**QUEEN'S
UNIVERSITY
BELFAST**

Machine learning for radiomics-based multi-modality and multi-parametric modeling

Wei, L., Osman, S., Hatt, M., & El Naqa, I. (2019). Machine learning for radiomics-based multi-modality and multi-parametric modeling. *The Quarterly Journal of Nuclear Medicine and Molecular Imaging* .
<https://doi.org/10.23736/S1824-4785.19.03213-8>

Published in:

The Quarterly Journal of Nuclear Medicine and Molecular Imaging

Document Version:

Peer reviewed version

Queen's University Belfast - Research Portal:

[Link to publication record in Queen's University Belfast Research Portal](#)

Publisher rights

Copyright 2019 The Quarterly Journal of Nuclear Medicine and Molecular Imaging

General rights

Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.

Machine learning for radiomics-based multi-modality and multi-parametric modeling

Lise Wei¹, Sarah Osman², Mathieu Hatt³, and Issam El Naqa¹

¹Department of Radiation Oncology, University of Michigan, Ann Arbor, MI, USA

²Centre for Cancer Research and Cell Biology, Queens' University, Belfast, UK

³LaTIM, INSERM, UMR 1101, Univ Brest, Brest, France

Abstract

Due to the recent developments of both hardware and software technologies, multimodality medical imaging techniques have been increasingly applied in clinical practice and research studies. Previously, the application of multimodality imaging in oncology has been mainly related to combining anatomical and functional imaging to improve diagnostic specificity and/or target definition, such as positron emission tomography/computed tomography (PET/CT) and single-photon emission computed tomography (SPECT)/CT. More recently, the fusion of various images, such as multi-parametric magnetic resonance imaging (MRI) sequences, different PET tracer images, PET/MRI, has become more prevalent, which has enabled more comprehensive characterization of the tumor phenotype. In order to take advantage of these valuable multimodal data for clinical decision making using radiomics, we present two ways to implement the multimodal image analysis, namely radiomic (handcrafted feature) based and deep learning (machine learned feature) based methods. Applying advanced machine (deep) learning algorithms across multi-modality images have shown better results compared with single modality modeling for prognostic and/or prediction of clinical outcomes. This holds great potentials for providing more personalized treatment for patients and achieve better outcomes.

Keywords: multimodality imaging, radiomics, deep learning, feature engineering, outcome modeling.

1. Introduction

The exponential growth in the use of more than one imaging modality for diagnostic, therapeutic and prognostic purposes in noninvasive and quantitative cancer studies has facilitated the development of multimodality techniques. In this review, we will focus on two approaches: based on handcrafted features (a.k.a. radiomics) and based on automatically or machine learned features from data (a.k.a. deep learning). The intuition of multimodality imaging is that different types of images can provide complementary information (for example anatomical and functional in the case of PET/CT) and combining them may result in more complete characterization of the disease (e.g., a tumor). Thus, initial efforts were made on combining functional and anatomical imaging types, such as SPECT/CT and PET/CT, which are today commonly applied in clinical practice. PET/CT is being used for diagnosis, staging and treatment planning for head and neck cancer [1, 2], lung cancer [3-6] and breast cancer [7]. PET/MRI can be superior to PET/CT in specific applications, due to better soft tissue contrast of MRI, such as for liver cancer [8, 9]. Another study of ^{68}Ga -PSMA-11 PET/MRI showed improved diagnostic accuracy over the individual modalities (multi-parametric MRI and PET) for localization of primary prostate cancer [10]. Now with more types of functional images available, which, for example, could provide, metabolic, vascular, and cellularity information, and better spatial and temporal resolution for CT and MRI, integration of various types of functional images (e.g. multi-parametric MRI, various tracers PET) has become possible for cancer research. For instance, PET, with ^{18}F -FDG (fluoro-2-deoxy-d-glucose), a glucose metabolism analog, could reflect the metabolic activity (assessing the aggressiveness of disease), with ^{18}F -fluoromisonidazole (FMISO), could quantify tissue oxygen status. Diffusion-weighted MRI (DW-MRI) maps the regional variations in apparent diffusion coefficient (ADC), which is inversely correlated with tissue cellularity. In Figure 1, we show an example of the usage of multimodality imaging in clinical routine, with a case where FDG-PET/CT images were used for the diagnosis of a metastasis of non-small-cell lung cancer (NSCLC).

Fig. 1 Multimodality imaging in the diagnosis of a metastasis in a patient with NSCLC. a. A hyper-metabolic focus in the right hemipelvis was found in the axial attenuation-corrected FDG-PET image, which is suggestive of a metastatic lesion, but difficult to localize anatomically. b. A corresponding unenhanced axial CT image. c. The lesion was conclusively defined as a metastasis of the right ischial tuberosity in the merged PET/CT image [11].

Radiomics, which is a field that aims to provide a comprehensive quantification of the imaging phenotype using automated data characterization algorithms [12], is a powerful tool for taking advantages of multimodality images. Conventional radiomics-based approach involves extracting a large amount of handcrafted features (e.g., intensity, shape and texture parameters), capturing different characteristics of the regions of interest, some of which may be difficult or even impossible to discern by human vision, even the expertly trained one [13]. With the objective of outcome modeling (classification, e.g. toxicity, response to treatment, overall or disease-free survival), radiomic features have been related to clinical and biological endpoints by feature selection and subsequent machine learning model construction. Recently, the rise of computational power due to advanced hardware (e.g., GPU, cluster or cloud computing) combined with better algorithms for training neural networks, have facilitated the breakthrough of successful application of deep learning algorithms in many computer vision and analysis tasks, for both natural and medical images [14-17]. However, machine learning, either radiomic or deep learning based methods in multimodality imaging data is still in developmental stage in terms of investigation from a systematic perspective about how we can fuse and explore the most out of the multi-representations.

Quantitative analyses of these multimodality images possess the potential to reveal underlying biological mechanisms and to enable a more precise and personalized patient health care. However, the prerequisite is to not only obtain these images, but also apply proper algorithms to extract the clinically-relevant information. Thus, in this review, we give a brief introduction to radiomic and deep learning based methods by first explaining the radiomic feature extraction, conventional machine learning algorithms (feature selection and model construction), followed by an introduction to deep learning algorithms. Then, some applications in each category are discussed to shed lights on possible ways to deal with multimodality data.

2. Radiomics

2.1 Handcrafted features and conventional machine learning

Before feature extraction, preprocessing of imaging data is often necessary for enhancing image quality (e.g., signal-to-noise ratio improvement), determining regions of interest, enabling the repeatable and comparable radiomic analysis. This is important especially in cases when different parameter settings or scanners are used across multi-centers, different protocols, or multimodalities. For different imaging modalities, the images should be co-registered relying on rigid and/or deformable registration or relying on inherent hardware registration. Moreover, some imaging modalities may need to undergo a conversion into a standardized (and more informative) scale, such as standardized uptake values (SUV) in PET in order to take into account the injected amount of radiotracer and the patient's weight [18]. MRI images preprocessing requires a bit more efforts due to two main reasons: (1) various sequences available for MRI (T1, T2, DWI, etc.) with changes in magnetic field strength, image definition, and weighting of tissue intensities, lead to the lack of a standard and quantifiable interpretation of image intensities; (2) some acquisition-related artifacts, such as the bias field. MRI scale standardization and N3/N4 bias correction processing are

typical preprocessing implemented steps for MRI images to account for the mentioned issues [19-21]. In addition to these modalities, specific preprocessing, some general procedures before radiomic feature extraction should also be applied. Isotropic voxel size is recommended for some textural features, which can be obtained by either down-sampling or up-sampling with some interpolation algorithms, such as polynomial and spline interpolation [22]. Prior to the calculation of textural features, the grey levels should be discretized to reduce noise and make the computation tractable [23]. There are several ways to perform discretization, two of which are described by the IBSI: fixed bin number N or fixed bin width B . Each approach has advantages and drawbacks, as described by the IBSI [24]. The IBSI recommends to avoid relying on the fixed bin width approach when the modality has no quantified values with physical meaning (such as some MRI sequences). On the other hand, the fixed bin number method has been recommended when the modality is not well calibrated [25, 26]. Though there are some preprocessing methods mentioned above that could standardize the radiomic features, the robustness of these features is still affected by multiple factors, such as the scanner type, the mathematics of extracting the features, the imaging settings. Thus, the test-retest analysis of the extracted features and phantom studies are very helpful for obtaining reliable radiomics models in practice.

ROIs for tumors are usually heterogeneous spatially and temporally. Thus, it is natural to divide radiomic features into static (time invariant) and dynamic (time variant) features.

2.1.1 Static features

Static features are based on intensity, shape, size (volume), texture and wavelet, describing the geometric property and the distribution of intensities of the ROIs in relation to its spatial distribution.

a) Morphological features: These are geometrical shape characteristics of ROIs, such as compactness (representing how compact the region is), eccentricity (a measure of non-circularity, describing tumor growth directionality); Euler number (the number of connected objects in a region minus the solidity (this is a measurement of convexity), which may be a characteristics of benign lesions [27, 28].

b) first order intensity features

First-order features are based on first-order histograms that shows the distribution of the voxel intensities in the ROIs. These features summarize the large number of voxel values in ROIs into single values, such as mean, minimum, skewness, etc.

c) Texture features: Broadly speaking, there are three categories for texture analysis: statistical (e.g. second and higher order features), model-based (e.g. Gaussian Markov random fields, Gabor filter and wavelet) and structural methods (e.g. Topological texture descriptors, Invariant histogram). The performance of texture approaches is not affected by tumor position, orientation, size, and brightness. It takes into account the local intensity-spatial distribution, and is also invariant to translation, rotation, affine and perspective transform [29-32]. Among these features, statistical methods have been widely used in the field of radiomics for cancer outcome modeling. Second-order features provide statistical interrelationships between voxels and capture special patterns in the ROIs, which make up for the loss of information associated with the first-order features. Haralick et al. introduced the idea of using textural features for image classification [33]. Several texture matrices formed the basis of statistical approaches: the grey level co-occurrence matrix (GLCM) [33], neighborhood gray tone difference matrix (NGTDM) [34], run-length matrix (RLM) [35], and grey level size-zone matrix (GLSZM) [36]. GLCM illustrates the distribution of the combinations of grey levels of neighboring voxels (pixels) along certain direction, entropy, angular second moment, correlation, contrast, inverse difference moment are

commonly used [30]. GLRLM was first proposed by Galloway. It is defined as the frequency of occurrence of contiguous voxels with some run length along certain direction that have the same grey level. It characterizes the distribution of combination of grey levels in different directions. Example features includes short/long run emphasis, low/high grey level run emphasis, etc. GLSZM gives the statistics of groups of voxels that are connected and have a specific grey level, with features such as small/large size zone emphasis. The NGTDM is thought to provide more human-like perception of texture such as: coarseness, contrast, busyness, and complexity.

2.1.2 Dynamic features

For dynamic imaging protocols, such as 4D MRI, CT and PET, features based on kinetic analysis using tissue compartment models and parameters related to transport and binding rates can be extracted [37]. Compartmental modeling is used to describe systems varying in time but not in space. In the case of FDG, a 3-compartment model could be used to depict the trapping of FDG-6-Phosphate (FDG6P) in tumor [38-40]. Glucose metabolic uptake rate could be evaluated from compartmental modeling. Values of influx rate constant from compartment modeling was shown to be able to offer an assessment for inflammations at different locations of the body for non-small cell lung carcinoma [41]. For dynamic contrast enhanced MRI, the Toft and Kermode (TK) model is one of the most popular compartment models, providing information about the influx forward volume transfer constant from plasma into the extravascular-extracellular space [42-45]. Lee et al. found that three dynamic parameters were correlated with the dose of radiation delivered to the parotid gland and the degree of radiation-induced parotid atrophy during the treatment of head and neck cancer [46].

The radiomic features are in general defined independently on the image modality. Yet, there are some variations in the nomenclature used depending on the different imaging techniques. For instance, the use of SUV in PET image quantitative analysis, which are used instead of the raw counts intensity values. Therefore, basic features such as maximum, minimum, mean, standard deviation (SD), and coefficient of variation (CV) are usually

expressed as SUV_{max} , SUV_{mean} , etc. Total lesion glycolysis (TLG) is defined as the product of volume and mean SUV [47-49].

Radiomic modeling consists of three main steps: feature selection and/or extraction, model construction, and evaluation. Feature selection and/or extraction is necessary due to the high dimensionality of radiomic features, which focuses on obtaining the optimal feature subset or representation that is relevant and non-redundant to reduce the risk of overfitting. There are supervised and unsupervised ways for feature selection and extraction. Supervised methods contain three categories -- filter, wrapper and embedded methods. Filter method is a ranking-based method using a relevance index, which is a metric that measures how relevant the feature is to the data or the outcome. Correlation based (e.g. Spearman, Pearson), relief, and classical test statistics are some examples of this category. Wrapper method, e.g. SVM-based feature selection methods select features based on a single classifier performance of certain feature. Embedded method, e.g. LASSO (least absolute shrinkage and selection operator), implements feature selection and model building simultaneously [50]. Common unsupervised methods include principle component analysis (PCA) [51], clustering and t-Distributed Stochastic Neighbor Embedding (t-SNE) [52]. PCA uses an orthogonal linear transformation to convert the data into a new coordinate system so that large variances are projected to orthogonal coordinates. Clustering is another feature extraction algorithm which aims at finding relevant features and combining them by their cluster centroids based on some similarity measure, such as K-means and hierarchical clustering. tSNE is a dimension reduction method that is capable of retaining the local structure (pairwise similarity) of data, while revealing some important global structure.

Varieties of machine learning algorithms can be applied to the extracted feature representation based on the type of the problem, time dependent (survival) or independent (classification) [53]. Classification algorithms include but are not limited to logistic regression, support vector machines (SVM), random forests (RF), and neural networks. Logistic regression is a linear method that fits the coefficients of the variables to predict a logit transformation of the probability of the presence of the event [54, 55]. SVM tries to

find an optimal hyperplane that separates the classes as wide as possible with acceptable misclassified cases [56]. Different kernel functions (e.g. radial basis function) can be used to perform non-linear classification by mapping to higher dimensional feature space. Decision trees are popular in medical application due to sequential reasoning similar to human thinking [57, 58]. Random forests solved the high variance issue of decision trees by applying bootstrap aggregation [58]. Semi-supervised learning is available for the rich unlabeled data in medical field. Rosenberg et al. presented a semi-supervised learning which utilizes additional labelled data obtained from its predictions for self-training [59]. Transductive SVM (TSVM) tries to keep the unlabeled data as far away from the margin as possible [60]. Graph-based methods construct a graph connecting similar observations and enable the class information being transported through the graph [61]. Cox regression [62], random survival forests [63] and support vector survival [64] methods, as well as recently proposed deep learning based survival models [65, 66] are available to investigate the presence of a set of variables that may affect survival time.

Before applying any models in clinical practice, large cohort external validation and preferably clinical trials should be implemented. However, for a lot of studies, even the collection of a relatively large retrospective training dataset can be challenging [67]. Thus, proper validation techniques are vital for the evaluation of constructed models. Bootstrapping technique is usually relied upon to estimate confidence intervals for statistical models. Cross-validation is another widely applied technique for both hyperparameter optimization and prediction error estimation. The receiver operating characteristic (ROC) curve, or the area under the ROC curve (AUC) quantifies the sensitivity and specificity of the model and represents the probability that a randomly selected case is correctly distinguished with a larger probability than a randomly selected case is incorrectly distinguished for the classification task. C-index is a measure of the goodness of fit for survival models, which is the probability of concordance between predicted and observed survival. Details of different types of multivariate predictive models and validation can be found in the TRIPOD guidelines, which should be adhered to as much as possible in any radiomics study [68]. In addition, it is essential to ensure that there is no “double dipping” into the training data for both feature selection, hyper-

parameter optimization and model selection, which will lead to overfitting and optimistic results that will not generalize well for new data [58, 69-73]. Nested cross-validation is recommended if no test set is available. Except for the lack of data, imbalance is another challenge in radiotherapy modeling. There are some approaches that could help with this issue, such as upsampling the minority class, downsampling the majority class or generate synthetic data by learning the distribution. Synthetic Minority Over-sampling Technique (SMOTE) and generative adversarial network (GAN) are two popular methods that can be utilized [74, 75].

2.2 Machine-derived features and deep learning methods

Handcrafted features are designed as to incorporate prior expert knowledge, however, the tedious designing process and the unavoidable loss of underlying information from the raw images may limit the discriminative power of handcrafted feature-based methods. Alternatively, the development of both hardware (GPU) and deep learning technologies based on multi-layer neural networks, especially the convolutional neural networks (CNN) has enabled the automated discovery of relevant features in the input images using a general learning procedure. Unlike the conventional ML approach, the processes of data representation and prediction (e.g., classification or regression) are performed jointly in deep learning. Typical architectures contain multilayers of varying modules (e.g., convolution or pooling) with usually non-linear activation functions (e.g. Relu, sigmoid) for learning the representations of data and subsequent fully connected layers for classification/regression tasks. Multiple levels of abstraction could be obtained by the data representation CNN layers (for example, the shallow levels may represent edges, motifs in an image, deeper layers amplify characteristics that are important for discrimination and suppress irrelevant variations) [76]. Deep learning methods also bypass the limitation of possible feature selection bias in conventional feature-based methods. The design of CNN networks could be from-scratch or leveraged from pre-trained networks, such as VGG [77] or ResNet [14]. These pre-trained networks can be used as fixed feature extractors for small datasets, or they can be fine-tuned for larger ones using the technique called transfer learning, which is very commonly used in medical image analysis.

Similar to conventional machine learning methods, there are supervised, unsupervised and semi-supervised approaches. Compared with vanilla neural networks, CNNs apply local filters to the input space, which exploit the strong local correlation in images and reduce the number of weights significantly by sharing weights for each filter. Recurrent neural networks (RNN) are suitable for sequential inputs and use a 'state vector' to contain information of the history of the past elements of the sequence. Two types of RNN – Long short-term memory (LSTM) and Gated recurrent units (GRU) are widely used in practice to solve the problem of vanishing or exploding gradients for long sequences by internal gates that are able to learn which data in the sequence is important to keep or discard [78, 79]. Autoencoders (AEs) consists of an unsupervised architecture, which is used for latent representative feature extraction. Variations of the AEs are proposed for better performance in specific tasks, such as variational autoencoders that resemble the original AE, variational Bayesian methods to learn a probability distribution that represents the data [80] and convolutional autoencoders that could be fed with images as inputs [81]. Another unsupervised network - Restricted Boltzmann machine (RBM), obtains a joint probability distribution of inputs and activations, with the forward pass learning the probability of activations given the inputs, whereas the backward pass estimates the probability of inputs given activations [82]. Deep belief networks are probabilistic generative models that are composed of multiple layers of stochastic, latent variables, which can also be viewed as a stack of RBMs. Similar as the conventional methods, semi-supervised approaches combining unsupervised and supervised learning by training the supervised network with an additional loss component from the unsupervised generative models (e.g. AEs, RBMs) could be applied for deep learning scheme [83].

Another advantage of deep learning methods is the flexibility, enabling structured inputs (sequential data), structured outputs (image-to-image) or other structures based on specific tasks. For instance, in addition to classification, segmentation, registration, and lesion detection are widely explored by deep learning techniques. Fully CNN (FCN) can merge features learned at different stages in the encoders and then upsample low resolution feature maps by deconvolutions [84]. U-Net is similar to FCN, but has a nearly symmetric U-shaped (encoding/decoding) network architecture [85].

3 Multimodality radiomics

Multimodality imaging is able to provide better ROIs delineation, thus can benefit the radiation therapy planning. Moreover, accurate and carefully designed fusion of anatomic and functional imaging could add great value to the outcome modeling [86]. Indeed, it allows the capture of complementary features from target regions. In practice, clinicians exploit and combine information from different modalities to reach a diagnosis. The goal of machine learning approaches in that context is to extract quantitative features from all these available modalities and rely on conventional machine learning methods (or design novel architectures for deep learning methods) in order to provide complementary biomarkers to facilitate a more efficient and accurate diagnosis or prognosis. The potential applications of multimodality imaging could be summarized in two aspects: better delineation of tumors, thus improving diagnosis and staging, and more precise outcome modeling. These two applications will further enable better radiotherapy treatment delivery and more personalized, adaptive treatment planning.

Better definition of ROIs is the basis of diagnosis, staging, accurate radiation treatment and reliable radiomics modeling. For instance, a study of fractionated stereotactic radiotherapy for meningioma patients by Milker-Zabel et al. demonstrated improved target definition by combining physiological information from PET, anatomical structure from CT and soft tissues contrasts from MRI, resulting in alterations of the original contour definitions in 73 percent of the cases [87]. Fusion of different modality information is the key to achieve better performance. James et al. made a survey for medical image fusion, in which they proposed the fusion to be a process of registering and combining multiple images from multiple modalities to improve the imaging quality and reduce randomness to increase the clinically applicability for medical problems [88]. They discussed some major image fusion methods, including morphological, knowledge-based, wavelet-based, neural network based methods. Motivated by related work, Guo et al. came up with an abstraction of image fusion methods that includes fusing at the feature level, the classifier level and decision-making level [89]. Fusing at feature level means that multimodality images are used together to learn a combined feature set; fusing at classifier level means that images

from each modality are used as separate inputs to learn individual feature sets and then learn a multimodal classifier; fusing at decision-making level means that each modality is analyzed separately to build a single-modality model, and then the outputs are fused (e.g. with a voting scheme). Guo et al. suggested that for a specific application, one should consider all these three strategies and select the most suitable one. In this review, multimodality methods are categorized into radiomic based and deep learning based. Yet, both methods could be divided into the three strategies summarized in Guo's work. We gave some application examples below to offer readers a comprehensive view of possible pathways to solve the multimodality modeling problem.

1. Radiomic-based modeling

One straightforward way to build models is to combine all the radiomic features extracted separately from different modalities. Beukinga et al. extracted texture features from both PET and CT, combining with clinical and geometric features, using LASSO to build/develop a classifier for predicting response to neoadjuvant chemoradiotherapy in esophageal cancer, they found the multimodality model was superior to the standard SUVmax method [90]. Vaidya et al. used pre-treatment FDG-PET/CT from twenty-seven NSCLC patients for local and loco-regional recurrence [91]. Thirty-two tumor region features, intensity-volume-histogram (IVH) and textural characteristics were extracted and further analyzed using Spearman's correlation (r_s) and multivariate logistic regression. A two-parameter model of PET-V80 and CT-V70 yielded $r_s = 0.4854$ ($p = 0.0067$) and $r_s = 0.5908$ ($p = 0.0013$), respectively for loco-regional and local failures. They showed that multimodality image-feature modeling provides better performance compared to existing metrics. Fig. 2 and 3 showed the PET/CT fused images and the imaging features being used in this work [91]. Lartizien et al. reported that the best performance was achieved by the SVM classifier combined with the 12 most discriminant PET and CT features with a value of the area under the receiver operating curve of 0.91 for the staging of lymphoma patients [92]. Desseroit et al. found that PET entropy and CT zone percentage can be used to create a nomogram with higher stratification power than staging alone in non-small-cell lung cancer stage I–III [93]. Lucia et al. showed that entropy and gray-level non-uniformity from diffusion weighted imaging (DWI) MRI and PET images were independent predictors

of recurrence and locoregional control for locally advanced cervical cancer treated with chemoradiotherapy, with significantly higher prognostic power than clinical parameters [94]. However, directly combining radiomic features extracted separately from each modality might not make a full use of the underlying biological correlation. Thus, coming up with novel ways to fuse the information from different modalities is a key and challenging step to obtain potentially better prediction models. Vallières et al. developed a joint FDG-PET and MRI texture based model for the early evaluation of lung metastasis risk in soft-tissue sarcomas (STSs) [95]. For a cohort of 51 patients with pre-treatment FDG-PET and MRI scans (T1-weighted and T2-weighted fat-suppression), 9 non-texture features and 41 texture features were extracted from the tumor regions of separate scans (PET, T1, T2) and fused scans (PET/T1, PET/T2). Wavelet transform was used for the fusion of scans. The logistic regression with imbalance adjusted bootstrap resampling was used for the modeling. They found that texture features extracted from fused scans significantly outperformed those from separate scans in terms of lung metastases prediction estimates. The best performance was obtained using a combination of four texture features extracted from FDG-PET/T1 and FDGPET/T2FS scans. This model reached an area under the receiver-operating characteristic curve of 0.984 ± 0.002 , a sensitivity of 0.955 ± 0.006 , and a specificity of 0.926 ± 0.004 in bootstrapping evaluations. Fig. 4 presented the comparison PET, MRI scans (T1, T2) for patients who did not develop (top row) and developed (bottom row) lung metastases. Fig. 5 showed the performance of PET scan features alone, separate scans features and fused scans features. In their study, a wavelet transform for the fusion of PET and MRI scans was applied. Chaddad et al. proposed to derive radiomic features from joint intensity matrix (JIM) and grey level co-occurrence matrix (GLCM) of the T2-Weighted (T2-WI) and apparent diffusion coefficient (ADC) images. Combined JIM and GLCM features provided the best performing AUC for predicting the Gleason score for prostate cancer patients. JIM computes the joint intensity distribution between 3D images of different modalities [96]. Mu et al. proposed to extract radiomic features from both original and fused PET/CT images to predict immunotherapy response in non-small cell lung cancer (NSCLC) patients. PET and CT were first normalized and fused (added voxel-wise with a weight parameter for PET images). They found the fused images and corresponding features have improved the classification results

compared to single modality images [97]. Vallières et al. published a significant study on enhancing the multimodality radiomics based model via optimization of PET and MR images acquisition protocols. Using simulated images, they found that it is feasible to identify an optimal set of image acquisition and reconstruction parameters that could lead to improve the final prediction performance of extracted features [98]. For these kinds of fusion, there is a prerequisite that the different modality images should be perfectly aligned. Integrated hardware systems such as PET/CT scanners or PET/MRI scanners can provide better aligned datasets; otherwise, software solutions, such as rigid or deformable registration techniques need to be deployed [99] [100]. Yang et al. developed several tools for this purpose such as MIASYS [101] and DIRART [102].

Fig. 2 Example images for fused PET/CT of one NSCLC patient [91].

Fig. 3 Top row: Example of Intensity volume histogram from GTV and PTV of PET and CT; bottom row: example of texture map for CT and PET images [91].

Fig. 4 FDG-PET and MRI diagnostic images of two patients with soft-tissue sarcomas of the extremities. Top row: patient that did not develop lung metastases. Bottom row: patient that eventually developed lung metastases. 1st column: FDG-PET images, axial plane. 2nd column: T1-weighted images, axial plane. 3rd column: T2- weighted fat-saturated images, axial plane. 4th column: short tau inversion recovery images, sagittal plane. The green lines in the two images of the 4th column correspond to the plane shown in the three other respective images [95].

Fig. 5 Estimation of prediction performance of multivariable models constructed from FDG-PET scans, SEPARATE scans, and FUSED scans using optimal degrees

of freedom on texture extraction parameters, for model orders of 1–10 [95].

2. Deep learning based modeling

Carneiro et al. used craniocaudal (CC) and mediolateral oblique (MLO) views and their respective mass and micro-calcification segmentations of the breast for training separate CNN models through an ImageNet pre-trained model. Then, using the features learnt from each segmentation map and unregistered views, a final CNN classifier to estimate the patient's risk of developing breast cancer was built [103]. Chang et al conducted an interesting multimodality study using four MRI sequences (FLAIR, T2, T1 precontrast and T1 postcontrast) and residual convolutional neural network for the determination of IDH status in low- and high- grade gliomas. They exploited a dataset of 703 patients from three hospitals. MRI images were first resampled to isotropic 1 mm voxels, followed by registration of T1 precontrast, T2, and FLAIR images to T1 postcontrast images. N4 bias correction was applied to remove low frequency intensity nonuniformity. Image intensities were normalized by subtracting the median intensity of normal brain and then divided by the interquartile intensity of normal brain. Slices containing tumor from three spatial dimensions were fed into the network. For the fusion of multimodality images, they implemented three training heuristics. In the combined network, all sequences and dimensions were input to a single ResNet with input size $12 \times 142 \times 142$. In the dimension network, a single ResNet with input size $4 \times 142 \times 142$ was trained for each dimension and the sigmoid probability of each network was combined with a logistic regression. In the sequence network, a separate network with input size $3 \times 142 \times 142$ for each MRI sequence was trained and the outputs were combined using logistic regression the same as the second heuristic. The network structure is shown in Fig. 6. They found that combining the sequence network and age resulted in the highest performance, with AUC of 0.95 in the validation and test sets [104].

Fig. 6 The training heuristics tested include a single combined network (A), dimensional networks (B), and sequence networks (C). In the single combined network training heuristic, all sequences and dimensions were inputted into a single network. In the dimensional networks training heuristic, a separate network was trained for each dimension. In the sequence networks training heuristics, a separate network was trained for each MR sequence [104].

Zhang et al. segmented infant brain tissue images into white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF) using T1, T2, and fractional anisotropy (FA) images as inputs for the deep CNN network. They directly input image patches from three modalities into the CNN and then obtain the prediction for the center pixel of each patch. It can be observed that the combination of different image modalities invariably yielded higher performance than any of the single image modality [105]. As mentioned above, Guo et al. experimented on architectures of fusing at feature-level, classifier level and decision-level for segmentation using CNN based on publicly available soft-tissue sarcoma (STS) dataset. Type I: patches from different modalities are transformed into 3-D tensor ($28 \times 28 \times k$, k the number of modalities) and input to the network. Type II: features were learnt separately through each modality's CNN and then classified by a single fully connected network. Type III: a single-modal CNN was trained and the patch-wise probability were ensembled using random forests. Average DICE of Type-I, II, and III fusion networks on PET + CT + T1 was 82%, 80%, and 77%, respectively; of Type-I, II, and III fusion networks on PET + CT + T2 is 85%, 85%, and 84%, respectively. Average performance of a single modality network is 76%, 68%, 66%, and 80% for PET, CT, T1,

and T2 images. The multimodality fusion networks were statistically better than single-modality networks. The segmentation architectures results were shown in Fig. 7.

Fig. 7 Illustration of the structure for (a) Type-I fusion networks, (b) Type-II fusion network and (c) Type-III fusion network [89].

Nie et al. used contrast-enhanced T1 MRI, diffusion tensor imaging (DTI) and resting-state functional MRI (rs-fMRI), for computing multiple metric maps (including various diffusivity metric maps derived from DTI, and also the frequency-specific brain fluctuation amplitude maps and local functional connectivity anisotropy-related metric maps derived from rs-fMRI) to predict the survival time for 68 high-grade glioma patients [106]. Both rs-fMRI and DTI images are co-registered to the T1 MRI of the same subject. 6 diffusivity metric maps: fractional anisotropy (FA), mean diffusivity (MD), the first/second/third eigenvalue of the tensor (λ_1 , λ_2 , λ_3), and radial diffusivity (RD) together with B0 ($b=0$

s/mm²) map were obtained from DTI. These 7 metric maps constitute 7 channels. For the rs-fMRI modality, frequency-specific BOLD fluctuation power maps calculated in five non-overlapping frequency bands, resulted in 5 metric maps (freq-fMRI). The functional connectivity tensor (fTensor-fMRI) was used to provide functional information in white matter. The bounding box containing tumors were applied for all channels of the metric maps. 8 non-overlapping patches within the bounding box of each metric map were extracted and fed into the 3D CNN network. Supervised features were then learned and combined with other demographic features. PCA and sparse representation were used for feature selection and reduction before final step of using SVM to do the prediction. They found that about 7% improvement of accuracy can be reached by fusing multi-modal images. Fig. 8 is their network architecture. Fig. 9 is the comparison of single modality and multimodality performance.

Fig. 8 Schematic description of the proposed survival prediction framework for high grade glioma patients by using (1) 3D CNN-based deep learning to conduct feature learning and (2) an SVM for final prediction [106].

Fig. 9 OS prediction results using different imaging modalities. “Proposed” denotes the prediction result by combining all modalities together [106].

Antropova et al. came up with a method that extracted low- to mid-level features using a pre-trained CNN and fused them with handcrafted radiomic features through a SVM for classification of lesions in three different image modalities (mammography, ultrasound and DCE-MRI images) [107]. They showed that the fusion-based method obtained significant improvements to previous breast cancer computer-aided diagnosis methods (CADx) across all three imaging modalities in terms of predictive performance in the task of estimating lesion malignancy. Fig. 10 shows a diagonal classifier agreement plot between the CNN-based classifier and the conventional CADx classifier for FFDM. The x-axis denotes the output from the CNN-based classifier, and the y-axis denotes the output from the conventional CADx classifier. Each point represents a ROI for which predictions were made. Points near or along the diagonal from bottom left to top right indicate high classifier agreement; points far from the diagonal indicate low agreement. ROI pictures of extreme examples of agreement/disagreement are included.

Fig. 10 A diagonal classifier agreement plot between the CNN-based classifier and the conventional CADx classifier for FFDM [107].

Lao et al. extracted a total of 1403 handcrafted features and 98304 deep features from preoperative multi-modality MR images for the prediction of overall survival (OS) in patients with Glioblastoma Multiforme (GBM) [108]. The image modalities being used were T1-weighted, T1-weighted Gadolinium contrast-enhanced, T2-weighted, and T2-weighted FLAIR images (short for T1, T1C, T2, and FLAIR). After feature selection, a six-deep-feature signature was constructed by using the least absolute shrinkage and selection operator (LASSO) Cox regression model. They showed their proposed signature achieved better performance for prediction of OS (C-index: 0.739 [95% CI: 0.686, 0.792], and significant stratification of patients into prognostically distinct groups ($P < 0.001$ HR =

4.608 [95% CI: 1.884, 11.270]). Fig. 11 showed the work flow for this study and Fig. 12 showed the CNN used for deep learning feature extraction.

Fig. 11 The workflow of radiomics analysis [108].

Fig. 12 Illustration of deep features extraction. LRN is short for Local Response Normalization [108].

4 Discussion

Intuitively, multimodality-based models are likely to outperform single modality ones, since different modalities will emphasize on different characteristics of the tumor. In all the studies we have gone through as relevant examples, the superiority of multimodality based models over the single modality models was indeed observed. However, there are some challenges in multimodality modeling. A major one is how to optimally combine different images and to achieve the best results. There is indeed a large flexibility in the design of fusion algorithms which makes it nontrivial and challenging. Guo et al. provided

an abstraction of fusion, which is to fuse at feature level, classifier level or decision level. This offers a general idea of fusion techniques in practice, though for specific task, we still cannot avoid the variation and should consider carefully to fit the task better.

Another barrier is the preprocessing for multimodality images. In most of the above mentioned studies, images were co-registered to one of them before the analysis, which will be a very time-consuming step for relatively large patient cohort for manual effort. The co-registration is beneficial, since it can provide the same coordinate for different images, thus ease the learning process for DNN or for transferring the tumor contours across images to avoid manual contouring of tumors, so that either radiomic or machine learnt features could be extracted from the ROIs. Some auto-registration methods might be helpful in this scheme [109-113].

Another challenge is the choice of the learning model, the radiomic based or the DNN based methods, which remains an open question for different cases. In general, radiomic features are able to encode prior expert knowledge, but suffer from the need to select a relevant subset, that could introduce statistical bias into the modeling. On the other hand deep learning methods are very powerful without the tedious feature designing process. However, they are very data hungry, which is a difficult condition to satisfy in the medical imaging fields, especially for multimodality images. With the widely usage of multimodality imaging technique in clinical practice, this hopefully will be less a problem in the future. Finally, we also emphasized on several studies that tried to combine radiomic features and DNN to achieve better prediction results.

5 Conclusions

The increasing use of multimodality imaging in the clinic, the tremendous advances in the field of radiomics and deep learning, and the breakthroughs in hardware (GPU) have enabled the research in multimodality based medical services such as diagnosis, outcome prediction, treatment planning, which holds the potential to provide more personalized treatment and better patient health care. However, there still exists some challenges in the availability of data (small sample size and incompleteness), image registration and preprocessing, and the increased complexity with the design of optimal models.

Nonetheless, this review has shown that the multimodality research already resulted in improved performance in different aspects including ROI delineation and response prediction. More innovation in algorithms as well as the hardware, is expected to continue in this field in the future.

6 Acknowledgements

This work was partly supported by the National Institute of Health (NIH) grants R37-CA222215, and R01-CA233487.

7 References

1. Goel R, Moore W, Sumer B, Khan S, Sher D, Subramaniam RM. Clinical practice in PET/CT for the management of head and neck squamous cell cancer. *American Journal of Roentgenology*. 2017;209:289-303.
2. Castaldi P, Leccisotti L, Bussu F, Micciche F, Rufini V. Role of 18F-FDG PET-CT in head and neck squamous cell carcinoma. *Acta Otorhinolaryngologica Italica*. 2013;33:1.
3. Kligerman S, Digumarthy S. Staging of non-small cell lung cancer using integrated PET/CT. *American Journal of Roentgenology*. 2009;193:1203-11.
4. Greenspan BS. Role of PET/CT for precision medicine in lung cancer: perspective of the Society of Nuclear Medicine and Molecular Imaging. *Translational lung cancer research*. 2017;6:617.
5. Chao F, Zhang H. PET/CT in the staging of the non-small-cell lung cancer. *BioMed Research International*. 2012;2012.
6. Hochegger B, Alves GRT, Irion KL, Fritscher CC, Fritscher LG, Concato NH, et al. PET/CT imaging in lung cancer: indications and findings. *Jornal Brasileiro de Pneumologia*. 2015;41:264-74.
7. Yazarbas U, Avci NC, Yeniay L, Argon AM. The value of 18F-FDG PET/CT imaging in breast cancer staging. *Bosnian journal of basic medical sciences*. 2018;18:72.
8. Brendle C, Schwenzer N, Rempp H, Schmidt H, Pfannenberger C, La Fougère C, et al. Assessment of metastatic colorectal cancer with hybrid imaging: comparison of reading performance using different combinations of anatomical and functional imaging techniques in PET/MRI and PET/CT in a short case series. *European journal of nuclear medicine and molecular imaging*. 2016;43:123-32.
9. Kirchner J, Sawicki LM, Deuschl C, Grüneisen J, Beiderwellen K, Lauenstein TC, et al. 18 F-FDG PET/MR imaging in patients with suspected liver lesions: Value of liver-specific contrast agent Gadobenate dimeglumine. *PloS one*. 2017;12:e0180349.

10. Eiber M, Weirich G, Holzapfel K, Souvatzoglou M, Haller B, Rauscher I, et al. Simultaneous ⁶⁸Ga-PSMA HBED-CC PET/MRI improves the localization of primary prostate cancer. *European urology*. 2016;70:829-36.
11. Yankeelov TE, Abramson RG, Quarles CC. Quantitative multimodality imaging in cancer research and therapy. *Nature Reviews Clinical Oncology*. 2014;11:670.
12. Aerts HJ. The potential of radiomic-based phenotyping in precision medicine: a review. *JAMA oncology*. 2016;2:1636-42.
13. Hatt M, Tixier F, Visvikis D, Le Rest CC. Radiomics in PET/CT: more than meets the eye? *Journal of Nuclear Medicine*. 2017;58:365-6.
14. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*; 2016. p. 770-8.
15. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*. 2015;115:211-52.
16. Cheng J-Z, Ni D, Chou Y-H, Qin J, Tiu C-M, Chang Y-C, et al. Computer-aided diagnosis with deep learning architecture: applications to breast lesions in US images and pulmonary nodules in CT scans. *Scientific reports*. 2016;6:24454.
17. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542:115.
18. Conti P, Strauss L. The applications of PET in clinical oncology. *J Nucl Med*. 1991;32:623-48.
19. Chirra P, Leo P, Yim M, Bloch BN, Rastinehad AR, Purysko A, et al. Empirical evaluation of cross-site reproducibility in radiomic features for characterizing prostate MRI. *Medical Imaging 2018: Computer-Aided Diagnosis: International Society for Optics and Photonics*; 2018. p. 105750B.
20. Nyúl LG, Udupa JK, Zhang X. New variants of a method of MRI scale standardization. *IEEE transactions on medical imaging*. 2000;19:143-50.
21. Kahali S, Adhikari SK, Sing JK. On estimation of bias field in MRI images: polynomial vs Gaussian surface fitting method. *Journal of Chemometrics*. 2016;30:602-20.
22. Thibault G, Angulo J, Meyer F. Advanced statistical matrices for texture characterization: application to cell classification. *IEEE transactions on bio-medical engineering*. 2014;61:630-7. doi:10.1109/tbme.2013.2284600.
23. Yip SS, Aerts HJ. Applications and limitations of radiomics. *Physics in medicine and biology*. 2016;61:R150-66. doi:10.1088/0031-9155/61/13/r150.
24. Zwanenburg A, Leger S, Vallières M, Löck S. Image biomarker standardisation initiative. *arXiv preprint arXiv:161207003*. 2016.
25. van Velden FH, Kramer GM, Frings V, Nissen IA, Mulder ER, de Langen AJ, et al. Repeatability of Radiomic Features in Non-Small-Cell Lung Cancer [(18)F]FDG-PET/CT Studies: Impact of Reconstruction and Delineation. *Molecular imaging and biology : MIB : the official publication of the Academy of Molecular Imaging*. 2016;18:788-95. doi:10.1007/s11307-016-0940-2.
26. Leijenaar RT, Nalbantov G, Carvalho S, van Elmpt WJ, Troost EG, Boellaard R, et al. The effect of SUV discretization in quantitative FDG-PET Radiomics: the need

for standardized methodology in tumor texture analysis. *Sci Rep*. 2015;5:11075. doi:10.1038/srep11075.

27. Jain AK. *Fundamentals of digital image processing*. Englewood Cliffs, NJ: Prentice Hall; 1989.

28. O'Sullivan F, Roy S, O'Sullivan J, Vernon C, Eary J. Incorporation of tumor shape into an assessment of spatial heterogeneity for human sarcomas imaged with FDG-PET. *Biostat*. 2005;6:293-301. doi:10.1093/biostatistics/kxi010.

29. Castleman KR. *Digital image processing*. Englewood Cliffs, N.J.: Prentice Hall; 1996.

30. Haralick R, Shanmugam K, Dinstein I. Texture Features for Image Classification. *IEEE Trans on Sys Man and Cyb SMC*. 1973;3:610-21.

31. Zhang J, Tan T. Brief review of invariant texture analysis methods. *Pattern Recognition*. 2002;35:735-47. doi:[http://dx.doi.org/10.1016/S0031-3203\(01\)00074-7](http://dx.doi.org/10.1016/S0031-3203(01)00074-7).

32. Castellano G, Bonilha L, Li LM, Cendes F. Texture analysis of medical images. *Clinical Radiology*. 2004;59:1061-9. doi:<http://dx.doi.org/10.1016/j.crad.2004.07.008>.

33. Haralick RM, Shanmugam K. Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics*. 1973:610-21.

34. Amadasun M, King R. Textural features corresponding to textural properties. *IEEE Transactions on systems, man, and Cybernetics*. 1989;19:1264-74.

35. Galloway MM. Texture analysis using grey level run lengths. NASA STI/Recon Technical Report N. 1974;75.

36. Thibault G, Fertil B, Navarro C, Pereira S, Cau P, Levy N, et al. Shape and texture indexes application to cell nuclei classification. *International Journal of Pattern Recognition and Artificial Intelligence*. 2013;27:1357002.

37. Watabe H, Ikoma Y, Kimura Y, Naganawa M, Shidahara M. PET kinetic analysis--compartmental model. *Ann Nucl Med*. 2006;20:583-8.

38. Graham MM, Peterson LM, Hayward RM. Comparison of simplified quantitative analyses of FDG uptake. *Nuclear Medicine and Biology*. 2000;27:647-55.

39. Patlak CS, Blasberg RG. Graphical evaluation of blood-to-brain transfer constants from multiple-time uptake data. Generalizations. *J Cereb Blood Flow Metab*. 1985;5:584-90.

40. Morris ED, Endres CJ, Schmidt KC, Christian BT, Muzic RF, Fisher RE. Kinetic modeling in positron emission tomography. *Emission Tomography: The Fundamentals of PET and SPECT Academic*, San Diego. 2004.

41. Yang Z, Zan Y, Zheng X, Hai W, Chen K, Huang Q, et al. Dynamic FDG-PET imaging to differentiate malignancies from inflammation in subcutaneous and in situ mouse model for non-small cell lung carcinoma (NSCLC). *PloS one*. 2015;10:e0139089.

42. Chikui T, Obara M, Simonetti AW, Ohga M, Koga S, Kawano S, et al. The principal of dynamic contrast enhanced MRI, the method of pharmacokinetic analysis, and its application in the head and neck region. *International journal of dentistry*. 2012;2012.

43. Tofts PS. Modeling tracer kinetics in dynamic Gd - DTPA MR imaging. *Journal of magnetic resonance imaging*. 1997;7:91-101.
44. Tofts PS, Brix G, Buckley DL, Evelhoch JL, Henderson E, Knopp MV, et al. Estimating kinetic parameters from dynamic contrast - enhanced T1 - weighted MRI of a diffusable tracer: standardized quantities and symbols. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*. 1999;10:223-32.
45. Leach M, Morgan B, Tofts P, Buckley D, Huang W, Horsfield M, et al. Imaging vascular function for early stage clinical trials using dynamic contrast-enhanced magnetic resonance imaging. *European radiology*. 2012;22:1451-64.
46. Lee FK-h, King AD, Ma BB-Y, Yeung DK-w. Dynamic contrast enhancement magnetic resonance imaging (DCE-MRI) for differential diagnosis in head and neck cancers. *European journal of radiology*. 2012;81:784-8.
47. Benz MR, Allen-Auerbach MS, Eilber FC, Chen HJJ, Dry S, Phelps ME, et al. Combined Assessment of Metabolic and Volumetric Changes for Assessment of Tumor Response in Patients with Soft-Tissue Sarcomas. *J Nucl Med*. 2008;49:1579-84. doi:10.2967/jnumed.108.053694.
48. Erdi YE, Macapinlac H, Rosenzweig KE, Humm JL, Larson SM, Erdi AK, et al. Use of PET to monitor the response of lung cancer to radiation treatment. *Eur J Nucl Med*. 2000;27:861-6.
49. Larson SM, Erdi Y, Akhurst T, Mazumdar M, Macapinlac HA, Finn RD, et al. Tumor Treatment Response Based on Visual and Quantitative Changes in Global Tumor Glycolysis Using PET-FDG Imaging. The Visual Response Score and the Change in Total Lesion Glycolysis. *Clin Positron Imaging*. 1999;2:159-71. doi:S1095039799000163 [pii].
50. Tibshirani R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B (Methodological)*. 1996:267-88.
51. Wold S, Esbensen K, Geladi P. Principal component analysis. *Chemometrics and intelligent laboratory systems*. 1987;2:37-52.
52. Maaten Lvd, Hinton G. Visualizing data using t-SNE. *Journal of Machine Learning Research*. 2008;9:2579-605.
53. Parmar C, Grossmann P, Bussink J, Lambin P, Aerts HJ. Machine learning methods for quantitative radiomic biomarkers. *Scientific reports*. 2015;5:13087.
54. Vallières M, Kay-Rivest E, Perrin LJ, Liem X, Furstoss C, Aerts HJ, et al. Radiomics strategies for risk assessment of tumour failure in head-and-neck cancer. *Scientific reports*. 2017;7:10117.
55. Wu S, Zheng J, Li Y, Wu Z, Shi S, Huang M, et al. Development and validation of an MRI-based radiomics signature for the preoperative prediction of lymph node metastasis in bladder cancer. *EBioMedicine*. 2018;34:76-84.
56. Zhang Y, Oikonomou A, Wong A, Haider MA, Khalvati F. Radiomics-based prognosis analysis for non-small cell lung cancer. *Scientific reports*. 2017;7:46349.
57. El Naqa I, Murphy MJ. What is machine learning? *Machine Learning in Radiation Oncology*: Springer; 2015. p. 3-11.

58. Hastie T, Tibshirani R, Friedman J, Franklin J. The elements of statistical learning: data mining, inference and prediction. *The Mathematical Intelligencer*. 2005;27:83-5.
59. Rosenberg C, Hebert M, Schneiderman H. Semi-Supervised Self-Training of Object Detection Models. *WACV/MOTION*. 2005;2.
60. Joachims T. Transductive inference for text classification using support vector machines. *ICML*; 1999. p. 200-9.
61. Blum A, Lafferty J, Rwebangira MR, Reddy R. Semi-supervised learning using randomized mincuts. *Proceedings of the twenty-first international conference on Machine learning*: ACM; 2004. p. 13.
62. Cox DR. Regression models and life - tables. *Journal of the Royal Statistical Society: Series B (Methodological)*. 1972;34:187-202.
63. Ishwaran H, Kogalur UB, Blackstone EH, Lauer MS. Random survival forests. *The annals of applied statistics*. 2008;2:841-60.
64. Van Belle V, Pelckmans K, Van Huffel S, Suykens JA. Support vector methods for survival analysis: a comparison between ranking and regression approaches. *Artificial intelligence in medicine*. 2011;53:107-18.
65. Katzman JL, Shaham U, Cloninger A, Bates J, Jiang T, Kluger Y. DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. *BMC medical research methodology*. 2018;18:24.
66. Ching T, Zhu X, Garmire LX. Cox-nnet: An artificial neural network method for prognosis prediction of high-throughput omics data. *PLoS computational biology*. 2018;14:e1006076.
67. Mayo CS, Kessler ML, Eisbruch A, Weyburne G, Feng M, Hayman JA, et al. The big data effort in radiation oncology: Data mining or data farming? *Advances in radiation oncology*. 2016;1:260-71.
68. Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Annals of internal medicine*. 2015;162:W1-W73.
69. Wood IA, Visscher PM, Mengersen KL. Classification based upon gene expression data: bias and precision of error rates. *Bioinformatics (Oxford, England)*. 2007;23:1363-70.
70. Singhi SK, Liu H. Feature subset selection bias for classification learning. *Proceedings of the 23rd international conference on Machine learning*: ACM; 2006. p. 849-56.
71. Neumann U, Riemenschneider M, Sowa J-P, Baars T, Kalsch J, Canbay A, et al. Compensation of feature selection biases accompanied with improved predictive performance for binary classification by using a novel ensemble feature selection approach. *BioData mining*. 2016;9:36.
72. Krawczuk J, Łukaszuk T. The feature selection bias problem in relation to high-dimensional gene data. *Artificial intelligence in medicine*. 2016;66:63-71.
73. Ambroise C, McLachlan GJ. Selection bias in gene extraction on the basis of microarray gene-expression data. *Proc Natl Acad Sci U S A*. 2002;99:6562-6. doi:10.1073/pnas.102102699.

74. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*. 2002;16:321-57.
75. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial nets. *Advances in neural information processing systems*; 2014. p. 2672-80.
76. LeCun Y, Bengio Y, Hinton G. Deep learning. *nature*. 2015;521:436.
77. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:14091556*. 2014.
78. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural computation*. 1997;9:1735-80.
79. Cho K, Van Merriënboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv:14061078*. 2014.
80. Kingma DP, Welling M. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*. 2013.
81. Li F, Qiao H, Zhang B. Discriminatively boosted image clustering with fully convolutional auto-encoders. *Pattern Recognition*. 2018;83:161-73.
82. Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. *science*. 2006;313:504-7.
83. Kingma DP, Mohamed S, Rezende DJ, Welling M. Semi-supervised learning with deep generative models. *Advances in neural information processing systems*; 2014. p. 3581-9.
84. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. *Proceedings of the IEEE conference on computer vision and pattern recognition*; 2015. p. 3431-40.
85. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. *International Conference on Medical image computing and computer-assisted intervention*: Springer; 2015. p. 234-41.
86. Reuzé S, Schernberg A, Orlhac F, Sun R, Chargari C, Dercle L, et al. Radiomics in nuclear medicine applied to radiation therapy: methods, pitfalls, and challenges. *International Journal of Radiation Oncology* Biology* Physics*. 2018;102:1117-42.
87. Milker-Zabel S, Zabel-du Bois A, Henze M, Huber P, Schulz-Ertner D, Hoess A, et al. Improved target volume definition for fractionated stereotactic radiotherapy in patients with intracranial meningiomas by correlation of CT, MRI, and [68Ga]-DOTATOC-PET. *Int J Radiat Oncol Biol Phys*. 2006;65:222-7.
88. James AP, Dasarathy BV. Medical image fusion: A survey of the state of the art. *Information Fusion*. 2014;19:4-19.
89. Guo Z, Li X, Huang H, Guo N, Li Q. Deep Learning-Based Image Segmentation on Multimodal Medical Imaging. *IEEE Transactions on Radiation and Plasma Medical Sciences*. 2019;3:162-9.
90. Beukinga RJ, Hulshoff JB, van Dijk LV, Muijs CT, Burgerhof JG, Kats-Ugurlu G, et al. Predicting response to neoadjuvant chemoradiotherapy in esophageal cancer with textural features derived from pretreatment 18F-FDG PET/CT imaging. *Journal of nuclear medicine*. 2017;58:723-9.

91. Vaidya M, Creach KM, Frye J, Dehdashti F, Bradley JD, El Naqa I. Combined PET/CT image characteristics for radiotherapy tumor response in lung cancer. *Radiother Oncol.* 2012;102:239-45.
92. Lartizien C, Rogez M, Niaf E, Ricard F. Computer-aided staging of lymphoma patients with FDG PET/CT imaging based on textural information. *IEEE journal of biomedical and health informatics.* 2013;18:946-55.
93. Desseroit M-C, Visvikis D, Tixier F, Majdoub M, Perdrisot R, Guillemin R, et al. Development of a nomogram combining clinical staging with 18 F-FDG PET/CT image features in non-small-cell lung cancer stage I–III. *European journal of nuclear medicine and molecular imaging.* 2016;43:1477-85.
94. Lucia F, Visvikis D, Desseroit M-C, Miranda O, Malhaire J-P, Robin P, et al. Prediction of outcome using pretreatment 18 F-FDG PET/CT and MRI radiomics in locally advanced cervical cancer treated with chemoradiotherapy. *European journal of nuclear medicine and molecular imaging.* 2018;45:768-86.
95. Vallières M, Freeman CR, Skamene SR, El Naqa I. A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities. *Physics in Medicine & Biology.* 2015;60:5471.
96. Chaddad A, Kucharczyk M, Niazi T. Multimodal radiomic features for the predicting gleason score of prostate cancer. *Cancers.* 2018;10:249.
97. Mu W, Qi J, Lu H, Schabath M, Balagurunathan Y, Tunali I, et al. Radiomic biomarkers from PET/CT multi-modality fusion images for the prediction of immunotherapy response in advanced non-small cell lung cancer patients. *Medical Imaging 2018: Computer-Aided Diagnosis: International Society for Optics and Photonics;* 2018. p. 105753S.
98. Vallières M, Laberge S, Diamant A, El Naqa I. Enhancement of multimodality texture-based prediction models via optimization of PET and MR image acquisition protocols: a proof of concept. *Physics in Medicine & Biology.* 2017;62:8536.
99. Leibfarth S, Mönnich D, Welz S, Siegel C, Schwenzer N, Schmidt H, et al. A strategy for multimodal deformable image registration to integrate PET/MR into radiotherapy treatment planning. *Acta oncologica.* 2013;52:1353-9.
100. Hajnal JV, Hawkes DJ, Hill DLG. *Medical image registration.* Boca Raton: CRC Press; 2001.
101. Yang D, Zheng J, Nofal A, Wu Y, Deasy J, El Naqa I. Techniques and software tool for 3D multimodality medical image segmentation *Journal of radiation oncology informatics.* 2009:accepted.
102. Yang D, Brame S, El Naqa I, Aditya A, Wu Y, Goddu SM, et al. Technical note: DIRART--A software suite for deformable image registration and adaptive radiotherapy research. *Med Phys.* 2011;38:67-77.
103. Carneiro G, Nascimento J, Bradley AP. Unregistered multiview mammogram analysis with pre-trained deep learning models. *International Conference on Medical Image Computing and Computer-Assisted Intervention: Springer;* 2015. p. 652-60.
104. Chang K, Bai HX, Zhou H, Su C, Bi WL, Agbodza E, et al. Residual convolutional neural network for the determination of IDH status in low-and high-grade gliomas from MR imaging. *Clinical Cancer Research.* 2018;24:1073-81.

105. Zhang W, Li R, Deng H, Wang L, Lin W, Ji S, et al. Deep convolutional neural networks for multi-modality isointense infant brain image segmentation. *NeuroImage*. 2015;108:214-24.
106. Nie D, Trullo R, Petitjean C, Ruan S, Shen D. Medical Image Synthesis with Context-Aware Generative Adversarial Networks. *ArXiv e-prints*; 2016.
107. Antropova N, Huynh BQ, Giger ML. A deep feature fusion methodology for breast cancer diagnosis demonstrated on three imaging modality datasets. *Medical physics*. 2017;44:5162-71.
108. Lao J, Chen Y, Li Z-C, Li Q, Zhang J, Liu J, et al. A deep learning-based radiomics model for prediction of survival in glioblastoma multiforme. *Scientific reports*. 2017;7:10353.
109. Cao X, Yang J, Wang L, Xue Z, Wang Q, Shen D. Deep learning based inter-modality image registration supervised by intra-modality similarity. *International Workshop on Machine Learning in Medical Imaging*: Springer; 2018. p. 55-63.
110. Hu Y, Modat M, Gibson E, Li W, Ghavami N, Bonmati E, et al. Weakly-supervised convolutional neural networks for multimodal image registration. *Medical image analysis*. 2018;49:1-13.
111. Zampieri A, Charpiat G, Girard N, Tarabalka Y. Multimodal image alignment through a multiscale chain of neural networks with application to remote sensing. *Proceedings of the European Conference on Computer Vision (ECCV)*; 2018. p. 657-73.
112. Balakrishnan G, Zhao A, Sabuncu MR, Guttag J, Dalca AV. An unsupervised learning model for deformable medical image registration. *Proceedings of the IEEE conference on computer vision and pattern recognition*; 2018. p. 9252-60.
113. Zhu X, Ding M, Huang T, Jin X, Zhang X. PCANet-Based Structural Representation for Nonrigid Multimodal Medical Image Registration. *Sensors*. 2018;18:1477.