Artificial Intelligence Will Improve Molecular Imaging, Therapy and Theranostics. Which Are the Biggest Advantages for Therapy?

12

Georgios Kaissis and Rickmer Braren

Contents

12.1 Introduction

Artificial Intelligence (AI) approaches in medical imaging have witnessed significant evolution over the past years. The reasons for this are manifold: The field of computer vision has arguably seen the most drastic advance in its state of the art facilitated by the increasingly widespread application of deep learning [1], the introduction of large, curated data sets facilitating transfer learning approaches [2], the substantial research and industry interest in the domain and the availability of both hardware accelerators (mainly graphics processing units) and software frameworks providing pretrained algorithms and approachable application programming interfaces lowering the barrier to entry to the field. Furthermore, medical imaging represents an excellent target for machine learning applications as it is widely available in standardized data exchange formats and stored electronically [3]. Also, the availability of images alongside medical/radiological reports provide inbuilt human ground-truth assessments of relevant findings.

G. Kaissis (✉)
Technical University of Munich, Munich, Germany

Imperial College London, London, UK
e-mail: g.kaissis@tum.de

R. Braren
Technical University of Munich, Munich, Germany
e-mail: rbraren@tum.de

The trend of large dataset accrual has increasingly also manifested in the medical field, with large databases of medical imaging data being assembled as national efforts attempting to provide a cross-sectional assessment of large populations of both healthy volunteers and patients. The German National Cohort Health Study (*NAKO Gesundheitsstudie, www.nako.de*) and the United Kingdom Biobank [4] are examples of this development, providing access to thousands of imaging data sets to researchers and practitioners in the field, which can be used for the development of machine learning algorithms. These efforts supplement initiatives such as the [5], representing curated collections of oncology-specific material including medical imaging but also digital histopathology or genomic sequence data. The increasing roll-out of partially or fully electronic patient records signifies a further important step towards the collection of relevant metadata, which can be included in predictive models alongside image-based information. However, such data repositories are not without specific challenges: Large-scale data collection signifies an increased importance of privacy protection, for which next-generation methods have only recently been introduced [6]. Moreover, data quality is paramount for the development of predictive algorithms, thus care needs to be taken that images and clinical metadata are generated and expertly curated with high standards of quality assurance. Algorithms need to be trained and validated on diverse and representative patient collectives to ascertain not only their validity when applied to unseen data from new sources, but also to assert their fairness, control their bias and render them reproducible and interpretable. The deployment of machine learning algorithms to clinical routine poses great challenges of its own, necessitating interdisciplinary cooperation and continuous monitoring and improvements. Finally, the reimbursement of algorithm-based diagnostic services remains largely unresolved. Issues such as these represent but a limited subset of the parameters which need to be taken into account in the design of artificial intelligence algorithms for medical use and are discussed in other parts of this book, as well as touched upon later in this chapter.

Expectedly for a novel field, most of the literature published on the topic of artificial intelligence applications in medical imaging has focused on diagnostic applications in the field of oncology such as the prediction of tumor subtypes, genetic features, metastatic behavior or patient survival. Algorithms targeted at diagnosis often provide objectively verifiable outputs (e.g. by comparison of the algorithm's prediction to a histopathologic result), and can be compared to the performance of human experts (e.g. true/false positive/negative rates), facilitating their validation. The field of therapy monitoring and theranostics, that is, the image-based expression quantification of relevant therapeutic targets, has however not yet witnessed the same level of research activity. Several reasons emerge, such as the following:

1. Treatment represents a heterogenous clinical process characterized by the application of several therapeutic approaches, often simultaneously. For instance, oncologic therapy consists of surgical, pharmacologic, radiotherapeutic, and other supplemental interventions. Establishing causal relationships between a certain treatment and its effect is therefore often a difficult undertaking.
2. The interplay between treatment and disease is hard to accurately quantify. For example, tumors demonstrate therapy escape phenomena leading to treatment resistance, which can be hard to distinguish from inefficacy or primary failure of the treatment.
3. Cancer imaging is influenced by systemic effects such as individual toxicity or comorbidities that can have a modulating effect on local findings (e.g. perfusion effects of anti-vascular agents versus decrease in cardiovascular output causing tissue mal-perfusion) and which are hard to deconvolve from specific treatment outcomes.
4. Effects mediating treatment response are also functions of the complex genetic, transcriptomic, epigenetic, and environmental tumor

landscape in which causes and effects can be impossible to distinguish.

5. Novel treatments are continuously introduced, thus retrospectively collected data, often the bedrock of oncological machine learning applications, might not be applicable as algorithm training material.

6. Finally, cancer is insufficiently understood and represents a disease as individual as the patients themselves. Intra- and inter-tumoral heterogeneity thus pose hindrances to the applicability of algorithmic tools aimed foremost at generalization, drastically increasing the difficulty of training such algorithms.

In attempting to taxonomically classify the current literature about machine learning and artificial intelligence approaches for treatment response prediction and assessment as well as theranostics, two patterns emerge:

- The majority of studies focus on the prediction of therapy response from a single timepoint and single surrogate. Such studies attempt to capture information from a singular imaging study, often the baseline examination, to predict differences in treatment outcome by characterizing a specific tumor phenotype.

- Studies focusing on longitudinal/integrative monitoring of findings, for example, integrating the features of the tumor alongside relevant metadata and/or their evolution over the treatment period to predict the course of therapy.

With respect to the defining tumor features, research can be stratified into studies aiming at the quantification of tumor volume, either purely morphological or morphological and metabolic, for example, by the definition and automated tracking of metabolic tumor volume, and into studies concerned with higher-order descriptors of disease features or treatment targets. Such features can be derived from the tumor itself, for example, histogram metrics, texture features etc.

and/or incorporate other data, such as clinical record information.

Finally, from a methodological point of view, research can be divided into studies applying traditional computer vision techniques by utilizing predefined mathematical descriptors of the image (features) alongside machine learning-methods typically used for tabular data analysis such as regression models, tree-based algorithms etc. and studies applying deep neural networks directly to the imaging data. For the former, the term radiomics is often used. We would like to point out that this distinction is not formal, and the term radiomics is used for deep-neural-network-based algorithms as well. Due to its ill definition, we eschew the usage of this term altogether and refer instead to the techniques and algorithms in question by their technical description, which we believe to be more both clearer and more informative.

The methodological concerns applied to a study are also a function of the data used for algorithm development. Unlike pure anatomic imaging, which typically takes the form of a three-dimensional stack of images in black and white, hybrid and functional imaging usually provides at least two congruent images for the same anatomical location. In case of dynamic acquisitions, such as multiple contrast media phases, the dimensionality of the data further increases. This data is often heterogeneous with respect to its spatial resolution (e.g. the technical resolution of the scanner or the effects resulting from interactions of radionuclides with the tissue leading to, for example, the actual resolution of PET differing from the nominal resolution of the detector elements). These factors need to be taken into account and potentially corrected for in quantitative imaging studies.

In the following sections we will highlight and contrast relevant literature findings regarding the application of machine learning to therapy response evaluation with a focus on hybrid oncological imaging and provide recommendations and future directions for practitioners and researchers in the field.

12.2 Literature Review

12.2.1 Morphological and Metabolic Tumor Volume Tracking

12.2.1.1 Volumetry-Based Oncological Response Assessment Frameworks

The conceptually simplest automated therapy surveillance approaches rely on the quantification of the reduction in tumor volume using automated methods, thus mirroring human evaluation, for example, by application of the Response Evaluation Criteria in Solid Tumors (RECIST). RECIST was among the first attempts to quantify tumor response to treatment in imaging. However, it relies on two-dimensional evaluation and on the definition of so-called target lesions, which necessarily limits its scope and potential representativeness, since individual tumor manifestations are employed as surrogates of disease burden. RECIST evaluation suffers from further notable limitations, mainly in tumor entities with ill-defined margins (e.g. pancreatic cancer) and can be a poor correlate of therapy response due to phenomena such as pseudo-progression, whereby tumor volume initially increases in response to therapy due to inflammatory changes. The 2009 position paper by Wahl et al. introduced a systematic framework combining previous guidelines for incorporating metabolic and functional imaging-derived information into tumor response assessment called PERCIST (PET response criteria in solid tumors). The PERCIST framework stipulates the categories complete and partial metabolic response, stable metabolic disease, and progressive metabolic disease by measurement of lean body mass-adjusted standardized uptake value (SUL). Similar frameworks have been proposed by other working groups, such as the EORTC, as well as combined functional/morphologic criteria such as the Lugano criteria proposed in 2014, incorporating elements of both RECIST and radionuclide uptake information.

The quantitative nature of PET allows the calculation of absolute radionuclide activity per volume tissue, offering benefits over the standardized uptake value, which has been shown to depend on several extraneous parameters. Thus, more recently, parameters like the total lesion glycolysis (TLG) and metabolic tumor volume (MTV) have been proposed as more precise biomarkers of disease activity. These however require a definition of the tumor volume itself, also termed segmentation.

12.2.1.2 Automated Segmentation-Based Volumetry Techniques

The evolution of automated volumetry methods thus closely follows the evolution of automated tumor segmentation methods. Earlier studies [7–9] rely on legacy segmentation techniques such as region-growing, nearest-neighbor or probabilistic graphical methods [10]. Hybrid imaging provides a benefit in this regard by providing a form of pre-segmentation mask via the high-SUV tumor region, helping to guide algorithm behavior. Such iso-contour-based segmentation methods [11] have been demonstrated, for example, in sarcoma. Similar approaches can also be applied directly to metabolic tumor volume (MTV) tracking without the associated morphological imaging. This approach has shown promise in several tumor entities such as rectal cancer [12], lymphoma [13], gynecological tumors [14], or esophageal cancer [15]. However, it has been noted that MTV lacks standardization and large-scale external validation and thus cannot be assumed to be a universal gold standard for therapy surveillance in comparison to, for example, the standardized uptake value (SUV) [16].

12.2.1.3 Evolution of Automated Segmentation Using Neural Networks

Automated segmentation has witnessed a substantial evolution with the introduction of neural network-based segmentation methods. Earlier methods, based on fully convolutional neural networks [17] have more recently been superseded by encoder–decoder architectures with transverse short-circuits, such as the UNet architecture proposed by Ronneberger et al. in 2015 [18] and their conceptual evolutions such as Feature Pyramid Networks (FPNs) [19]. A common trait of these architectures is the utilization of image

information captured at multiple scales and the transmission of high spatial frequency (i.e. high detail) image information from early to late parts of the network with corresponding feature map sizes. Encoder–decoder architectures have dominated the segmentation literature since ca. 2015, and can be applied both in two and three dimensions. Fully automatic segmentation has been proposed as a solution to the aforementioned standardization problem [20] and been successfully applied to both treatment response assessment, for example, in breast cancer [21], where it has been shown to outperform dynamic contrast-enhanced MRI, and treatment planning, for example, for brain tumor radiotherapy [22].

12.3   Quantitative Image and Texture Analysis in Oncological Therapy Response Monitoring

The advent of quantitative image analysis workflows within the past 5 years has generated significant interest in the utilization of image-derived data for tumor characterization. Such approaches rely on either the bulk extraction of tumor-related image features, their preprocessing and modeling using machine learning (also termed radiomics), or the end-to-end analysis of image data using neural networks. As discussed above, we will not terminologically differentiate between these approaches, believing them to not be mutually exclusive. However, it is expected that the numerous shortcomings of the so-called radiomics workflow will eventually lead to its replacement by algorithms and techniques based on more robust techniques and models, and not susceptible to the same technical limitations we will describe below. The typical workflow of quantitative image analysis studies is common to both approaches, consisting of a volume of interest definition step and a modeling step. For volume of interest definition i.e. segmentation, both manual and all above-mentioned automatic methods are applicable and commonly used. For details on the various techniques, we refer to the chapters in Part I of this book.

The research developments in the field of treatment supervision in hybrid imaging have closely followed the main oncologic application areas of PET.

12.3.1 Neuro-Oncology

In neuro-oncologic applications, for example, studies have focused on the identification of molecular phenotypes with relevance for therapy and prognosis, such as isocitrate dehydrogenase status [23] from amino acid (fluoroethyl tyrosine, FET) PET scans in gliomas. The authors found that the inclusion of radiomic parameters improved diagnostic accuracy compared to PET-derived metrics alone. Similarly, a recent study by Hotta et al. found image texture parameters derived from 11C-methionine PET to yield excellent discriminative performance between recurrence of malignant brain tumors and radiation necrosis [24], a topic of critical relevance for steering treatment decisions. A multitude of works (see e.g. overview in [25]) have focused on brain metastases, amongst others for differentiation of primary brain tumors from metastases, pinpointing the origin of metastatic lesions to the brain and for differentiating treatment-related changes from recurrence. Recent studies have also focused specifically on treatment, with studies by Cha et al. demonstrating strong performance of convolutional neural network ensembles in the prediction of metastatic lesion response to radiotherapy [26] from baseline imaging examinations.

12.3.2 Head and Neck Cancers

In head and neck cancers, several studies have demonstrated the benefits of integrating quantitative imaging features with morphological tumor descriptors for predictive modeling workflows. For instance, Fujima et al. showed that in patients who underwent chemoradiation treatment for pharyngeal cancer, tumor shape and texture features were highly predictive of progression-free and overall patient survival [27]. They note that

clinical parameters alone were not sufficient for discriminating survival subgroups in their study. Feliciani et al. employed texture metrics derived from pretherapeutic FDG-PET and found these imaging biomarkers highly predictive of local chemoradiation therapy failure [28]. Crispin-Ortuzar and colleagues aimed at predicting head and neck tumor hypoxia, which is usually assessed, for example, with specific hypoxia radiotracers such as 18F-FMISO, using FDG-PET-derived texture parameters. They report substantial improvements over baseline FDG-PET performance alone and note that quantitative imaging biomarkers can provide an alternative to hypoxia-specific radiotracers where such are unavailable [29].

12.3.3 Lung Cancer

In lung cancer, the relevance of including FDG-PET into patient workup was shown in the 2002 PLUS trial [30], demonstrating a 20% reduction in unnecessary surgical interventions. Consequently, several studies have investigated quantitative imaging features, for example, in the prediction of histological subtypes [31] or post-treatment survival [32]. Oikonomou et al. studied the association of quantitative image features with several outcomes, including local and distant disease control, recurrence-free probability and survival metrics and found image-derived features to represent the only predictors of overall survival, disease-specific survival and regional disease control [33]. A recent multicenter trial by Dissaux et al. demonstrated that FDG-PET-derived texture features predict local disease control in patients undergoing stereotactic radiotherapy for early-stage non-small-cell lung cancer and highlighted the potential value of such algorithms for therapeutic decision-making. The large body of research into machine learning and quantitative imaging biomarker applications in lung cancer has also provided insight into key challenges associated with such applications. Yang et al. note that the widespread application of texture-derived image features as prognostic predictors is impeded by a lack of quality control

and robustness and proceed to demonstrate high inter-rater variability impacting the reproducibility of texture parameters [34]. Such challenges are of course not immanent to thoracic imaging workflows and have been repeatedly noted in previous studies irrespective of imaging modality applied [35, 36] with PET-specific solutions recently proposed [37].

12.3.4 Prostate Cancer

The role of hybrid imaging in prostate cancer is continuously evolving and expanding with the application of Gallium or Fluorine-labeled PSMA supported by recent meta-analyses [38, 39] and having been demonstrated to impact patient management in a majority of cases [40]. The first randomized prospective trial testing the influence of PSMA PET/CT on prostate patient outcome was announced in early 2019 [41]. Quantitative imaging feature studies have recently provided promising results applied to PSMA PET. For example, Zamboglou et al. demonstrate PSMA-PET-derived quantitative features to discriminate between cancer- and non-cancer-affected prostate tissue, as well as differentiate between Gleason scores of 7 and ≥ 8 and between patients with and without nodal involvement [42]. PSMA expression is an excellent example of a theranostic application, i.e. the specific expression monitoring of a therapy-relevant target: since Lutetium-PSMA can be used for radioligand treatment in advanced prostate cancer [43], machine-learning applications predicting response to such therapy directly from the images could hence represent a promising next step.

12.3.5 Breast Cancer

The field of breast cancer research has witnessed among the strongest advances in the utilization of quantitative imaging workflows and the application of machine intelligence, likely due to the high quality of image acquisition because of the lack of motion artifacts, the universal

implementation of standardized reporting in the form of BIRADS and the high incidence. Hence, several studies have proposed image-derived features for the noninvasive characterization of breast cancer. For example, Antunovic et al. utilized pretreatment FDG-PET/CT of breast cancer and found histogram features to be associated with histopathological, molecular, and receptor expression subtypes [44]. Similarly, Huang et al. found image features derived from PET/MRI data to be associated with tumor grading, stage, subtype, recurrence, and survival [45]. Ou et al. utilize machine learning to differentiate between breast carcinoma and breast lymphoma based on texture features derived from FDG-PET/CT [46]. Focused on therapy response prediction, Antunovic and colleagues noted the association of molecular breast cancer subtypes with distinct responses to neoadjuvant chemotherapy and developed machine learning algorithms on FDG-PET/CT to predict pathological complete response in locally advanced breast cancer [47]. Ha et al. also utilized FDG-PET/CT to develop machine learning-derived metabolic signatures of breast cancer associated with Ki67 gene expression, pathological complete response to neoadjuvant chemotherapy and recurrence risk [48]. As noted above, however, such workflows are not without challenges and it was recently noted in the work by Sollini et al. that *most evidence on the utility […] is at the feasibility level.* The authors recommend harmonization, validation on representative datasets and the establishment of guidelines for the application of quantitative imaging parameters in breast imaging [49].

12.3.6 Gastrointestinal Oncology

The largest body of work regarding therapy prediction using quantitative image-derived parameters in hybrid imaging has arguably been produced in the area of gastrointestinal oncology. In esophageal cancer for instance, several studies on radiomics workflows have highlighted the significance of heterogeneity-related image features and have derived models predictive of prognosis

and therapy response [50–52]. Yip et al. included longitudinally acquired datasets in their model and found a decrease in tumor heterogeneity-related texture and histogram features to be associated with tumor response and patient survival [53]. Ypsilantis et al. employed convolutional neural networks on PET scans and found them to outperform radiomics models in the prediction of therapy response in esophageal cancer [54]. Furthermore, sub-regional analyses, taking into account intra-tumoral heterogeneity are being assessed for their impact on the survival of esophageal cancer patients treated with chemoradiation, shown, for example, in the study by Xie et al. [55].

In pancreatic cancer, multiparametric imaging and machine learning have been investigated for differentiation of inflammatory and neoplastic processes [56]. The added utility of hybrid fusion imaging for the delineation of tumors has been noted by Belli et al. in a recent study [57] with applications in quantitative imaging workflows. In our own work, we note the importance and potential benefits of multiparametric data integration for accurate prognostic prediction in the field of pancreatic cancer [58]. Cui et al. identified quantitative parameters prognostic of stereotactic radiation therapy in pancreatic cancer from FDG-PET/CT imaging [59]. With the evolving role of hybrid imaging for therapy planning in pancreatic cancer [60, 61] especially with respect to neoadjuvant treatment regimens, as well as the advances in molecular subtyping including the distinction of differentially activated metabolic pathways, [62–64], it must be assumed that the scope of quantitative imaging workflows will soon expand further to hybrid imaging.

In rectal cancer, several studies have investigated the utility of pretreatment quantitative imaging biomarkers in the prediction of therapy response. The study by Lovinfosse and colleagues found texture parameters derived from pretreatment FDG-PET/CT predictive of survival in a cohort of patients with locally advanced rectal cancer treated with neoadjuvant chemoradiation, noting that these features outperformed volume-based parameters in predictive performance [65]. Amorim et al. compared FDG-PET- and diffu-

sion-weighted MRI-derived parameters and observed the information gained from these modalities to be independent and complementary, underscoring the relevance of multiparametric hybrid imaging workflows in oncology [66]. The importance of tumor heterogeneity was noted by Bundschuh et al., who note that heterogeneity-related image features are relevant both early in the course of therapy and after its completion [67]. A similar dual timepoint study was performed by Jeon et al., who performed multiparametric modeling including clinical parameters and MRI-derived texture features and observed changes in these features to be associated with distinct risk phenotypes. The authors note that their results would be applicable to and benefit from the inclusion of functional imaging [68].

12.4 Discussion and Outlook

In this chapter, we review the applications of machine learning and artificial intelligence to therapy monitoring in the domain of molecular and hybrid imaging, as well as theranostics. Despite its somewhat earlier stage of evolution compared to applications purely focused on diagnosis, such as tumor detection or subtype classification, the multitude of studies presented showcase the intense research interest in the field and provide an outlook on the main objectives of techniques, algorithms, and applications aimed at therapy monitoring and response prediction. Evidently, diagnostic and theranostic applications are closely related. For example, specific tumor subtypes are associated with distinct therapy response, providing space for exploration of novel therapy targets and specific therapeutic agents. The clinical utilization of theranostic radiotracers is also expected to expand beyond the current main routine application of prostate imaging with PSMA: initial studies report successes, for example, in the application of texture analysis in neuroendocrine tumors [69]. The combined application of diagnostic and theranostic radiotracers has also been reported, with very recent results showcasing their complementary value in the outcome prediction of pancreatic neuroendocrine neoplasms [70], expanding on previous studies reporting on combined radiotracer application [71]. We believe machine learning techniques to herald a transition towards integrated theranostic applications which will likely blur the current borders between diagnosis- and therapy-response-focused studies. This evolution will obviously not remain without challenges. Foremost, it will be predicated on the development and availability of emerging and novel theranostic radiotracers beyond the above-mentioned fields of prostate cancer and neuroendocrine tumors, as well as the understanding of their interaction with biological targets and their unique challenges and pitfalls [72], to enable their utilization in AI-guided and precision medicine applications [73].

Reviewing the current literature findings, a clear trend can be observed from tumor tissue and metabolic volume tracking applications towards image texture analysis which can be ascribed to the above-mentioned rise of quantitative imaging workflows [74] within the past few years. We however still observe specific challenges, several of which are unmet in current literature:

Nearly all of the studies outlined above utilize hybrid imaging-based texture analysis workflows. A more thorough investigation on the differential contribution of each modality to the predictive model, or an analysis of the added benefit of hybrid imaging over a single modality were not routinely performed. Anatomic and functional imaging have been shown to present specific and individual challenges with respect to texture analysis, rendering such a differentiated assessment necessary [75]. Furthermore, the difficulties of harmonizing quantitative imaging workflows and rendering them robust towards variances between diagnostic equipment vendors, differences in human performance and unstandardized texture feature specifications have been noted extensively in the literature [36], mostly aimed at anatomical imaging modalities. However, recent works have focused on harmonizing texture features specifically in functional imaging [37] alongside efforts for protocol standardization and guidelines aimed at hybrid imaging studies [76]. Ultimately, we

believe handcrafted quantitative imaging features and the field of radiomics to represent an intermediate step in the evolution of machine learning application in medical imaging towards deep learning-based workflows. The latter offer greater representational flexibility and robustness, obviating post-processing and harmonization requirements in favor of data diversity and larger patient cohorts and rendering them inherently more suitable for multicentric studies [77–80]. The advent of deep learning and associated advances in image registration [81] will also signify greater facility in integrating additional information from studies acquired at multiple timepoints. Longitudinal imaging has been shown to offer deeper insight into therapy-related changes in tumor biology [82]; however, it was only performed in a small fraction of the studies presented above due to the difficulties of acquiring multi-timepoint imaging and the escalated requirements towards selection of time-stable and reproducible image features [83]. Lastly, many of the studies presented base their assessment of therapy response on surrogate measurements, for example, on tumor volume decrease or on associations between therapy response and a decrease in image heterogeneity believed to mirror biological phenomena, which cannot always be objectively validated. Furthermore, therapy response is a multifactorial process greatly dependent on clinical parameters, which should be included in the modeling process [58]. The introduction of algorithms enabling the direct prediction of patient survival from images and the associated clinical data [84] will thus improve the capabilities for pre-therapeutic risk stratification and provide higher confidence for guiding therapy decisions.

In conclusion, this chapter discusses the applications of machine learning-based medical image analysis workflows, their applications to therapy response monitoring and theranostics in a hybrid imaging setting, as well as current and future research directions. We believe that the concurrent evolution and innovations in the fields of oncologic hybrid imaging, theranostics, and computer vision will fuel scientific discovery in the field and provide the opportunity for clinical translation and improvements to patient care.

References

1. Sejnowski TJ. The unreasonable effectiveness of deep learning in artificial intelligence. Proc Natl Acad Sci. 2020;117(48):30033–8.
2. Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L. ImageNet: a large-scale hierarchical image database. In: CVPR09. 2009.
3. Gibaud B. The quest for standards in medical imaging. Eur J Radiol. 2011;78:190–8.
4. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med. 2015;12:e1001779.
5. Prior F, Smith K, Sharma A, Kirby J, Tarbox L, Clark K, et al. The public cancer radiology imaging collections of the cancer imaging archive. Scient Data. 2017;4:170124.
6. Kaissis GA, Makowski MR, Rückert D, Braren RF. Secure, privacy-preserving and federated machine learning in medical imaging. Nature Mach Intell. 2020;2:305–11.
7. Clark M, Hall L, Goldgof D, Velthuizen R, Murtagh F, Silbiger M. Automatic tumor segmentation using knowledge-based techniques. IEEE Trans Med Imaging. 1998;17:187–201.
8. Vokurka EA, Herwadkar A, Thacker NA, Ramsden RT, Jackson A. Using bayesian tissue classification to improve the accuracy of vestibular schwannoma volume and growth measurement. AJNR Am J Neuroradiol. 2002;23:459–67.
9. Haney SM, Thompson PM, Cloughesy TF, Alger JR, Toga AW. Tracking tumor growth rates in patients with malignant gliomas: a test of two algorithms. AJNR Am J Neuroradiol. 2001;22:73–82.
10. Zhao Z, Yang G, Lin Y, Pang H, Wang M. Automated glioma detection and segmentation using graphical models. PLoS One. 2018;13:e0200745.
11. Lee H, Paeng JC, Hong SH, Yoo HJ, Cheon GJ, Lee DS, et al. Appropriate margin thresholds for isocontour metabolic volumetry of fluorine-18 fluorodeoxyglucose PET in sarcoma. Nucl Med Commun. 2016;37:1088–94.
12. Bang J-I, Ha S, Kang S-B, Lee K-W, Lee H-S, Kim J-S, et al. Prediction of neoadjuvant radiation chemotherapy response and survival using pretreatment [18F] FDG PET/CT scans in locally advanced rectal cancer. Eur J Nucl Med Mol Imaging. 2015;43:422–31.
13. Kostakoglu L, Chauvie S. Metabolic tumor volume metrics in lymphoma. Semin Nucl Med. 2018;48:50–66.
14. Erdogan M, Erdemoglu E, Evrimler S, Hanedan C, Şengül SS. Prognostic value of metabolic tumor volume and total lesion glycolysis assessed by 18F-FDG PET/CT in endometrial cancer. Nucl Med Commun. 2019;40:1099–104.
15. Yildirim BA, Torun N, Guler OC, Onal C. Prognostic value of metabolic tumor volume and total lesion gly-

colysis in esophageal carcinoma patients treated with definitive chemoradiotherapy. Nucl Med Commun. 2018;39:553–63.

16. Gallamini A, Kostakoglu L. Metabolic tumor volume: we still need a platinum-standard metric. J Nucl Med. 2016;58:196–7.

17. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. arXiv e-prints [Internet]. 2014. Available from: http://arxiv.org/abs/1411.4038v2

18. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. arXiv e-prints [Internet]. 2015. Available from: http://arxiv.org/abs/1505.04597v1

19. Wu X, Sahoo D, Zhang D, Zhu J, Hoi SC. Single-shot bidirectional pyramid networks for high-quality object detection. Neurocomputing. 2020;401:1–9.

20. Beichel RR, Smith BJ, Bauer C, Ulrich EJ, Ahmadvand P, Budzevich MM, et al. Multi-site quality and variability analysis of 3D FDG PET segmentations based on phantom and clinical image data. Med Phys. 2017;44:479–96.

21. Andreassen MMS, Goa PE, Sjøbakk TE, Hedayati R, Eikesdal HP, Deng C, et al. Semi-automatic segmentation from intrinsically-registered 18F-FDGPET/MRI for treatment response assessment in a breast cancer cohort: comparison to manual DCEMRI. MAGMA. 2019;33:317–28.

22. Rundo L, Stefano A, Militello C, Russo G, Sabini MG, D'Arrigo C, et al. A fully automatic approach for multimodal PET and MR image segmentation in gamma knife treatment planning. Comput Methods Prog Biomed. 2017;144:77–96.

23. Lohmann P, Lerche C, Bauer EK, Steger J, Stoffels G, Blau T, et al. Predicting IDH genotype in gliomas using FET PET radiomics. Sci Rep. 2018;8:13328.

24. Hotta M, Minamimoto R, Miwa K. 11C-methionine-PET for differentiating recurrent brain tumor from radiation necrosis: radiomics approach with random forest classifier. Sci Rep. 2019;9:15666.

25. Lohmann P, Kocher M, Ruge MI, Visser-Vandewalle V, Shah NJ, Fink GR, et al. PET/MRI radiomics in patients with brain metastases. Front Neurol. 2020;11:1.

26. Cha YJ, Jang WI, Kim M-S, Yoo HJ, Paik EK, Jeong HK, et al. Prediction of response to stereotactic radiosurgery for brain metastases using convolutional neural networks. Anticancer Res. 2018;38:5437–45.

27. Fujima N, Hirata K, Shiga T, Li R, Yasuda K, Onimaru R, et al. Integrating quantitative morphological and intratumoural textural characteristics in FDG-PET for the prediction of prognosis in pharynx squamous cell carcinoma patients. Clin Radiol. 2018;73(1059):e1–8.

28. Feliciani G, Fioroni F, Grassi E, Bertolini M, Rosca A, Timon G, et al. Radiomic profiling of head and neck cancer: 18F-FDG PET texture analysis as predictor of patient survival. Contrast Media Mol Imaging. 2018;2018:1–8.

29. Crispin-Ortuzar M, Apte A, Grkovski M, Oh JH, Lee NY, Schöder H, et al. Predicting hypoxia status

using a combination of contrast-enhanced computed tomography and [18F]-fluorodeoxyglucose positron emission tomography radiomics features. Radiother Oncol. 2018;127:36–42.

30. van Tinteren H, Hoekstra OS, Smit EF, van den Bergh JH, Schreurs AJ, Stallaert RA, et al. Effectiveness of positron emission tomography in the preoperative assessment of patients with suspected non-small-cell lung cancer: the PLUS multicentre randomised trial. Lancet. 2002;359:1388–92.

31. Hyun SH, Ahn MS, Koh YW, Lee SJ. A machine-learning approach using PET-based radiomics to predict the histological subtypes of lung cancer. Clin Nucl Med. 2019;44:956–60.

32. Ahn H, Lee H, Kim S, Hyun S. Pre-treatment 18F-FDG PET-based radiomics predict survival in resected non-small cell lung cancer. Clin Radiol. 2019;74:467–73.

33. Oikonomou A, Khalvati F, Tyrrell PN, Haider MA, Tarique U, Jimenez-Juan L, et al. Radiomics analysis at PET/CT contributes to prognosis of recurrence and survival in lung cancer treated with stereotactic body radiotherapy. Sci Rep. 2018;8:4003.

34. Yang F, Simpson G, Young L, Ford J, Dogan N, Wang L. Impact of contouring variability on oncological PET radiomics features in the lung. Sci Rep. 2020;10:369.

35. van Timmeren JE, Carvalho S, Leijenaar RTH, Troost EGC, van Elmpt W, de Ruysscher D, et al. Challenges and caveats of a multi-center retrospective radiomics study: an example of early treatment response assessment for NSCLC patients using FDG-PET/CT radiomics. PLoS One. 2019;14:e0217536.

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

37. Orlhac F, Boughdad S, Philippe C, Stalla-Bourdillon H, Nioche C, Champion L, et al. A postreconstruction harmonization method for multicenter radiomic studies in PET. J Nucl Med. 2018;59:1321–8.

38. Hope TA, Goodman JZ, Allen IE, Calais J, Fendler WP, Carroll PR. Meta analysis of 68Ga-PSMA-11 PET accuracy for the detection of prostate cancer validated by histopathology. J Nucl Med. 2018;60:786–93.

39. Perera M, Papa N, Christidis D, Wetherell D, Hofman MS, Murphy DG, et al. Sensitivity, specificity, and predictors of positive 68 gaProstate-specific membrane antigen positron emission tomography in advanced prostate cancer: a systematic review and meta-analysis. Eur Urol. 2016;70:926–37.

40. Calais J, Fendler WP, Eiber M, Gartmann J, Chu F-I, Nickols NG, et al. Impact of 68Ga-PSMA-11 PET/CT on the management of prostate cancer patients with biochemical recurrence. J Nucl Med. 2017;59:434–41.

41. Calais J, Czernin J, Fendler WP, Elashoff D, Nickols NG. Randomized prospective phase III trial of 68Ga-PSMA-11 PET/CT molecular imaging for prostate cancer salvage radiotherapy planning [PSMA-SRT]. BMC Cancer. 2019;19:18.

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

43. Hofman MS, Violet J, Hicks RJ, Ferdinandus J, Thang SP, Akhurst T, et al. [177 lu]-PSMA-617 radionuclide treatment in patients with metastatic castration-resistant prostate cancer (LuPSMA trial): A single-centre, single-arm, phase 2 study. Lancet Oncol. 2018;19:825–33.

44. Antunovic L, Gallivanone F, Sollini M, Sagona A, Invento A, Manfrinato G, et al. [18F]FDG PET/CT features for the molecular characterization of primary breast tumors. Eur J Nucl Med Mol Imaging. 2017;44:1945–54.

45. Huang S-y, Franc BL, Harnish RJ, Liu G, Mitra D, Copeland TP, et al. Exploration of PET and MRI radiomic features for decoding breast cancer phenotypes and prognosis. npj Breast Cancer. 2018;4:24.

46. Ou X, Zhang J, Wang J, Pang F, Wang Y, Wei X, et al. Radiomics based on 18 f-FDG PET/CT could differentiate breast carcinoma from breast lymphoma using machine-learning approach: a preliminary study. Cancer Med. 2019;9:496–506.

47. Antunovic L, Sanctis RD, Cozzi L, Kirienko M, Sagona A, Torrisi R, et al. PET/CT radiomics in breast cancer: promising tool for prediction of pathological response to neoadjuvant chemotherapy. Eur J Nucl Med Mol Imaging. 2019;46:1468–77.

48. Ha S, Park S, Bang J-I, Kim E-K, Lee H-Y. Metabolic radiomics for pretreatment 18F-FDG PET/CT to characterize locally advanced breast cancer: histopathologic characteristics, response to neoadjuvant chemotherapy, and prognosis. Sci Rep. 2017;7:1556.

49. Sollini M, Cozzi L, Ninatti G, Antunovic L, Cavinato L, Chiti A, et al. PET/CT radiomics in breast cancer: mind the step. Methods. 2021;188:122–32.

50. Yip C, Landau D, Kozarski R, Ganeshan B, Thomas R, Michaelidou A, et al. Primary esophageal cancer: heterogeneity as potential prognostic biomarker in patients treated with definitive chemotherapy and radiation therapy. Radiology. 2014;270:141–8.

51. van Rossum PSN, Fried DV, Zhang L, Hofstetter WL, van Vulpen M, Meijer GJ, et al. The incremental value of subjective and quantitative assessment of 18F-FDG PET for the prediction of pathologic complete response to preoperative chemoradiotherapy in esophageal cancer. J Nucl Med. 2016;57:691–700.

52. Xiong J, Yu W, Ma J, Ren Y, Fu X, Zhao J. The role of PET-based radiomic features in predicting local control of esophageal cancer treated with concurrent chemoradiotherapy. Sci Rep. 2018;8:9902.

53. Yip C, Davnall F, Kozarski R, Landau DB, Cook GJR, Ross P, et al. Assessment of changes in tumor heterogeneity following neoadjuvant chemotherapy in primary esophageal cancer. Dis Esophagus. 2014;28:172–9.

54. Ypsilantis P-P, Siddique M, Sohn H-M, Davies A, Cook G, Goh V, et al. Predicting response to neoadjuvant chemotherapy with PET imaging using convolutional neural networks. PLoS One. 2015;10:e0137036.

55. Xie C, Yang P, Zhang X, Xu L, Wang X, Li X, et al. Sub-region based radiomics analysis for survival prediction in oesophageal tumours treated by definitive concurrent chemoradiotherapy. EBioMedicine. 2019;44:289–97.

56. Zhang Y, Cheng C, Liu Z, Wang L, Pan G, Sun G, et al. Radiomics analysis for the differentiation of autoimmune pancreatitis and pancreatic ductal adenocarcinoma in 18 f-FDG PET/CT. Med Phys. 2019;46:4520–30.

57. Belli ML, Mori M, Broggi S, Cattaneo GM, Bettinardi V, Dell'Oca I, et al. Quantifying the robustness of [18 f]FDG-PET/CT radiomic features with respect to tumor delineation in head and neck and pancreatic cancer patients. Phys Med. 2018;49:105–11.

58. Kaissis GA, Jungmann F, Ziegelmayer S, Lohöfer FK, Harder FN, Schlitter AM, et al. Multiparametric modelling of survival in pancreatic ductal adenocarcinoma using clinical, histomorphological, genetic and image-derived parameters. J Clin Med. 2020;9:1250.

59. Cui Y, Song J, Pollom E, Shirato H, Chang D, Koong A, et al. Radiomic analysis of FDG-PET identifies novel prognostic imaging biomarkers in locally advanced pancreatic cancer patients treated with SBRT. Int J Radiat Oncol Biol Phys. 2015;93:S4–5.

60. Dholakia AS, Chaudhry M, Leal JP, Chang DT, Raman SP, Hacker-Prietz A, et al. Baseline metabolic tumor volume and total lesion glycolysis are associated with survival outcomes in patients with locally advanced pancreatic cancer receiving stereotactic body radiation therapy. Int J Radiat Oncol Biol Phys. 2014;89:539–46.

61. Mellon EA, Jin WH, Frakes JM, Centeno BA, Strom TJ, Springett GM, et al. Predictors and survival for pathologic tumor response grade in borderline resectable and locally advanced pancreatic cancer treated with induction chemotherapy and neoadjuvant stereotactic body radiotherapy. Acta Oncol. 2016;56:391–7.

62. Kaissis G, Ziegelmayer S, Lohöfer F, Steiger K, Algül H, Muckenhuber A, et al. A machine learning algorithm predicts molecular subtypes in pancreatic ductal adenocarcinoma with differential response to gemcitabine-based versus FOLFIRINOX chemotherapy. PLoS One. 2019;14:e0218642.

63. Kaissis GA, Ziegelmayer S, Lohöfer FK, Harder FN, Jungmann F, Sasse D, et al. Image-based molecular phenotyping of pancreatic ductal adenocarcinoma. J Clin Med. 2020;9:724.

64. Kaissis G, Ziegelmayer S, Lohöfer F, Algül H, Eiber M, Weichert W, et al. A machine learning model for the prediction of survival and tumor subtype in pancreatic ductal adenocarcinoma from preoperative diffusion-weighted imaging. Eur Radiol Exp. 2019;3:41.

65. Lovinfosse P, Polus M, Daele DV, Martinive P, Daenen F, Hatt M, et al. FDG PET/CT radiomics for predicting the outcome of locally advanced rectal

cancer. European Journal of Nuclear Medicine and Molecular Imaging. 2017;45:365–75.

66. Amorim BJ, Torrado-Carvajal A, Esfahani SA, Marcos SS, Vangel M, Stein D, et al. PET/MRI radiomics in rectal cancer: a pilot study on the correlation between PET- and MRI-derived image features with a clinical interpretation. Mol Imag Biol. 2020;22(5):1438–45.

67. Bundschuh RA, Dinges J, Neumann L, Seyfried M, Zsoter N, Papp L, et al. Textural parameters of tumor heterogeneity in 18F-FDG PET/CT for therapy response assessment and prognosis in patients with locally advanced rectal cancer. J Nucl Med. 2014;55:891–7.

68. Jeon SH, Song C, Chie EK, Kim B, Kim YH, Chang W, et al. Delta-radiomics signature predicts treatment outcomes after preoperative chemoradiotherapy and surgery in rectal cancer. Radiat Oncol. 2019;14:43.

69. Werner RA, Ilhan H, Lehner S, Papp L, Zsótér N, Schatka I, et al. Pre-therapy somatostatin receptor-based heterogeneity predicts overall survival in pancreatic neuroendocrine tumor patients undergoing peptide receptor radionuclide therapy. Mol Imaging Biol. 2018;21:582–90.

70. Mapelli P, Partelli S, Salgarello M, Doraku J, Pasetto S, Rancoita PM, et al. Dual tracer 68Ga-DOTATOC and 18F-FDG PET/computed tomography radiomics in pancreatic neuroendocrine neoplasms: an endearing tool for preoperative risk assessment. Nucl Med Commun. 2020;41(9):896–905.

71. Hindié E. The NETPET score: combining FDG and somatostatin receptor imaging for optimal management of patients with metastatic well-differentiated neuroendocrine tumors. Theranostics. 2017;7:1159–63.

72. Werner RA, Thackeray JT, Pomper MG, Bengel FM, Gorin MA, Derlin T, et al. Recent updates on molecular imaging reporting and data systems (MI-RADS) for theranostic radiotracersNavigating pitfalls of SSTR- and PSMA-targeted PET/CT. J Clin Med. 2019;8:1060.

73. van der Meel R, Sulheim E, Shi Y, Kiessling F, Mulder WJM, Lammers T. Smart cancer nanomedicine. Nat Nanotechnol. 2019;14:1007–17.

74. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. Radiology. 2016;278:563–77.

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

76. Prenosil GA, Weitzel T, Fürstner M, Hentschel M, Krause T, Cumming P, et al. Towards guidelines to harmonize textural features in PET: Haralick textural features vary with image noise, but exposure-invariant domains enable comparable PET radiomics. PLoS One. 2020;15:e0229560.

77. Kuhl CK, Truhn D. The long route to standardized radiomics: unraveling the knot from the end. Radiology. 2020;295:339–41.

78. Truhn D, Schrading S, Haarburger C, Schneider H, Merhof D, Kuhl C. Radiomic versus convolutional neural networks analysis for classification of contrast-enhancing lesions at multiparametric breast MRI. Radiology. 2019;290:290–7.

79. Lou B, Doken S, Zhuang T, Wingerter D, Gidwani M, Mistry N, et al. An image-based deep learning framework for individualising radiotherapy dose: a retrospective analysis of outcome prediction. Lancet Digit Health. 2019;1:e136–47.

80. Sun Q, Lin X, Zhao Y, Li L, Yan K, Liang D, et al. Deep learning vs. radiomics for predicting axillary lymph node metastasis of breast cancer using ultrasound images: Don't forget the peritumoral region. Front Oncol. 2020;10:53.

81. Haskins G, Kruger U, Yan P. Deep learning in medical image registration: a survey. Mach Vis Appl. 2020;31:8.

82. Bak SH, Park H, Sohn I, Lee SH, Ahn M-J, Lee HY. Prognostic impact of longitudinal monitoring of radiomic features in patients with advanced non-small cell lung cancer. Sci Rep. 2019;9:8730.

83. Leijenaar RTH, Carvalho S, Velazquez ER, van Elmpt WJC, Parmar C, Hoekstra OS, et al. Stability of FDG-PET radiomics features: an integrated analysis of test-retest and inter-observer variability. Acta Oncol. 2013;52:1391–7.

84. Katzmann A, Mühlberg A, Sühling M, Nörenberg D, Maurus S, Holch JW, et al. Computed tomography image-based deep survival regression for metastatic colorectal cancer using a non-proportional hazards model. Predict Intell Med. 2019:73–80.