyperbolic tangent or sigmoid, or a function called a rectified linear unit that can be represented as the maximum of the product of the coefficient and zero or one. A representation of a single neuron, including the activation function, can be seen in Figure 6. Multiple neurons can then be stacked to create a single layer referred to as a “hidden layer” and hidden layers (where inputs and outputs all connect) can be stacked to create larger networks, see Figure 7. The term deep learning is used to describe a neural network that has many layers, which is considered deep. For a binary classifier or regression, the final layer should contain only a single neuron and use a sigmoid activation function to make a prediction with a binary outcome (zero or one). If the problem is categorical, the network's final layer should contain the same number of neurons as there are categories to be classified and the final activation will be a “softmax”.

Figure 6. The architecture of a single neuron with a transfer function and a sigmoid activation function visualised.

Figure 7. A three layer neural network that is a binary classifier with three inputs. Nodes with Xn refer to inputs while other nodes refer to activation functions. The connecting lines between the nodes represent weights.
function, which is the average of the exponentials of the inputs, yielding the probabilities of each category. Deep learning for image vision employs convolutional neural networks (CNNs) which are a type of ANN that have an automated feature extractor designed specifically for images. CNNs employ a filtering technique, which convolves the image with a kernel (sliding window), creating a new pixel/voxel value (and hence new image) by sliding a matrix of numbers over the image, see Figure 8. It is possible to make a variety of different filters using these types of convolutions, such as blurring, sharpening, edge detection, and gradient detection, and CNNs are able to learn filters that are best suited to extracting features needed for making predictions.

ANNs do have some drawbacks compared to using handcrafted features alongside other machine learning techniques. The main drawback is the intrinsic need for much larger datasets to train the models, since feature creation is contingent on the training data, as opposed to handcrafted radiomics. Another drawback to using ANNs is interpretability. ANNs build ultracomplex functions that can be extremely difficult for practitioners to make sense of. Although CNNs have performed very well in image recognition, they have been less successful learning texture features, since texture information inherently has a higher dimensionality compared to other types of data sets, making them more difficult for neural networks to master. According to Basu et al, a redesign of neural network architectures is required to extract features in a similar manner as GLCM and other features based on spatial correlation.

Currently, the main application of deep learning in the radiomics workflow still lies in the automated detection and localisation of organs and lesions, removing the major burden in data set cura- tion. While there is no algorithm that can solve every problem, deep learning still has its place and is able to work as additional methods for delineation and feature extraction that compliments handcrafted radiomics. There is active research in combining both deep learning features and radiomics features that shows improved results.

### POTENTIAL CLINICAL APPLICATIONS

#### Radiomics in oncology

Radiomics has been widely studied for application in diagnosis and treatment prognosis/selection in oncology, primarily due to the existence of large imaging data sets used for staging, often containing delineations of tumours and organs at risk necessary for radiation treatment planning. These data sets can be used to train diagnostic and prognostic models for a variety of cancer types and sites. Using clinical reports, pathology/histology, and genetic information along with radiomics analysis can give a global outlook on the biology of the disease. In this section, an overview of notable studies published in this area will be discussed.

#### Lung

Lung cancer is by far the leading cause of cancer-related deaths among both males and females worldwide. Recent studies have shown that radiomics can determine the risk of lung cancer from screening scans. Radiomic features found to have a strong association to decode tumour heterogeneity for risk stratification, concluding that patients with heterogeneous tumours tend to have a worse prognosis. In addition to that, Yoon et al were able to show the association of radiomic analysis with gene expression. Radiomic features were also found to correlate with TNM staging for lung and head-and-neck cancer. Later studies further validated the strong predictive power of radiomics for distant metastasis.

Radiomics may also play a role in lung cancer treatment planning by evaluating tumour response to a specific treatment. Several studies focused on analysing the tumour response to radiation therapy. For instance, Mattonen et al developed a radiomics signature for treatment response to stereotactic ablative radiation therapy that was able to predict lung cancer recurrence post-therapy, while Fave et al used multiple time point information referred to as delta-radiomic analysis to evaluate the change of radiomic features as a predictor for tumour response to radiation therapy. The results suggest that 6 radiomic features are in fact a good indicator of treatment response. Another interesting study by Mattonen et al found that radiomic analysis can identify features associated with local recurrence of lung cancer after
radiation therapy, while physicians usually have great difficulty to distinguish local recurrence from radiation-induced sequelae.

Besides the traditional handcrafted feature extraction approach followed in the radiomics pipeline, deep learning radiomics is also gaining popularity among researchers. A deep learning-based approach followed by Shen et al yielded more accurate malignancy prediction of nodules compared to previous methods. Pham et al used a two-step deep learning approach for evaluating lymph node metastases with accurate cancer detection. Instead of using data from a single time point, deep recurrent convolutional network architectures can be used to analyse data from multiple time points to monitor treatment response.

Brain

Brain tumours are usually graded based on clinical or pathological analysis to define their malignancy. Radiomics may be able to non-invasively perform grade assessment, as reported by Coroller et al in meningioma patients, suggesting a strong correlation between certain imaging features and histopathological grade. Zhang et al were able to classify between low-grade gliomas and high-grade gliomas with high accuracy. Chen et al investigated the prediction of brain metastases in T1 lung adenocarcinoma patients and found that the predictive performance for the radiomics model was significantly better compared to clinical models and could potentially be used for brain metastases screening. Petit et al performed radiomic analysis for the classification of brain tumours in childhood suggesting that radiomics can aid in the classification of tumour subtype. However, the scalability of the techniques used in these studies needs to be assessed further by extensions to multicentric cohorts using different acquisition protocols and vendors.

Radiation therapy can lead to necrosis, which is difficult to distinguish from tumour recurrence on imaging. Larroz & al were able to develop a high classification accuracy model to distinguish between brain metastasis and radiation necrosis using radiomic analysis. Some radiomic studies successfully investigated the treatment response in recurrent glioblastoma patients with a radiomics approach. An iterative study by radiomic researchers found strong evidence of radiomic features in predicting survival and treatment response of patients with glioblastoma using pre-treatment imaging data.

Deep learning has also made some other interesting contributions in this area. Chang et al used residual deep convolutional network for predicting the genotype in Grade II-IV glioma with high accuracy. Deep learning can also be used complementary to traditional handcrafted radiomics studies. For example, studies focused on using deep networks for segmentation, followed by radiomics analysis for survival prediction.

Breast

Among females, breast cancer is the second leading cause of death for cancer worldwide. However, earlier diagnosis can lead to a better prognosis. Radiomics in the field of breast cancer has been applied to several imaging modalities including (PET)-MRI, (contrast-enhanced) mammography, ultrasound, and digital breast tomosynthesis focusing on tumour classification, molecular subtypes, tumour response prediction to neoadjuvant systemic therapy (NST), lymph node metastasis, overall survival, and recurrence risks. For example, a large number of radiomics studies have been used for the prediction of malignant breast cancers. Besides the prediction of tumour malignancy, several radiomics studies examined the prediction of breast cancer molecular subtypes with the aim of leaving out liquid biopsies in the future. Lymph node metastasis identification is an important prognostic factor and often determines treatment. In all clinically node negative patients, a sentinel lymph node procedure is the basis of the axillary treatment. Dong et al was able to provide an alternative to this invasive approach by successfully applying radiomics for the prediction of lymph node metastasis in the sentinel lymph node using imaging data.

In addition to the prediction of breast tumour malignancy, tumour molecular subtypes and sentinel lymph node metastasis identification, radiomics studies have also made some significant contributions to treatment planning. Chan et al investigated the power of radiomics to discriminate between patients with low and high treatment failure risk on pre-treatment imaging data. There are multiple studies that predict tumour response to NST using radiomic analysis. For instance, Bramer et al found a combination of intratumoral and peritumoral radiomics features as a robust and strong indicator for pathologic complete tumour response using pre-treatment imaging data. Two other studies found similar evidence on serial imaging data containing follow-up scans. The use of multiparametric MRI for the prediction of tumour response to NST showed promising results.

Deep learning approaches have also been adopted in breast cancer research. The study of Huyn & al investigated tumour classification capacity of deep features extracted from convolutional networks trained on a different data set to analytically extracted features. The results suggested a higher performance of deep features. Similarly, another study used deep learning for risk assessment and found higher performance compared to conventional texture analysis.

Other sites and diseases

While cancers of the lung, brain, and breast have received wide attention from the radiomics research community, any site is open to QIA research. Diagnostic and prognostic radiomics research is ongoing for cancers of the head-and-neck, ovaries, prostate, kidney, liver, colon and rectum, and many other sites. The main requirements for a radiomics study are the presence of a radiologic phenotype which allows for the clustering of patients based on differences within that phenotype or some correlation to the underlying biology, and the availability of imaging and clinical data. While not nearly as prevalent, this has meant that non-oncological diseases which require medical imaging as part of the standard of care have also been the subject of radiomics analysis, such as in the fields of neurology, ophthalmology, and dentistry.

Limitations of radiomics and future directions towards precision medicine

While radiomics facilitates new possibilities in the field of personalised medicine, some challenges remain. One of the
primary obstacles is the lack of big and standardised clinical data. Although large amounts of medical imaging data are stored, these data are dispersed across different centres and acquired using different protocols. Access for research purposes is highly restricted by law and ethics. An exhaustive data curation and harmonisation process is still necessary to make it usable for research. Radiomics will potentially enable imaging-based clinical decision support systems, however, the current black box approach, particularly in deep learning, makes it less acceptable for clinical application. In certain cases, handcrafted radiomic features have already been correlated with biological processes, but it is essential to work further in the direction of interpretable artificial intelligence (AI) to make it more accessible for clinical implementation. In recent years, various countries have already adopted many measures to control variability in clinical trial protocols, data acquisition, and analysis. For example, across Europe consistent protocol guidance was adopted with the help of European Association of Nuclear Medicine. The Quantitative Imaging Biomarker Alliance initiative also aims to achieve the same task in a much broader level. On the other hand, algorithmically, developments in deep learning allow for automated quality check, clustering of data, and automated detection and cueing of organs and lesions, vastly improving data curation times. Generative adversarial networks open up the possibility of generating synthetic data or domain adaptive algorithms might be able to deal with the shortage of standardised images. Techniques like distributed learning provide the ability to train machine learning models using distributed data without the data ever leaving their original locations. Distributed learning has already been applied across several medical institutions to build predictive and segmentation models. Furthermore, this approach can be coupled with other technologies such as blockchain to trace back data provenance and monitor the use of the final models. Various techniques to visualise deep features have already been put forward by researchers to generate an intuitive understanding. A completely new research area of AI called explainable AI aims to track the decisions made by the intelligent algorithms so that it can be better understood by humans. Companies like Google, IBM, Microsoft and Facebook are at the forefront in this research. This will not only help to build trust of AI systems among medical professionals but also unlocks new possibilities in understanding a disease.

The implementation of precision medicine itself has its own limitations and has drawn criticism due to the lack of a "transformation in therapeutic medicine" in the last two decades. So far, life expectancies or other public health measures have not shown any dramatic improvements, regardless of the vast amounts of precision medicine research being conducted. Contentious points remain such as excessive costs (e.g. gene therapy), although new developments such as radiomics promise to reduce costs in the long run. Furthermore, the diagnostic and prognostic power of complex "omics-driven" models is still to be determined in specific populations, and evidence needs to be produced that such methods improve health outcomes. Precision medicine is likely to mature and translate to clinical workflows over the next decade and will change the way health services are delivered and evaluated. Healthcare systems will need to adjust their methods and processes to accommodate for these changes.

CONCLUSION
Radiomics, whether handcrafted or deep, is an emerging field that translates medical images into quantitative data to give biological information and enable phenotypic profiling for diagnosis, theragnosis, decision support, and monitoring. Radiomics, in essence, allows personalisation of care by identifying features or signatures associated with a disease or a treatment response with high precision and in a non-invasive way. Recent developments in genomics and deep learning have pushed radiomics researchers to focus more on extracting deep features and explore new possibilities in AI modelling. In the future, radiomics will be a valuable addition to precision medicine workflows by facilitating earlier and more accurate diagnosis, providing prognostic information, aiding in treatment choice, monitoring disease and treatment non-invasively, and enabling routine dynamic treatment based on individual responses. But the road to this vision is long, and many technical, regulatory, and ethical problems still need to be solved.

CONFLICT OF INTEREST
Dr Philippe Lambin reports, within and outside the submitted work, grants/sponsored research agreements from Varian medical, Oncoradiomics, ptTheragnostic, Health Innovation Ventures and DualTpharma. He received an advisor/presenter fee and/or reimbursement of travel costs/external grant writing fee and/or in-kind manpower contribution from Oncoradiomics, BHV, Merck and Convert pharmaceuticals. Dr Lambin has shares in the company Oncoradiomics SA and Convert pharmaceuticals SA and is co-inventor of two issued patents with royalties on radiomics (PCT/NL2014/050248, PCT/NL2014/050728) licensed to Oncoradiomics and one issue patent on mtDNA (PCT/EP2014/059089) licensed to ptTheragnostic/DNAmito, three non-patentable invention (software) licensed to ptTheragnostic/DNAmito, Oncoradiomics and Health Innovation Ventures.

Dr Woodruff has (minority) shares in the company Oncoradiomics.

FUNDING
Authors acknowledge financial support from ERC advanced grant (ERC-ADG-2015, no 694812 - Hypoximmuno), ERC-2018-PoC (no 81320 - CL-IO). This research is also supported by the Dutch Technology Foundation STW (grant no P14-19 Radiomics STRaTegy), which is the applied science division of NWO, and the Technology Programme of the Ministry of Economic Affairs. Authors also acknowledge financial support from SME Phase 2 (RAIL - no 673780), EUROSTARS (DART, DECIDE), the European Program H2020-2015-17 (BD2Decide - PHC30-689715, ImmunoSABR - no 733008, PREDICT - ITN - no 766276), TRANSCAN Joint Transnational Call 2016 (JTC2016 "CLEARLY" - no UM 2017–8295), Interreg V-A Euregio Meuse-Rhine ("Euradiomics"). This work was supported by the Dutch Cancer Society (KWF Kankerbestrijding), Project number 12085/2018–2
REFERENCES

5. Miller GA. The magical number seven plus or minus two: some limits on our capacity for processing information. Psychol Rev 1956; 63: 81–97. doi: https://doi.org/10.1037/h0043158
15. Halabi B, Makandar A. Computer Aided Diagnosis - Medical Image Analysis Techniques [Internet. Breast Imaging 2018;
32. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J,


43. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology* 2016; 278: 563–77. doi: https://doi.org/10.1148/radiol.2015151169


59. Bengio Y, Delalleau O, Le Roux N. The curse of highly variable functions for local kernel


137. Mankoff DA. Quantitative imaging as a reproducibility study on renal clear cell carcinomas. *Eur J Radiol* 2019; 116; 2572–81. doi: 10.1016/j.ejrad.2019.03.017


