Transcending the Oncological Application of Radiomics—Towards Predictive Signatures from Routine Chest CTs

Article Type:

Review article

Summary statement:

There are ways to facilitate the expansion of the application of radiomics to diseases other than tumors, and thus, increase the predictive potential of radiomics.

Essentials:

To cope with the heterogeneity inherent to routine data, a close collaboration between radiologists and IT-specialists is necessary.

Deep-learning-based radiomics may help to cope with technical variability in routine imaging data.

Large study cohorts may provide insights transferable to a more heterogeneous routine population.

Keywords:

radiomics, routine data, chest, computed tomography

List of abbreviations:

BMD	bone mineral density
COPD	chronic obstructive pulmonary disease
CLAD	chronic lung allograft dysfunction
DXA	dual energy x-ray absorptiometry
GOLD	global initiative for obstructive lung disease

ILD	interstitial lung disease
IPF	idiopathic pulmonary fibrosis
IPSI	image biomarker standardization initiative
TNM	TNM classification of malignant tumors
UID	unique identifier

Abstract:

Radiomics studies for oncological applications, particularly in lung cancer, have been increasing exponentially over the last years. The automatic extraction and quantification of imaging features may not only characterize histological subtypes of cancer but also help to predict morbidity and overall mortality by including cardiovascular, pulmonary and metabolic imaging features of structures other than the tumor. However, an adequate sample size as a statistical necessity for radiomics studies is often difficult to achieve in prospective trials. By exploiting imaging data from clinical routine, a much larger amount of data could be utilized than in clinical trials. Still, there is only little literature on the implementation of radiomics in clinical routine imaging data. Reasons are heterogeneous computed tomography (CT) scanning protocols and the resulting technical variability in routine CT imaging data. In this review, we summarize the recent state of the art of studies aiming to develop quantifiable imaging biomarkers in chest CTs beyond cancer. We explain solutions to overcome heterogeneity in routine data such as the utilization of imaging repositories, the standardization of radiomic features, algorithmic approaches to improve feature stability, test-retest studies and the evolution of deep-learning for modelling radiomics features.

Introduction

The highly active field of radiomics poses an opportunity to extract comprehensive information from medical images for the prediction of disease course, survival or even treatment response in oncology (1). In addition to staging, histology, molecular analyses, and performance status, comorbidities of lung cancer (e.g., chronic obstructive pulmonary disease (COPD), additional lung fibrosis, or congestive heart disease) are also predictors of survival (2). An overly concentrated focus on tumor-specific parameters may lead to the exclusion of additional factors, and thus to a neglect of important individual traits that support true precision medicine.

Only recently, the fast-paced development of machine-learning approaches allowed tapping into patient records and medical images from the clinical routine for the deduction of predictive markers. In the USA, nearly 73.8 million computed tomography (CT) examinations are conducted every year (3). The ability to utilize such large amounts of imaging data can greatly benefit the evolution of powerful computational imaging analysis methods.

The prospect of radiomics

What is radiomics?

The definition of radiomics is the "high-throughput extraction and analysis of quantitative imaging features from medical images" with the aim to generate predictive models and aid in clinical decision support (1). These features are either hand-crafted (i.e., mathematically predetermined such as, texture-, shape- and size-based features, as well as first-order statistics), or may be deep learning-based (4). Each feature constitutes a potential biomarker by providing specific, quantifiable, information about the image. Thus, radiomics is not a specific methodology but describes an integrated workflow from image acquisition over feature extraction to biomarker discovery (figure 1). However, this integration of different principles that are evolving from year to year leads to an ambiguity in

definition. In the following paragraphs we provide a short clarification on the relations of radiomics, machine- and deep learning, as well as (quantitative) imaging biomarkers.

What is the relation of radiomics to (quantitative) imaging biomarkers?

Imaging biomarkers represent biological features of pathologies. For instance, cavitation in lung cancer may indicate necrosis due to high growth rate with insufficient blood supply in the center and spiculation represents a pattern of local spread of cancer. Some imaging biomarkers may be easily deduced from pathophysiological mechanisms (e.g., bronchiectasis in lung fibrosis due to volume reduction and traction). Other imaging biomarkers may have no such obvious underlying pathomechanism (e.g, the reversed-halo-sign as a pathological pattern of the lung) but are invaluable for diagnosis as they are associated with several diseases. Quantitative imaging biomarkers compared to qualitative ones are better suited for follow-up evaluations by allowing the comparison of metrics such as tumor size or density.

Radiomics is the approach to extract and analyze a multitude of different imaging features (or biomarker candidates), including those which are not accessible by human vision or intuitive understanding. To avoid the discovery of imaging features without relevance, they need to be selected depending on their capability of predicting, classifying, monitoring or measuring a clinical outcome. Thus, radiomics may accelerate the discovery of (quantitative) imaging biomarkers and facilitate the subsequent integration of multiple singular imaging biomarkers for a more complete characterization of a medical image. An example is given by the study of Humphries et al. (5) in which the addition of features learned with a clustering analysis (a method of unsupervised machine learning) improved the prediction of pulmonary function in patients with idiopathic pulmonary fibrosis compared to a model using a histogram-based densitometric method alone.

However, biomarkers derived from radiomics are still at the level of *analytical validation* (the demonstration of accuracy, precision, and feasibility (6)), which is the first step in the development of an imaging biomarker. Multicenter studies for *qualification* (the demonstration of an association with

a clinical endpoint (6)) and *utilization* (assessment of the biomarker's performance (6)) are still largely missing.

Radiomics in the world of machine learning

Whereas a radiomics workflow could, in theory, be performed without machine learning (e.g., manual segmentation, handcrafted features and statistical analysis), machine learning provides the computational methods that make radiomics efficient. Manually segmenting thousands of scans containing hundreds of thousands of images is not feasible. We do not have to knowledge for handcrafting every possibly relevant imaging feature, especially those that are only relevant in combination with others. Finally, machine learning is better suited for dealing with multi-variety and multi-dimensional data to make accurate predictions. Thus, machine learning can facilitate an efficient radiomics approach at all) steps.

With the success of deep learning based methods (e.g., neural networks), end-to-end machine learning has become an often used tool in data analysis and especially image processing. End-to-end machine learning means that all the imaging features are automatedly learned and it has shown to yield improved results compared to the two step approach of handcrafted feature extraction and subsequent learning of the target task based on these features (7). However, while yielding good classification and prediction performance, this approach introduced a black box, making it harder to identify and interpret quantifiable features that lead to the decision of the deep learning-based method. Imperative goals of biomedical data analysis, are both the learning and understanding of underlying mechanisms and the identification of quantifiable biomarkers. To this end, there are currently two viable strategies: first, the two step approach of handcrafted feature extraction and subsequent machine learning-based analysis with a focus on reproducibility, interpretability and transferability of features which might constitute a clinically comprehensible approach. Second, the increasingly successful efforts to render deep learning models transparent, offering means to trace from model decision to responible image features by approaches such as GradCAM. These approaches

might lead to the “best of both worlds” by exploiting both the superior ability of deep learning to identify predictive features, and providing explainability.

How can radiomics support daily clinical routine?

The mainly qualitative reports radiologists create from medical images are merely a small part of what could be deduced from imaging data by applying quantitative approaches. Similar to a blood sample that has not been analyzed yet, most of this type of information is not accessible for human recognition and is often limited to specialized techniques (see figure 2). Thus, medical images acquired during clinical routine are often excluded from quantitative analysis, possibly omitting an opportunity to support decision-making and improve diagnostic accuracy. Each step of this complex workflow needs to be rigorously evaluated to ensure the scientific integrity of imaging feature sets, especially if these sets are to act as future validated, qualified and utilizable biomarkers with a reliable and accurate predictive value that important decisions can be based upon.
In Table 1, several studies are listed that use radiomics approaches in pulmonary and coronary diseases for classification, stratification, and prediction of gene expression profiles, pulmonary function, disease severity, possible response to medication and differentiation of pathologies.

What is the benefit of extracting and analyzing large amounts of imaging features rather than a preset few with an underlying pathomechanism already hypothesized?

Using features that are likely to correlate with a desired outcome due to their known relevance has the advantage of including existing knowledge in the analysis (e.g., bronchial wall thickening in COPD exacerbations (8)). The multitude of lung reaction patterns in interstitial lung diseases (ILD) is an example where this holds true, as the integration of multivariate information (i.e., not only the pattern itself, but also its spatial and temporal distribution) gained from a CT scan helps in determining the most probable diagnosis and track or predict disease progression (9) (figure 3 and 4). However, this may introduce a selection bias and decrease the chances of detecting novel predictive imaging features that have not been used by clinicians thus far. By applying a radiomics approach, the relationship of

separate voxels can be analyzed to gain additional non-intuitive information with which to support the radiologist in decision-making (10).

Need to expand the application areas of radiomics:

Oncology has been the main application for radiomics, and cross-sectional imaging is the primary modality. There is much insight to be gained from analyzing tumor heterogeneity; however, comprehensive management of lung cancer patients needs to address all present conditions. This becomes even more important, as comorbidities are common in this population (e.g., COPD affects over half of the lung cancer patients) and constitute a major predictor of survival (2). Another predictor of survival and quality of life is sarcopenia (11). The quantitative measurement of skeletal muscle cross-sectional area has been shown to correlate with clinical outcome parameters (such as mortality and postoperative bleeding) in patients with aortic valve replacement (12). A first study investigating radiomics features of skeletal muscle tissue gained additional information compared to more simplistic imaging biomarkers by discussing the correlation with intramuscular fat deposition (13). Predictive risk scores calculated from the combination of imaging features and the patient history may enable a more thorough evaluation of an individual case, thus leading to a different treatment decision compared to the current staging methods (figure 5). Since the validity and reliability of these imaging features can be tested independently across several diseases, their benefit extends beyond oncological applications.

With a standard chest CT scan, both the thorax and parts of the upper abdomen are open for the quantification of essential findings (see figure 6 and table 1 in the supplemental material for a list of imaging biomarkers).

Examples of three diseases show the importance of (radiomics derived) imaging biomarkers in chest CTs:

COPD is considered one of the most prevalent pulmonary diseases in Europe, contributing in important ways to a morbidity rate of 4-10% of the adult population who are affected, and resulting in over €20 billion in direct healthcare costs per year (14). There are prominent changes in the chest CTs of patients who suffer from COPD, such as emphysema and airway wall thickening (figure 7). By quantitatively evaluating these changes, both temporally and spatially, a new way of personalized risk assessment may be developed.

As an example, emphysema and airway wall thickening found on chest CTs acquired for non-pulmonary indications have shown to be independent predictors of COPD exacerbations that led to hospitalization or death (8). In (12), the imaging features were well known radiological findings and were graded visually; however, other authors showed that with a specialized IT-infrastructure, an automatic extraction and quantification was also feasible (15).

Previous research was focused on the densitometry of emphysema and differing approaches to airway measurements (e.g., the full-width-at-half-maximum principle) as quantitative CT imaging correlates of COPD (16). Bronchiectasis and mucous plugging are further factors that influence the appearance of a lung with COPD that could be quantified, but which are excluded from evaluation by focusing on a specific pathological manifestation. The radiomics approach, on the other hand, aims to mine all the information from an image. Indeed, a texture-based analysis was able to achieve better results than a densitometric approach (17).

Whereas the above mentioned articles consider pathomorphological imaging changes in their feature analysis, it is also possible to visually capture patient clusters with similar pathophysiological information (e.g., lung function) (18). A further example for the discovery of new subgroups of COPD is given by a study that used an unsupervised approach to find distinct subtypes of patients that correlate with physiological parameters based on imaging features (19). Additional characterization of

COPD-groups may be achieved by integrating features that represent spatial information (20). Extrapulmonary imaging features, such as thoracic fat, cardiac features and osteoporosis, are also important for the characterization of a patient with COPD (21).

Osteoporosis is not only a comorbidity of COPD but also one of the most common metabolic disorders that constitutes a major risk factor for vertebral fractures (22). A standard way to assess bones is the measurement of bone mineral density (BMD), which is limited in its capability to define a population at risk (23). To adequately predict vertebral strength, it is essential to consider several factors, which has led to the introduction of the term “bone quality” that comprises bone architecture, mineralization, and turnover (24). With this, the focus lies on the efficient use of BMD, not only the absolute amount. With the help of micro-CT studies, a multitude of trabecular bone parameters has been established, such as intra-vertebral heterogeneity of BMD, which is able to predict vertebral strength and stiffness, and ultimately, vertebral failure patterns (25).

The evaluation of intra-vertebral heterogeneity in a larger population requires the corresponding features to be accessible in medium- or low-dose CTs. Indeed, a recent study of native chest CTs showed a high reliability for the manual quantification of vertebral attenuation values (26).

However, the problem with the dual use of CT scans (i.e., opportunistically screening patients for osteoporosis in CT scans acquired for another purpose) is the low standardization and consequent bias of density values. To overcome this problem, several studies used texture analysis, resulting in a superior classification performance by combining texture features and density values compared to density values alone (27). A retrospective analysis demonstrates the capabilities of a machine learning algorithm in predicting vertebrae that showed fractures from the baseline scan (28). Possible bias through the application of contrast agents was counteracted by choosing specific texture features with stable predictive performance (29).

Interstitial lung diseases (ILDs) comprise entities that are very hard to diagnose even for experienced radiologists. While inter-observer agreement is low between radiologists, the discussion of these cases

in interdisciplinary boards can reduce mortality by identifying patients with a disease course that may be positively influenced by the adequate treatment (9). Taking into consideration that patients with ILDs commonly receive chest-CT scans repeatedly over several years, this provides an ideal situation for applying radiomics analyses and enables the possibility of studying the temporal progression of relevant features. Based on images alone, Walsh et al. (30) were able to demonstrate a slight superiority of a deep learning based algorithm in diagnosing fibrotic ILDs compared to radiologists. Existing tools, such as the Computer-Aided Lung Informatics for Pathology Evaluation and Rating (CALIPER) software (31) were successfully utilized to extract additional information and stratify patients to groups that have similar clinical phenotypes (18). Stratifying patients may discover those that benefit from a treatment, for example by selecting glucocorticoid-sensitive patients in ILDs (32), predict survival in idiopathic pulmonary fibrosis (33), or aide in assessing the occurrence of radiation pneumonitis (34).

For ongoing research it will be important to determine possible biological and technical biases, the stability of radiomics features and how to deal with these issues.

Challenges and possible solutions:

Differences in extracted features across scanner manufacturers and scanner models were tested with physical phantoms in (35), and showed strong variations in the features extracted, but also differences based on the feature types. On the same data, it was shown that the features can be normalized across scanners and that deep-learning-based features can be trained for a certain degree of invariance to scanner types (36).

When retrospective studies require data from several different scanner types and generations, it is important to ensure that features extracted from the images are comparable. Learning from older, retrospective data is one of the main applications where radiomics features are considered very useful and it must be ensured that such results can be transferred correctly.

One source of retrospective data are clinical trials, even-though only 3% of chest CT scans for an oncological indication are from clinical trials (37). To develop reproducible and valid signatures from imaging features of the lung for a broad population predictors of variation need to be considered (e.g., scanner-, protocol- and subject-specific factors of which the influence is typically minimized or avoided altogether in clinical trials (38)), as discussed in the following sections.

Inter-scanner reliability:

Features of lung cancer extracted from a lung-cancer phantom and lung-cancer patients showed reliable results between the phantom and the patients. However, there was a high variability between scanners from different vendors (35). In the same study, the feature "texture strength" was proven to remain stable, even between different scanners.

Voxel size / slice thickness

Although voxel size and slice thickness introduce a large variability, voxel size resampling and the use of normalizing factors was shown to be a way to increase the reproducibility of features from CT scans with different voxel sizes (39).

Dose level

Dose levels and reconstruction algorithms may influence imaging features. Interestingly, the interaction of these two technical parameters can be exploited to increase the reliability of the imaging features, as the variability introduced by lower dose levels is balanced out by using smoother reconstruction algorithms (40). A more detailed evaluation of the tube currents influence on radiomics features showed only a low contribution to variability (41). Therefore, it may be plausible to attribute the change to other factors, such as the tube voltage.

Contrast agent and time after injection:

A distortion of the texture features of pulmonary nodules caused by variation in contrast enhancement may be avoided by acquiring images in the time frame between 60 and 150 seconds post injection (42).

Another study suggests an increase in diagnostic performance by not administering contrast agent at all for solid pulmonary nodules (43).

Reconstruction algorithm:

Whereas texture features have a strong variability with respect to image reconstruction algorithms (40), shape- and size-based features remain table (44), further showing the importance of choosing adequate imaging feature sets for different preconditions.

The analysis of such predictors, in turn, is possible only with an adequate standardization. Whereas biological variability between patients and disease phenotypes cannot be controlled, technical parameters can be easily adjusted before a scan. In the future, scanning protocols might not only be adjusted for human readers but also for specific machine learning algorithms that could also use raw scanner data as an input. To discover technical priority targets for standardization, routine CT scans pose a valuable source of data that should be utilized.

Standardization and quality control of data acquisition

In the clinical routine, standardization will occur at a lower level than in studies with a fixed design. Therefore, it is important to dedicate further research to investigate the stability and susceptibility of imaging features to varying acquisition parameters. One option is to resample voxel sizes (39). Another option may be to conduct a test-retest analysis specifically designed for the respective study (45).

Quality control to ensure the scientific integrity of radiomics studies can be easily found in the public domain, including a digital phantom of reference features (46), publicly available images from physical lung phantoms (47), and a quality score to rate the design of radiomics studies (48). A higher quality can be achieved by conducting segmentation (semi-) automatically, as this has a higher accuracy than manual segmentation (49). An additional benefit of automation is the reduced amount of working time, which is pivotal in the light of sample sizes that could possibly reach four or more digits.

Standardization of radiomics features

A key value of radiomics analysis is the comparability of features across feature implementations, studies, centers, or routine imaging institutions. Initiatives to enable this by standardizing feature sets exist in several directions. Implementations of visual features have been shown to vary in important ways (50). The IBSI (Image Biomarker Standardization Initiative) (51) has tried to limit differences in such implementations by iteratively reducing differences in the implementations of several participants, as well as finding the mistakes or differences in the implementations when comparing the values of features. This has begun for simple features and is underway for more complex frequency-based features.

Algorithmic approaches to improve feature stability despite image heterogeneity and test-retest-stability

A different approach is to train models to become invariant or robust in the face of imaging heterogeneity. The step after feature extraction is the identification of features or feature-sets that hold predictive information. To do so, selection algorithms use probabilistic frameworks or deep-learning techniques to generate a prognostic model with high-dimensional feature spaces linked to prediction targets. This way of machine-learning uses a first data set for training and at least a second one for validation, further increasing the potential need for large sample sizes. Optimization of such classification methods reduces performance variation by 30% (52), and therefore, increases the stability and predictive value of imaging features selected in this way.

Domain adaptation is a popular deep-learning approach to reduce the susceptibility of a classifier toward unwanted heterogeneous variation. During training of a domain adaptation network, a representation is learned that is discriminative toward the desired goal (e.g., detection between diseased and healthy) yet invariant to undesired heterogeneity (e.g., 1- and 3-mm slice thickness) or domains. Kamnitsas et al. (53) offer one example of such an approach. Their domain adaptation algorithm was able to segment traumatic brain injuries in scans that were recorded with a different

protocol than the images in the training dataset. Another way to increase feature stability in heterogeneous datasets is to guide feature selection based on the stability in test re-test experiments of similar datasets, as described in (54).

Evolution of radiomics and deep-learning

Domain knowledge is one of the most valuable inputs that radiologists have when collaborating with data scientists. In this regard, domain knowledge comprises both the knowledge about pathophysiological and technical factors that contributed to the final image. Traditional machine-learning approaches rely heavily on the design of "handcrafted" features, with the advantage that these handcrafted features may code domain knowledge and are often easier to interpret or even standardize than more abstract features (even though many standard texture features remain difficult to interpret). Alternatively, modern deep-learning methods aim at end-to-end (e.g., raw-image to desired classification) training, inferring image features automatically based on the available data and annotations (thus, the steps of feature extraction and feature modelling during the radiomics workflow are done inside of a "deep learning black box"). These features can usually be extracted among the last layers of a network but are often difficult to interpret. Traditional radiomics aims at translating medical images to quantitative form with features that can be interpreted, for example, texture features, such as heterogeneity and entropy based on co-occurrence matrices. Domain knowledge can thus be included directly, meaning that fewer training data are required to obtain good results.

Deep-learning-based radiomics, on the other hand, has several advantages (36), for example, building features to be invariant to scanner variations. The inconvenience of deep-learning is the necessity of large amounts of data to learn very complex models. Further, the amount of data required for training is difficult to estimate. Domain knowledge cannot easily be integrated, and at the moment deep-learning is seen as a black-box model with low interpretability, even though approaches to deal with this exist (55) and are rapidly gaining momentum driven by the necessity of explainability and interpretability of deep learning models (56). Thus a transition of the field to deep learning, while at

the same time improving the explainability, and the capacity to integrate domain knowledge in these approaches can be expected.

Image repositories to foster algorithmic advances

Efficient ways to utilize routine CT images can enable access to large amounts of imaging data. A further benefit of routine data is the possible investigation of important comorbidities that are common in most patients in conjunction with the primary disease and that may have an impact on the subsequent therapy. Sampling from the clinical routine population may also reduce the selection bias, which can be an inherent problem in designed studies.

However, study cohorts may initially be beneficial when a high standardization enables concentration on fewer unknown variables, such as the influence of technical parameters on imaging features. Promising sets of features with a prediction target may then be validated on datasets from the clinical routine. The transfer of such predictive models can be facilitated via domain adaptation techniques and transfer-learning, making the best possible use of all the knowledge gained from the source-data (the study cohort) to highlight features that are not applicable, need adaptation, or are stable in the target-data (the imaging data from the clinical routine).

Publicly accessible databases comprising several thousand CT scans (57) and similarly large trial cohorts (58) provide the necessary standardization and adequate sample size, and are open for the high-throughput exploitation of possible future imaging biomarkers (see Table 2 in the supplemental material). In these trial cohorts, outcome parameters are available for most patients, often for several follow-up evaluations, including cardiovascular and pulmonary comorbidities, diagnosis and stage of lung cancer, and mortality statistics. Exploiting these homogeneous data from organized data repositories and trial cohorts can greatly add to the feasibility of radiomics studies.

Conclusion:

Radiomics in chest CT studies provides unprecedented opportunities for the discovery of markers for detection, classification, monitoring and prediction of oncological and non- oncological diseases alike. However, merely reporting the accuracy of algorithms from a given imaging dataset does not allow useful conclusions to be drawn for future studies. More important than demonstrating feasibility is the analysis of under what circumstances such a feat is possible. This includes an analysis of which technical and biological parameters may influence the outcome. Are the different steps of radiomics stable between slice thicknesses or in different respiratory phases? Should regions prone to motion, such as near the diaphragm or the heart, be excluded from analysis? What is the influence of time after contrast agent application or flow? Even if there is no external validation of the results (i.e., from a different center or at least scanner), such an analysis will contribute knowledge that is important to move the field forward. External validation, however, is pivotal to ensure that the whole process of a radiomics analysis is reproducible under different circumstances and cannot be omitted for the final conclusion of clinical relevance of a specific application. Exploiting large-scale data repositories and existing study cohorts of screening trials may enable the utilization of radiomics for, and its implementation in, the analysis of data from the clinical routine.

As mentioned above, quality guidelines and scores are already available and support researchers in planning their radiomics analysis accordingly. By following a process of constant steps forward, rather than jumping to the supposedly final algorithm, it will be possible to create a knowledge base of technical and biological preconditions that are necessary for a stable and reproducible radiomics analysis. Such a knowledge base could then, for instance, be used to tailor scanning protocols for a specific clinical question.

Creating more available data sets for a variety of diseases is another important step to advance radiomics, as these data sets can be used for model training and for evaluating generalizability of features on independent data. TCIA (The Cancer Imaging Archive) and TCGA (The Cancer Genome Atlas) are important steps in this direction but it is important to frequently check with images from new machines to assure that the algorithms work well also on heterogeneous data.

References

1. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. Radiology. 2016 Feb;278(2):563–77.

2. Islam KMM, Jiang X, Anggondowati T, Lin G, Ganti AK. Comorbidity and Survival in Lung Cancer Patients. Cancer Epidemiol Biomarkers Prev. 2015 Jul;24(7):1079–85.

3. Computed tomography (CT) exams [Internet]. OECD; Available from: https://www.oecd-ilibrary.org/social-issues-migration-health/computed-tomography-ct-exams/indicator/english_3c994537-en

4. Afshar P, Mohammadi A, Plataniotis KN, Oikonomou A, Benali H. From Handcrafted to Deep-Learning-Based Cancer Radiomics: Challenges and Opportunities [Internet]. Vol. 36, IEEE Signal Processing Magazine. 2019. p. 132–60. Available from: http://dx.doi.org/10.1109/msp.2019.2900993

5. Humphries SM, Yagihashi K, Huckleberry J, Rho B-H, Schroeder JD, Strand M, et al. Idiopathic Pulmonary Fibrosis: Data-driven Textural Analysis of Extent of Fibrosis at Baseline and 15-Month Follow-up [Internet]. Vol. 285, Radiology. 2017. p. 270–8. Available from: http://dx.doi.org/10.1148/radiol.2017161177

6. Abramson RG, Burton KR, Yu J-PJ, Scalzetti EM, Yankeelov TE, Rosenkrantz AB, et al. Methods and Challenges in Quantitative Imaging Biomarker Development [Internet]. Vol. 22, Academic Radiology. 2015. p. 25–32. Available from: http://dx.doi.org/10.1016/j.acra.2014.09.001

7. Krizhevsky A, Sutskever I, Hinton GE. ImageNet Classification with Deep Convolutional Neural Networks. In: Pereira F, Burges CJC, Bottou L, Weinberger KQ, editors. Advances in Neural Information Processing Systems 25. Curran Associates, Inc.; 2012. p. 1097–105.

8. Jairam PM, van der Graaf Y, Lammers J-WJ, Mali WPTM, de Jong PA, PROVIDI Study group. Incidental findings on chest CT imaging are associated with increased COPD exacerbations and mortality. Thorax. 2015 Aug;70(8):725–31.

9. Walsh SLF, Wells AU, Desai SR, Poletti V, Piciucchi S, Dubini A, et al. Multicentre evaluation of multidisciplinary team meeting agreement on diagnosis in diffuse parenchymal lung disease: a case-cohort study. Lancet Respir Med. 2016 Jul;4(7):557–65.

10. Xu Y, van Beek EJR, Hwanjo Y, Guo J, McLennan G, Hoffman EA. Computer-aided classification of interstitial lung diseases via MDCT: 3D adaptive multiple feature method (3D AMFM). Acad Radiol. 2006 Aug;13(8):969–78.

11. Prado CMM, Lieffers JR, McCargar LJ, Reiman T, Sawyer MB, Martin L, et al. Prevalence and clinical implications of sarcopenic obesity in patients with solid tumours of the respiratory and gastrointestinal tracts: a population-based study. Lancet Oncol. 2008 Jul;9(7):629–35.

12. Nemec U, Heidinger B, Sokas C, Chu L, Eisenberg RL. Diagnosing Sarcopenia on Thoracic Computed Tomography: Quantitative Assessment of Skeletal Muscle Mass in Patients Undergoing Transcatheter Aortic Valve Replacement. Acad Radiol. 2017 Sep;24(9):1154–61.

13. de Jong EEC, Sanders KJC, Deist TM, van Elmpt W, Jochems A, van Timmeren JE, et al. Can radiomics help to predict skeletal muscle response to chemotherapy in stage IV non-small cell lung cancer? Eur J Cancer. 2019 Oct;120:107–13.

14. Mannino DM, Sonia Buist A. Global burden of COPD: risk factors, prevalence, and future trends.

Lancet. 2007;370(9589):765–73.

15. Mohamed Hoesein FA, de Jong PA. Landmark papers in respiratory medicine: Automatic quantification of emphysema and airways disease on computed tomography. Breathe (Sheff). 2016 Mar;12(1):79–81.

16. Mets OM, de Jong PA, van Ginneken B, Gietema HA, Lammers JWJ. Quantitative computed tomography in COPD: possibilities and limitations. Lung. 2012 Apr;190(2):133–45.

17. Sorensen L, Nielsen M, Lo P, Ashraf H, Pedersen JH, de Bruijne M. Texture-Based Analysis of COPD: A Data-Driven Approach. IEEE Trans Med Imaging. 2012;31(1):70–8.

18. Raghunath S, Rajagopalan S, Karwoski RA, Maldonado F, Peikert T, Moua T, et al. Quantitative stratification of diffuse parenchymal lung diseases. PLoS One. 2014 Mar 27;9(3):e93229.

19. Binder P, Batmanghelich NK, Estepar RSJ, Golland P. Unsupervised Discovery of Emphysema Subtypes in a Large Clinical Cohort. Mach Learn Med Imaging. 2016 Oct;10019:180–7.

20. Yang J, Angelini ED, Balte PP, Hoffman EA, Austin JHM, Smith BM, et al. Unsupervised Discovery of Spatially-Informed Lung Texture Patterns for Pulmonary Emphysema: The MESA COPD Study. Med Image Comput Comput Assist Interv. 2017 Sep;10433:116–24.

21. Bhatt SP, Washko GR, Hoffman EA, Newell JD Jr, Bodduluri S, Diaz AA, et al. Imaging Advances in Chronic Obstructive Pulmonary Disease. Insights from the Genetic Epidemiology of Chronic Obstructive Pulmonary Disease (COPDGene) Study. Am J Respir Crit Care Med. 2019 Feb 1;199(3):286–301.

22. O'Neill TW, Felsenberg D, Varlow J, Cooper C, Kanis JA, Silman AJ. The prevalence of vertebral deformity in european men and women: the European Vertebral Osteoporosis Study. J Bone Miner Res. 1996 Jul;11(7):1010–8.

23. Kanis JA, Johnell O, Oden A, De Laet C, Jonsson B, Dawson A. Ten-year risk of osteoporotic fracture and the effect of risk factors on screening strategies. Bone. 2002 Jan;30(1):251–8.

24. NIH Consensus Development Panel on Osteoporosis Prevention, Diagnosis, and Therapy, March 7-29, 2000: highlights of the conference. South Med J. 2001 Jun;94(6):569–73.

25. Hussein AI, Morgan EF. The effect of intravertebral heterogeneity in microstructure on vertebral strength and failure patterns. Osteoporos Int. 2013 Mar;24(3):979–89.

26. Pompe E, de Jong PA, de Jong WU, Takx RAP, Eikendal ALM, Willemink MJ, et al. Inter-observer and inter-examination variability of manual vertebral bone attenuation measurements on computed tomography. Eur Radiol. 2016 Sep;26(9):3046–53.

27. Valentinitsch A, Trebeschi S, Kaesmacher J, Lorenz C, Löffler MT, Zimmer C, et al. Opportunistic osteoporosis screening in multi-detector CT images via local classification of textures. Osteoporos Int. 2019 Jun;30(6):1275–85.

28. Muehlematter UJ, Mannil M, Becker AS, Vokinger KN, Finkenstaedt T, Osterhoff G, et al. Vertebral body insufficiency fractures: detection of vertebrae at risk on standard CT images using texture analysis and machine learning. Eur Radiol. 2019 May;29(5):2207–17.

29. Mookiah MRK, Rohrmeier A, Dieckmeyer M, Mei K, Kopp FK, Noel PB, et al. Feasibility of opportunistic osteoporosis screening in routine contrast-enhanced multi detector computed tomography (MDCT) using texture analysis. Osteoporos Int. 2018 Apr;29(4):825–35.

30. Walsh SLF, Calandriello L, Silva M, Sverzellati N. Deep learning for classifying fibrotic lung disease on high-resolution computed tomography: a case-cohort study. Lancet Respir Med. 2018 Nov;6(11):837–45.

31. Bartholmai BJ, Raghunath S, Karwoski RA, Moua T, Rajagopalan S, Maldonado F, et al. Quantitative Computed Tomography Imaging of Interstitial Lung Diseases [Internet]. Vol. 28, Journal of Thoracic Imaging. 2013. p. 298–307. Available from: http://dx.doi.org/10.1097/rti.0b013e3182a21969

32. Feng D-Y, Zhou Y-Q, Xing Y-F, Li C-F, Lv Q, Dong J, et al. Selection of glucocorticoid-sensitive patients in interstitial lung disease secondary to connective tissue diseases population by radiomics. Ther Clin Risk Manag. 2018 Oct 10;14:1975–86.

33. Maldonado F, Moua T, Rajagopalan S, Karwoski RA, Raghunath S, Decker PA, et al. Automated quantification of radiological patterns predicts survival in idiopathic pulmonary fibrosis. Eur Respir J. 2014 Jan;43(1):204–12.

34. Cunliffe A, Armato SG 3rd, Castillo R, Pham N, Guerrero T, Al-Hallaq HA. Lung texture in serial thoracic computed tomography scans: correlation of radiomics-based features with radiation therapy dose and radiation pneumonitis development. Int J Radiat Oncol Biol Phys. 2015 Apr 1;91(5):1048–56.

35. Mackin D, Fave X, Zhang L, Fried D, Yang J, Taylor B, et al. Measuring Computed Tomography Scanner Variability of Radiomics Features. Invest Radiol. 2015 Nov;50(11):757–65.

36. Andrearczyk V, Depeursinge A, Müller H. Standardization of features extracted from CT images of texture phantoms. In: SPIE Medical Imaging 2019. to appear 2019.

37. Oncology society rolls out big-data initiative, tells why radiology should care [Internet]. Imaging Biz. [cited 2018 Sep 17]. Available from: https://www.imagingbiz.com/topics/imaging-informatics/oncology-society-rolls-out-big-data-initiative-tells-why-radiology

38. Ruben T H, Defraene G, De Ruysscher D, Lambin P, van Elmpt W. Quantitative radiomics studies for tissue characterization: a review of technology and methodological procedures. Br J Radiol [Internet]. 2017 Feb [cited 2018 Sep 17];90(1070). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5685111/

39. Larue RTHM, van Timmeren JE, de Jong EEC, Feliciani G, Leijenaar RTH, Schreurs WMJ, et al. Influence of gray level discretization on radiomic feature stability for different CT scanners, tube currents and slice thicknesses: a comprehensive phantom study. Acta Oncol. 2017 Nov;56(11):1544–53.

40. Lo P, Young S, Kim HJ, Brown MS, McNitt-Gray MF. Variability in CT lung-nodule quantification: Effects of dose reduction and reconstruction methods on density and texture based features. Med Phys. 2016 Aug;43(8):4854.

41. Mackin D E al. Effect of tube current on computed tomography radiomic features. - PubMed - NCBI [Internet]. [cited 2018 Sep 17]. Available from: https://www.ncbi.nlm.nih.gov/pubmed/29403060

42. Kim H, Park CM, Park SJ, Song YS, Lee JH, Hwang EJ, et al. Temporal Changes of Texture Features Extracted From Pulmonary Nodules on Dynamic Contrast-Enhanced Chest Computed Tomography: How Influential Is the Scan Delay? Invest Radiol. 2016 Sep;51(9):569–74.

43. He L E al. Effects of contrast-enhancement, reconstruction slice thickness and convolution

kernel on the diagnostic performance of radiomics signature in soli... - PubMed - NCBI [Internet]. [cited 2018 Sep 17]. Available from: https://www.ncbi.nlm.nih.gov/pubmed/27721474/

44. Kim H, Park CM, Lee M, Park SJ, Song YS, Lee JH, et al. Impact of Reconstruction Algorithms on CT Radiomic Features of Pulmonary Tumors: Analysis of Intra- and Inter-Reader Variability and Inter-Reconstruction Algorithm Variability. PLoS One. 2016 Oct 14;11(10):e0164924.

45. van Timmeren JE, Leijenaar RTH, van Elmpt W, Wang J, Zhang Z, Dekker A, et al. Test–Retest Data for Radiomics Feature Stability Analysis: Generalizable or Study-Specific? Tomography. 2016 Dec;2(4):361.

46. Website [Internet]. [cited 2018 Sep 17]. Available from: Lambin P. Radiomics Digital Phantom, https://www.cancerdata.org/resource/doi:10.17195/candat.2016.08.1. 2016.

47. Kalendralis P, Traverso A, Shi Z, Zhovannik I, Monshouwer R, Starmans MPA, et al. Multicenter CT phantoms public dataset for radiomics reproducibility tests. Med Phys. 2019 Mar;46(3):1512–8.

48. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, et al. Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol. 2017;14(12):749–62.

49. Parmar C, Rios Velazquez E, Leijenaar R, Jermoumi M, Carvalho S, Mak RH, et al. Robust Radiomics feature quantification using semiautomatic volumetric segmentation. PLoS One. 2014 Jul 15;9(7):e102107.

50. Kalpathy-Cramer J, Mamomov A, Zhao B, Lu L, Cherezov D, Napel S, et al. Radiomics of Lung Nodules: A Multi-Institutional Study of Robustness and Agreement of Quantitative Imaging Features. Tomography. 2016 Dec;2(4):430–7.

51. Zwanenburg A, Leger S, Vallières M, Löck S, Initiative, for the Image Biomarker Standardisation. Image biomarker standardisation initiative [Internet]. 2016 [cited 2019 Jan 28]. Available from: http://arxiv.org/abs/1612.07003

52. Parmar C, Grossmann P, Rietveld D, Rietbergen MM, Lambin P, Aerts HJWL. Radiomic Machine-Learning Classifiers for Prognostic Biomarkers of Head and Neck Cancer. Front Oncol. 2015 Dec 3;5:272.

53. Kamnitsas K, Baumgartner C, Ledig C, Newcombe V, Simpson J, Kane A, et al. Unsupervised Domain Adaptation in Brain Lesion Segmentation with Adversarial Networks. In: Lecture Notes in Computer Science. 2017. p. 597–609.

54. Aerts HJWL, Velazquez ER, Leijenaar RTH, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun. 2014 Jun 3;5:4006.

55. Graziani M, Andrearczyk V, Müller H. Regression Concept Vectors for Bidirectional Explanations in Histopathology. In: Lecture Notes in Computer Science. 2018. p. 124–32.

56. Kendall A, Gal Y. What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision? In: Advances in Neural Information Processing Systems. 2017. p. 5574–84.

57. Armato SG 3rd, McLennan G, Bidaut L, McNitt-Gray MF, Meyer CR, Reeves AP, et al. The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): a completed reference database of lung nodules on CT scans. Med Phys. 2011 Feb;38(2):915–31.

58. Pastorino U, Rossi M, Rosato V, Marchianò A, Sverzellati N, Morosi C, et al. Annual or biennial CT screening versus observation in heavy smokers: 5-year results of the MILD trial. Eur J Cancer Prev. 2012 May;21(3):308–15.

59. Kolossváry M, Park J, Bang J-I, Zhang J, Lee JM, Paeng JC, et al. Identification of invasive and radionuclide imaging markers of coronary plaque vulnerability using radiomic analysis of coronary computed tomography angiography. Eur Heart J Cardiovasc Imaging. 2019 Nov 1;20(11):1250–8.

60. Kolossváry M, Karády J, Szilveszter B, Kitslaar P, Hoffmann U, Merkely B, et al. Radiomic Features Are Superior to Conventional Quantitative Computed Tomographic Metrics to Identify Coronary Plaques With Napkin-Ring Sign. Circ Cardiovasc Imaging [Internet]. 2017 Dec;10(12). Available from: http://dx.doi.org/10.1161/CIRCIMAGING.117.006843

61. Oikonomou EK, Williams MC, Kotanidis CP, Desai MY, Marwan M, Antonopoulos AS, et al. A novel machine learning-derived radiotranscriptomic signature of perivascular fat improves cardiac risk prediction using coronary CT angiography. Eur Heart J. 2019 Nov 14;40(43):3529–43.

62. Lafata KJ, Zhou Z, Liu J-G, Hong J, Kelsey CR, Yin F-F. An Exploratory Radiomics Approach to Quantifying Pulmonary Function in CT Images. Sci Rep. 2019 Aug 8;9(1):11509.

63. Yanling W, Duo G, Zuojun G, Zhongqiang S, Yankai W, Shan L, et al. Radiomics Nomogram Analyses for Differentiating Pneumonia and Acute Paraquat Lung Injury. Sci Rep. 2019 Oct 21;9(1):15029.

64. Shi W, Zhou L, Peng X, Ren H, Wang Q, Shan F, et al. HIV-infected patients with opportunistic pulmonary infections misdiagnosed as lung cancers: the clinicoradiologic features and initial application of CT radiomics. J Thorac Dis. 2019 Jun;11(6):2274–86.

65. Wang B, Li M, Ma H, Han F, Wang Y, Zhao S, et al. Computed tomography-based predictive nomogram for differentiating primary progressive pulmonary tuberculosis from community-acquired pneumonia in children. BMC Med Imaging. 2019 Aug 8;19(1):63.

66. Ryan SM, Fingerlin TE, Mroz M, Barkes B, Hamzeh N, Maier LA, et al. Radiomic measures from chest high-resolution computed tomography associated with lung function in sarcoidosis. Eur Respir J [Internet]. 2019 Aug;54(2). Available from: http://dx.doi.org/10.1183/13993003.00371-2019

67. Krafft SP, Rao A, Stingo F, Briere TM, Court LE, Liao Z, et al. The utility of quantitative CT radiomics features for improved prediction of radiation pneumonitis. Med Phys. 2018 Nov;45(11):5317–24.

68. Moran A, Daly ME, Yip SSF, Yamamoto T. Radiomics-based Assessment of Radiation-induced Lung Injury After Stereotactic Body Radiotherapy. Clin Lung Cancer. 2017 Nov;18(6):e425–31.

69. Szigeti K, Szabó T, Korom C, Czibak I, Horváth I, Veres DS, et al. Radiomics-based differentiation of lung disease models generated by polluted air based on X-ray computed tomography data. BMC Med Imaging. 2016 Feb 11;16:14.

70. Jacob J, Bartholmai BJ, Rajagopalan S, Kokosi M, Egashira R, Brun AL, et al. Serial automated quantitative CT analysis in idiopathic pulmonary fibrosis: functional correlations and comparison with changes in visual CT scores. Eur Radiol. 2018 Mar;28(3):1318–27.

71. Humphries SM, Yagihashi K, Huckleberry J, Rho B-H, Schroeder JD, Strand M, et al. Idiopathic Pulmonary Fibrosis: Data-driven Textural Analysis of Extent of Fibrosis at Baseline and 15-Month

Follow-up. Radiology. 2017 Oct;285(1):270–8.

72. Kloth C, Henes J, Xenitidis T, Thaiss WM, Blum AC, Fritz J, et al. Chest CT texture analysis for response assessment in systemic sclerosis. Eur J Radiol. 2018 Apr;101:50–8.

73. Kloth C, Blum AC, Thaiss WM, Preibsch H, Ditt H, Grimmer R, et al. Differences in Texture Analysis Parameters Between Active Alveolitis and Lung Fibrosis in Chest CT of Patients with Systemic Sclerosis: A Feasibility Study. Acad Radiol. 2017 Dec;24(12):1596–603.

Figure legends:

Figure 1: Radiomics workflow – radiomics starts with the image acquisition (including the reconstruction), followed by the localisation and segmentation of the region or volume of interest for the subsequent feature extraction (these can either be handcrafted or deep learning-based). By analysing the stability, discriminative power and redundancy of imaging features, the appropriate ones can be selected (this is often done with machine learning-based methods). The process of radiomics may end with a signature of imaging features that provide prognostic information. For an efficient implementation of radiomics in large datasets, and eventually, in routine clinical data, each step should be reproducible with a minimum of manual input.

Figure 2: A blood sample needs to be treated with additives, which is followed by a specific procedure for processing and analysing the sample to yield a quantitative laboratory report. Comparably, the acquisition parameters of a CT scan and the following procedure of segmentation, feature extraction, and selection (or alternatively, feature modelling) must be chosen depending on the desired quantitative radiological results. In the same way as the development of laboratory medicine took many years, with the advances of machine-learning techniques and rigorous quality management, a radiology report that includes quantitative information from automated feature extraction may be achievable.

Figure 3: Patterns of idiopathic interstitial pneumonias: a) traction bronchiectasis; b) ground glass; c) honeycombing; d) irregular reticulation.

Figure 4: Example of disease progression in IPF using radiomics features: a) axial CT-scan at baseline and b) at follow-up; c) visualization of radiomics derived features at baseline and d) at follow-up. The color-coding allows a comparison with known visual features, such as ground-glass opacifications, bronchiectasis or distorted pulmonary vessels. Radiomics allows to detect and quantify both features that may have a known correlate, and complex features that can only be explained mathematically

without a visually accessible correlate.

IPF = idiopathic pulmonary fibrosis

Figure 5: Radiomics as part of advanced feature modelling - quantitative imaging biomarkers that already exist may be integrated in such a model. The added benefit of using a multitude of features that can be extracted by using a radiomics approach is given by the possibility of analysing these features in conjunction with the information from other sources.

For instance, there may be imaging features that provide the same kind of information as other imaging- or clinical features, and thus, would be redundant. Such redundancy can be reduced by different machine learning techniques. Theoretically, known quantitative imaging biomarkers can be integrated into this model as well. However, this would constitute a prior selection of which features of the image are important and which are not, carrying the danger of omitting possibly relevant features.

Figure 6: Overview of structures and parameters in routine chest CTs that are already accessible for the development of biomarkers by machine-learning methods. This demonstrates the general possibility of finding additional biomarkers by “mining” imaging features with a radiomics approach, and the potential of machine-learning methods to facilitate the process.

Figure 7: a) Axial image of a healthy lung, b) a lung with centrilobular emphysema, and c) panlobular emphysema. Radiomic features may help to differentiate patients at risk of acute exacerbation or cardiovascular complications.

Study	# of patients	Anatomic site	Analyzed endpoint
Kolossváry et al. Nov 2019 (59)	25 (44 singular lesions)	Coronaries	Identification of unstable coronary plaques
Kolossváry et al. Dec 2017 (60)	30	Coronaries	Identification of unstable coronary plaques
Oikonomou et al. Nov 2019 (61)	167	Perivascular adipose tissue	Gene expression profiles of inflammation, fibrosis and vascularity in coronary heart disease
de Jong et al. Sep 2019 (13)	116	Skeletal muscle tissue	Prediction of sarcopenia
Lafata et al. Aug 2019 (62)	64	Lung	Quantification of pulmonary function
Yanling et al. Oct 2019 (63)	180	Lung	Differentiation of pneumonia vs. paraquat lung injury
Shi et al. Jun 2019 (64)	49	Lung	Identification of opportunistic pulmonary infections in HIV patients
Wang et al. Aug 2019 (65)	115	Lung	Differentiation of primary progressive pulmonary tuberculosis vs. community-acquired pneumonia
Ryan et al. Aug 2019 (66)	151	Lung	Classification of sarcoidosis stage
Krafft et al. Nov 2018 (67)	192	Lung	Prediction of radiation pneumonitis
Cunliffe et al. Apr 2015 (34)	106	Lung	Prediction of radiation pneumonitis
Moran et al. Nov 2017 (68)	14	Lung	Classification of severity of radiation-induced lung injury
Szigeti et al. Feb 2016 (69)	16 (mouse model)	Lung	Identification of pollution induced lung disease
Jacob et al. Mar 2018 (70)	66	Lung	Classification of idiopathic pulmonary fibrosis severity measured by pulmonary function
Maldonado et al. Jan 2014 (33)	55	Lung	Prediction of survival in idiopathic pulmonary fibrosis
Bartholmai et al. Sep 2013 (31)	14	Lung	Classification of interstitial lung disease severity as determined by various clinical parameters
Humphries et al. Oct 2017 (71)	280	Lung	Prediction of idiopathic pulmonary fibrosis severity measured by pulmonary function
Raghunath et al. Mar 2014 (18)	1322	Lung	Stratification of diffuse parenchymal lung diseases based on various clinical parameters
Feng et al. Oct 2018 (32)	416	Lung	Identification of glucocorticoid-sensitive patients in interstitial lung disease
Kloth et al. Apr 2018 (72)	23	Lung	Prediction of responders and non-responders of stem cell transplantation in systemic sclerosis
Kloth et al. Dec 2017 (73)	43	Lung	Differentiation of active alveolitis vs. fibrosis in patients with systemic sclerosis

Table 1 - Radiomics applications in chest imaging outside of tumors:
Most studies report on a small number of subjects in the 2-digit or low 3-digit range. Still, such studies are very important for the demonstration of feasibility as well as for disseminating methodological experience gained during the conduction of the study. For the development of valid, reliable and clinically relevant radiomics-derived biomarkers, larger, preferably externally validated projects are necessary.