PET Beyond Pictures

Eric Wolsztynski and Janet F. Eary

Conventional PET-Based Quantitation

Standardized Uptake Values

PET imaging data essentially consist of count and spatial information on photons emitted by positron decay of the PET radiotracer molecules within the body. The scanner output data represents the number of photons detected by the detectors at an imaging time point. This raw information is processed (or "reconstructed"), usually by the imaging device based hardware, to generate a final four-dimensional digital image of intensities that relate to these initial photon counts in the volumetric field of view, captured over a number of imaging time frames. We refer to a unit of this 3D image frame as a voxel (volume element). A higher intensity at a given voxel therefore indicates higher concentration of the administered radiotracer at that location.

In routine practice, the volumetric PET body tissue uptake information is converted into standardized uptake values (SUV) by adjusting the raw tissue radioactivity concentration measured within an image region of interest at time t, $C(t)$, for patient body weight W and injected dose D:

$$\text{SUV}(t) = C(t) / (D / W)$$

Tissue SUV values are commonly expressed in g/mL, with concentration level $C(t)$ in MBq/mL, and injected radioactive imaging agent dose per unit weight D/W in MBq/g. SUV measures are considered unitless, on the basis that $C(t)$ can be defined as concentration in soft tissue, itself with a mass density of about 1 g/mL.

The above uptake value standardization convention remains the predominant choice for image analysis in the literature and in clinical practice, although alternative methods have been considered for generation of initial tracer concentration than that provided by body weight W, which is very sensitive to patient physiology [1]. The most widespread of these alternatives is the lean SUV (SUL), using lean body mass W-BF, where BF is body fat, in place of W.

Volume of Interest Segmentation

PET-based quantitation consists of deriving statistical summaries of the image data-based volume of interest (VOI), many of which rely on accurate tumor size assessment. Delineation, or segmentation, of the VOI is therefore critical but

E. Wolsztynski (✉)
Insight Centre for Data Analytics, School of Mathematical Sciences, University College Cork, Cork, Ireland
e-mail: eric.w@ucc.ie

J. F. Eary
National Institutes of Health/NCI/DCTD, Bethesda, MD, USA
e-mail: janet.eary@nih.gov

can vary significantly with the method chosen. The use of hand-drawn VOI segmentation masks determined by trained staff is most commonly performed, but this process is tedious and prone to human error and interoperator variability. A wide range of semi- and fully automated alternatives are also available. Some approaches consist of identifying VOI voxels that have values above a segmentation threshold defined as a fraction of the maximum SUV in the image or as a multiplier of the average background SUV level [2]. Such thresholding schemes tend to be sensitive to image contrast and noise characteristics [1, 3]. Other techniques consist of identifying contours of the VOI adaptively from the imaging data, by evaluating the likelihood that a given voxel belongs to the tumor according to its direct neighborhood voxel characteristics [1, 2]. Multimodality image analysis techniques also exploit the additional information provided at higher transverse spatial resolution by a co-registered CT or MRI scan when available. For example, FMISO-PET VOI masks of glioma volumes may be determined using image appearance in FLAIR sequence MRI data [4, 5]. Reproducible segmentation seems more likely to be achieved using a combination of automatic edge-detection techniques and expert assessment of its output [2].

Conventional Summaries and Their Use at Baseline and Follow-Up

Once a tissue VOI is obtained, PET tracer uptake distribution is traditionally analyzed in terms of SUV summaries, namely, its maximum value (SUV_{max}), average value (SUV_{mean}), average value within the SUV_{max} tissue neighborhood (SUV_{peak}), metabolically active tumor volume (MATV), and total lesion glycolysis (TLG = $SUV_{mean} \times$ MATV) [1]. Note that the term "glycolysis" here stems from the use of FDG as a tracer for tissue metabolism. Other PET imaging agents report on different tissue processes or status. For example, F-18 fluoromisonidazole is used to determine the level of tissue oxygenation. In analyzing FMISO-PET data, a value for total lesion hypoxia can be

determined. Several variations exist for the calculation of SUV_{peak} using a number of PET imaging agents. One common approach uses the average SUV within a 1 cubic centimeter tissue sphere centered at the voxel with intensity SUV_{max}, but the neighborhood could be, e.g., the 9-voxel cube centered at SUV_{max} instead of a sphere. As for MATV, the tissue volume analyzed could be roughly approximated by the product of the number of voxels within the VOI and the unit volume of a single voxel. Some refinements exist to account for fractional voxels at VOI boundaries.

The above summaries are used essentially for characterization of primary lesions. Metastases assessments can also be performed. In most instances, similar tissue uptake determinations are performed. Typically, the number of lesions and parameters such as their size and SUV_{peak} values are reported. Methodologies vary with the image specialist and the type and nature of the disease.

For PET-based therapeutic response assessment, it is common to compare the most active lesions before and after therapy. Several sets of guidelines have been created. The Positron Emission Tomography Response Criteria in Solid Tumors (PERCIST) was published in 2009 in an effort to adapt the former WHO and (more MR- and CT-appropriate) Response Evaluation Criteria In Solid Tumors (RECIST) guidelines to the specifics of FDG PET (tissue metabolism) imaging data [1]. PERCIST classification rules determine whether there is "complete metabolic response," "partial metabolic response," "stable metabolic disease," or "progressive metabolic disease" following therapy, taking into account the number of lesions and the proportional changes in size and SUV_{peak} in all measurable lesions.

Risk Characterization and Predictive Validation

The prognostic value of PET-derived biomarkers is traditionally assessed by means of survival analyses, which can be carried out in a number of ways and are interpreted with respect to the

endpoint considered. Conventional endpoints used to describe patient risk, outcome, or therapeutic effectiveness are, most often, overall survival (OS) (duration of patient survival before they become lost to follow-up), progression-free survival (PFS), or disease-free survival (DFS). These endpoints are continuous in nature, in that all patients can be placed on a timeline that usually starts either on the day of diagnosis or baseline scan ($t = 0$), and allow for longitudinal survival analyses. Time to progression (TTP) is another commonly used reference, but OS and PFS are usually preferred. Other endpoints, such as fixed-term survival (e.g., 2-year survival: yes/no) or disease presence (yes/no), are discrete (binary or multilevel factors) and are used for patient classification into risk groups. The nature of the endpoint determines the type of statistical analysis that can be carried out for risk characterization.

Risk Characterization for Continuous Endpoints

For continuous endpoints, cohort information is longitudinal—it contains a record of duration along a timeline until the event of interest occurs or the patient becomes lost to follow-up (e.g., for OS or PFS). In this context, two complementary forms of statistical assessment are traditionally carried out, which evaluate (i) the predictive potential of risk variables of interest and (ii) the ability of predictive models based on these variables to segregate patients into risk groups. Hazard models (most often, Cox proportional hazard models [6]) are usually used for the former task; they provide a measure of the effect of an increase in a particular variable on patient risk (as defined by the endpoint considered). For the latter task, survival curve estimates (most often, Kaplan-Meier estimates [6]) are obtained and compared for subgroups of the cohort defined on the basis of a score determined from the risk variables.

Univariate hazard models focus on the individual effect of a variable on risk. For a given outcome type, they produce a hazard ratio (HR) that quantifies the change in risk for a unit increase of this variable, compared to a hypothet-ical baseline population within which the variable has a value of zero. A HR equal to 1 indicates no effect; HRs less than or greater than 1, respectively, indicate reduction and increase in risk following a change in the variable. Table 6.1 below illustrates an example where the risk variable x is *tumor grade*, taking values for "low," "intermediate," and "high" tumor grade. Here x = "low" is used as the baseline reference profile (a low-grade patient tumor). The HR measures the relative difference in risk between this reference and an intermediate-grade or high-grade patient tumor (with x = "intermediate" and "high," respectively). The case x = "high" yields a HR of 6.22, indicating a (very) large increase in risk for poor outcome compared to the reference case. The variable may also be continuous; for example, if x is (normalized) SUV_{max}, then 0 could be used as the reference value, and the hazard ratio measures the relative change in risk associated with a unit increase in SUV_{max} from a value of 0. In multivariate hazard models, a number of variables are considered together to describe the endpoint of interest. In this context the hazard ratio of a variable quantifies the change in risk for a unit increase of this variable when all other variables are held constant in the model.

The statistical significance of the hazard ratio associated with a risk variable (i.e., of the effect of the risk variable) is often the primary imaging biomarker performance indicator. A statistical test (usually either the Wald or Likelihood Ratio test) is carried out to determine whether the effect of the variable (in other words, the proportional change in risk incurred by a change in this variable) is significantly different from zero. A statistically significant effect indicates that the variable aptly describes some of the risks defined by the endpoint. Statistical significance of a variable is usually determined by its *p*-value. When the *p*-value falls below a set significant threshold (usually set at either 1% or 5%), then the variable is deemed a statistically significant risk factor. Although this result is derived from the clinical cohort under study, interpretation applies to the whole population of patients that this cohort represents. In the example of Table 6.1, the *p*-value for variable x = *'high'* is $p = 0.0004$, i.e., $p < 0.01$,

Table 6.1 Left: Cox proportional hazards analysis of tumor grade (low, intermediate, high) effect on overall patient survival duration. The hazard ratios (HR) indicate that the risk of death increases, respectively, by a factor of 2.76 (i.e. a 176% increase) and 6.22 (i.e. a 522% increase) respectively for intermediate- and high-grade tumors, compared to a reference low-grade profile. The p-value smaller than 1% (0.01) for the "high-grade" level confirms that the change in risk in a sarcoma population is statistically significant for high-grade patients, compared to low-grade patients. Right: logistic regression analysis of the effect on the same variable on 2-year survival status (i.e., whether death occurs within 2 years of diagnosis) provides similar assessment. The intercept parameter corresponds to the reference low-grade group, with overall lower risk of death than the other two groups, as indicated by the increase in variable effect across groups. Corresponding p-values indicate that high-grade tumors are associated with a significantly higher risk of death at the 2-year horizon for patients with sarcoma. The corresponding effect (or odds ratio) indicates this increase in risk of death is estimated at 9.39 times the risk for a low-grade patient. Dataset acquired at the University of Washington School of Medicine (August 1993 to January 2003, Dr. Janet F. Eary), after biopsy (202 patients, 91 events; 88 females; 114 males; 52 bone, 17 cartilage, 132 soft-tissue; 32 low-, 69 intermediate-, 101 high-grade)

Cox proportional hazard model (overall survival, continuous endpoint)			Logistic regression model (2-year survival, discrete endpoint)		
Grade level	HR	p-value	Grade level	Effect	p-value
–	–	–	Intercept	0.07	**0.0003**
Intermediate	2.76	0.0593	Intermediate	3.49	0.1154
High	6.22	**0.0004**	High	9.39	**0.0031**

and thus x is deemed a significant risk factor for overall survival of sarcoma patients at the 1% significance level (or "with 99% confidence"), that may be used to classify sarcoma patients into different risk subgroups. From this result we can also infer that any high-grade sarcoma patient has a significantly higher risk of death than a low-grade tumor, whether these patients belong to this cohort or any other.

Statistical reports aiming to validate pseudo-markers or other predictive variables should provide both the p-value and associated variable effect or HR provided by the model, in order to provide both evidence of the statistical significance and an evaluation of the magnitude of the effect of the variable. Confidence intervals for the variable effect or HR are also strongly desired to indicate variability of the statistical assessment.

Multivariate predictive models can also be used for risk stratification for outcome type, based on a single-score summary obtained from them. Risk groups are determined by comparing this score to threshold values, which are set by the analyst according to the context of the cohort, either arbitrarily or based on clinical experience. Survival curve estimates are used to evaluate this stratification, most often by means of a Kaplan-Meier analysis (here, "survival" is used loosely and describes duration-to-event, whatever this endpoint event may be) [6], such as that depicted in Fig. 6.1.

For each risk group, a curve is obtained that starts at 1 and monotonically decreases toward 0 as events occur within that group. One essential advantage of these analyses is that they allow subject censoring, i.e., loss-to-follow-up, to be taken into account. A statistical test (usually the log-rank test) is then used to evaluate the statistical significance of the separation observed between different survival curves. A very small test p-value (at the 1% significance level, any $p < 0.01$) indicates that the survival curves are statistically significantly well separated, i.e., that there are significantly different survival outcomes on average for at least some of the corresponding sub-cohorts.

Risk Characterization for Discrete Endpoints

For discrete endpoints, the problem of interest becomes a classification problem. Duration data for a patient cohort may be available but the goal is to assign patients to different groups, according to risk level, disease stage, disease subtype, etc., on the basis of their information. In this context, statistical models such as logistic regression, decision trees, or more elaborate forms (random forests, support vector machines, neural networks) are used to determine a classification rule based on the available data. The model allows summarizing a patient's multivariate profile into one single-valued score, which is compared
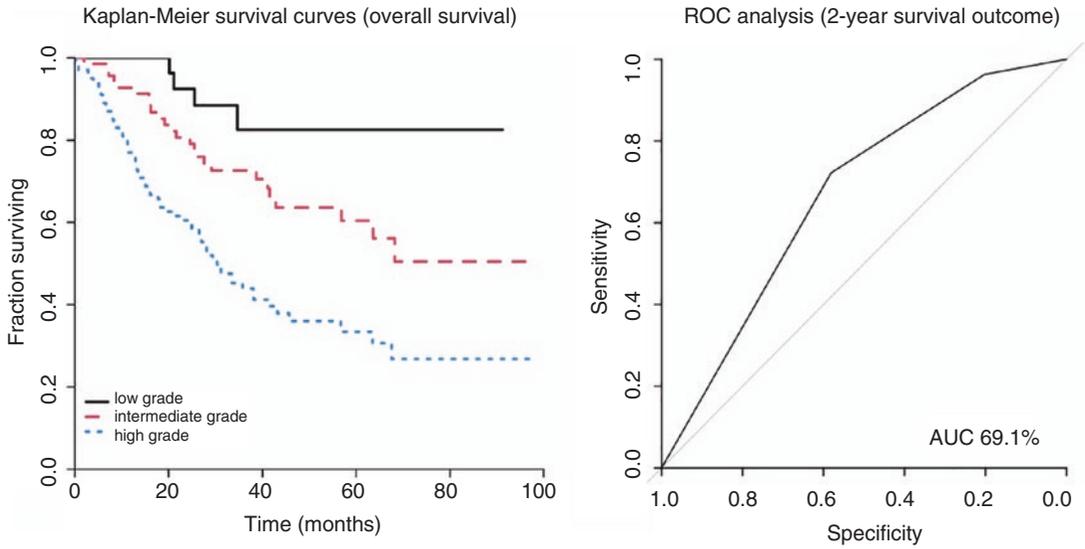
Fig. 6.1 Overall survival risk analysis based on variable grade used in Table 6.1 for the same sarcoma cohort. Left: Kaplan-Meier survival curve estimates as a function of survival duration (a continuous endpoint) obtained for each grade group clearly indicate higher risk (lower chance of survival) as grade increases. The log-rank test p-value $p = 0.000004$ indicates significant difference between the survival curves for each tumor-grade group. Right: ROC analysis of patient survival status (0: dead, 1: alive; a discrete endpoint) predicted by a univariate logistic regression model using the same grade variable. The faint diagonal line indicates the prediction performance obtained by flipping a fair coin. The thick curve indicates the relative gain in prediction accuracy (both in sensitivity and specificity) obtained using the univariate model to predict patient outcome. The associated AUC of 69.1% indicates mediocre overall performance from this model

against a set threshold value in order to classify the case [6, 7]. Just like in a Cox model (which itself relies on a form of linear regression), a classification model yields an assessment of statistical significance for each of its variables. This significance is again evaluated in terms of a p-value compared to a predefined significant threshold, as demonstrated in Table 6.1.

Predictive performance of the classification rule is then evaluated in terms of standard detection theory tools. For binary classification (by far the most common scenario), sensitivity and specificity of the classifier are measured, with values within the [0,1] interval, i.e., between 0% and 100%. Any classification model achieves a trade-off between these two performance metrics and tends to be stronger in one of these two aspects. One controls this trade-off by tuning the above-mentioned classification threshold. Performance of the classifier is commonly assessed by means of an ROC analysis (receiver operating characteristics), which provides a plot of sensitivity (or true positive rate) against 1-specificity (or false-positive rate) for varying classification thresholds, as illustrated in Fig. 6.1. The area under the ROC curve (AUC) yields a single-number summary of the classifier's predictive capacity, but many other related metrics are available for refined performance evaluation. A common rule of thumb is that AUCs between 60–69%, 70–79%, 80–89%, and 90–100%, respectively, indicate poor, fair, good, and excellent model fits, but this is not prescriptive. An AUC of 0.5 (i.e., 50%) corresponds to a decision based on flipping a fair coin [6–8].

Influential Factors in PET-Based Quantitation

Nature of PET Data and Impact of Imaging Protocol

A number of operational parameters such as scanning time, injected dose, patient position, scanner calibration, and other elements of

imaging protocols can vary considerably between clinical sites. Some of these aspects directly impact the data quality of the output image. Injected dose is a major contributor to image contrast with some imaging agents resulting in differences in the signal-to-noise ratio in the image. Often with a lower imaging agent dose, there is greater image signal noise. Use and choice of attenuation correction techniques also affect the statistical nature of the final PET imaging data, with significant differences between PET-MR and PET-CT [2, 6, 9]. PET scanners yield different spatial and timing resolutions, depending on manufacturers and especially across generations of tomographs. This can particularly affect the visibility, detection, and characterization of small uptake foci. Slice thickness also has an impact on apparent image noise [10].

PET imaging data consists of reconstructions of photon pairs registered in the scanner detectors. As a result, scanner architectures also have an impact on the image output. PET imaging systems yield lower transverse spatial resolution than MRI and CT systems, due to the nature of the detection mode and information density of each modality. Software treatment of the raw data, either in-built or applied post-acquisition for all these systems, also has an impact. The raw acquired system input are positive integer values, but due to the tomography process, their spatial distribution must be reconstructed in a 3D spatial coordinate domain to produce a final image. During that step, the imaging input events become continuous values as opposed to integers, due to the application of image reconstruction algorithms. Choice and calibration of these algorithms has an impact on the statistical characteristics of the reconstructed image output data. As an example, previous reconstruction methods based on filtered back projection (FBP) produce images where (mostly background) voxels may have negative intensity values, but lower overall bias, whereas more current methods relying on ordered-subset expectation maximization (OSEM) ensure positive uptake values in all areas of the image, but result in larger reconstruc-

tion bias [11]. Image reconstruction may be carried out iteratively over acquired 2D slices or overall across the imaged 3D volume.

Segmentation and Post-Processing

We mentioned earlier that the choice of segmentation algorithm can drastically impact image quantitation. For example, in a case where peak tissue VOI uptake is found at the tumor boundary, improper volume delineation may remove some of the voxels contributing to the 1 cc sphere placed around the VOI designated for determination of tumor SUV_{max}. This would result in an error in SUV_{peak} determination. Tumor of other tissue uptake focus volume assessment and other parameters are also very sensitive to image segmentation approaches.

Image interpolation procedures are also known to have a potentially great influence on analysis output variables [3]. Image interpolation is carried out when the resolution of the image needs to be adapted in order to apply image-based data analysis types such as radiomics, described further in Sect. 3. In some of these schemes, approximate voxel intensities are calculated in tissue locations with a finer or coarser image array size. In a simple example consisting of a function for halving slice thickness, voxel intensities in each slice k in the interpolated image would be predicted on the basis of its neighboring slices $k - 1$ and $k + 1$, e.g., by averaging adjacent voxels.

PET spatial resolution induces another limitation on quantitative image analysis, due to the partial volume effect (PVE) for some tissue uptake foci in various body locations. Large voxel sizes imply that the information contained within a voxel partially captures the activity of its surrounding voxels [12–15]. PVE correction techniques are oftentimes used in order to correct this spillover effect and ensure that the voxel intensity reflects the activity in the tissue at that location only. This correction takes scanner specifications (its point-spread function, PSF) and voxel resolution into account. VOI interpolation should generally also be followed by PVE correction.

Quantitation with Short-Half-Life Tracers Radiotracers

Important considerations in PET imaging are radiotracer physical and biological half-lives. Using radiotracers with short uptake and clearance times often reduces the overall duration of the imaging procedure, resulting in less radioisotope decay which increases the number of photon-induced count input from each imaging plane. This factor, in turn, can give a positive impact on image quality. Some imaging agents can also result in enhanced image contrast by yielding low tracer uptake in background tissue [16]. Several studies with short half-life tracers have been reported. In melanoma, amino acid imaging agents for tissue protein synthesis radiolabeled with carbon ^{11}C (20.4-minute half-life) have been evaluated for imaging prediction of therapeutic response. ^{11}C-alpha-methyltryptophan (^{11}C-AMT) was evaluated in NIH clinical trials in metastatic melanoma, and ^{11}C-methionine (^{11}C-MET) was studied in patients with mucosal malignant melanoma [17]. Gallium (^{68}Ga) has a 67.7-minute half-life, and imaging agents that use this radiolabel have been considered for study in several cancers including sarcoma [16]. Guidance on image quantitative analysis for these imaging agents is not yet clear, since the statistical characteristics of the PET imaging data output will strongly depend upon imaging agent characteristics, including administered dose.

Quantifying Tumor Heterogeneity

Motivation for PET-Based Heterogeneity Assessment

The level of biologic heterogeneity in subpopulations of cancer cells within a tumor has been identified as a key driver of cancer patient outcome, in many forms of cancer including sarcoma and melanoma [6, 14, 18–20]. Intratumoral biologic heterogeneity is most commonly assessed by tissue biopsy, in a process that requires that the optimal sampling area be identi-

fied, and sampled for histopathologic assessment. This is a challenging task, and there is a significant risk that the histopathologic assessment may not represent the underlying phenotypic or genetic landscape of the disease fully or adequately [21].

The promising potential of imaging biological assessment for personalized cancer care motivates the use of noninvasive imaging, and in particular PET, to replace biopsy-based histopathology assessments with a more convenient, faster, reliable, and accurate evaluation process. This approach relies on the assumption that tumor biological heterogeneity drives imaging agent uptake heterogeneity. Results from current work can be interpreted to suggest that macroscopic uptake heterogeneity observable on PET images is likely to reflect the microscopic heterogeneity of cancerous tissue. The prognostic potential of PET tumor uptake heterogeneity has been established for several cancer types, predominantly using FDG-PET imaging [13, 19, 22, 23]. Interpretation of PET-derived heterogeneity quantitation must be made in the context of the PET tracer used [6, 14]. Interpretation of the uptake variations at macroscopic level in the PET tracer uptake pattern must be made in terms of the specific agent used for imaging different tumor biologic parameters such as glycolysis, hypoxia, or cell proliferation. Specific PET agents for imaging these parameters are fluorodeoxyglucose (FDG), fluoromisonidazole (FMISO), and fluorothymidine (FLT).

Heterogeneity Characterization in Sarcoma

The main methodological approaches to image-based evaluation of intratumoral heterogeneity in radiotracer uptake distribution in the literature can be organized into two generic groups [6]. One strategy that has received consideration relies on the analysis of image texture, using well-established tools from the (non-radiological) image processing community. Texture analysis produces a large number of image-derived variables, or features, that capture different aspects of the image content, summarizing variations in

image intensity and relief, among other things. Following a trend that started about 15 years ago, a growing number of studies characterize intratumoral radiotracer uptake distribution heterogeneity in terms of combinations of a number of these texture features. Novel multivariate predictive models, or nomograms, for staging and prognosis can be built by incorporating such combinations with existing clinical variables used routinely, such as patient age and tumor SUV_{max} (in FDG-PET imaging) [14]. However, selecting an adequate number and subset of texture variables is a considerable task. This prompts advanced model selection procedures relying on modern statistical learning techniques, and for this reason the underlying concept and use of texture analysis for image heterogeneity assessment is further described in Sect. 3.

The second methodology consists in comparing the observed tracer uptake distribution to reference shapes or patterns. Motivated by clinical experience with sarcoma, intratumoral radiotracer uptake spatial heterogeneity has been measured conceptually as a degree of conformity of the spatial PET tracer uptake distribution with an idealized ellipsoidal pattern. Less conformity of the tumor spatial uptake pattern with the idealized ellipsoidal object implies more heterogeneity in tumor radiotracer spatial uptake [6, 19]. The 3D spatial map of tracer uptake in the tumor VOI can be described using a mathematical model that represents each voxel in terms of its radial position within the idealized ellipsoidal pattern:

$$SUV_{ideal} \approx g\left(\text{voxel radial position}\right)$$

where g is a function of voxel location that represents the uptake profile signature going from the VOI core out toward its boundary. Figure 6.2 illustrates two examples of sarcoma studies acquired at the University of Washington with a GE Advance PET, presenting different levels of tumor FDG spatial uptake heterogeneity, and shows the corresponding uptake profile g obtained following this approach, where the uptake is shown on the y-axis as a function of the radial voxel position on the x-axis. Clinical experience indicates that a tumor mass with an FDG-avid (viable tissue) central region is usually observed in a histologically low-grade tumor, whereas a tumor that has a photopenic central region that likely represents a necrotic zone is a histologically high-grade tumor. This image analysis spatial model was designed with the purpose of capturing these tumor features and used to derive a measure of conformity to the "idealized" pattern of a homogeneous uptake mass, thus defining FDG uptake spatial heterogeneity in terms of the scaled distance between idealized and observed SUV data:

$$\text{Het} \approx \left(SUV_{obs} - SUV_{ideal}\right)^2 / \text{Variance}\left(SUV_{obs}\right)$$

Based on the above definition, a low value of Het is assigned to studies for which the difference $SUV_{obs} - SUV_{ideal}$ is small. In the examples of Fig. 6.2, this difference corresponds to the average distance of the dots (the observed uptake data, SUV_{obs}) to the model line (the idealized SUV, SUV_{ideal}).

Baseline prognostic potential of this FDG spatial uptake heterogeneity assessment was validated for sarcoma patient overall survival [6, 19]. This method can be applied to PET/CT and PET/MR data. Variations using a tubular reference structure were also considered, to provide a more flexible structure for characterization [24].

Further Characterization of Spatial Tumor Uptake Patterns

Statistical modeling of intratumoral uptake creates opportunities for a detailed quantitative characterization of tumor imaging agent uptake parameters beyond standard heterogeneity evaluation. In particular, local spatial variations in uptake may also be assessed from modeling of the spatial uptake distribution, to evaluate differences in uptake at any location within the VOI. Mathematically, this difference can be defined as the derivative (or gradient) of the uptake profile function g defined above, and a gradient value can be derived for any voxel:

$$\text{Gradient} \approx -g'\left(\text{voxel radial position}\right)$$

This measure quantifies the steepness of the profile uptake curve of Fig. 6.2, along that radial
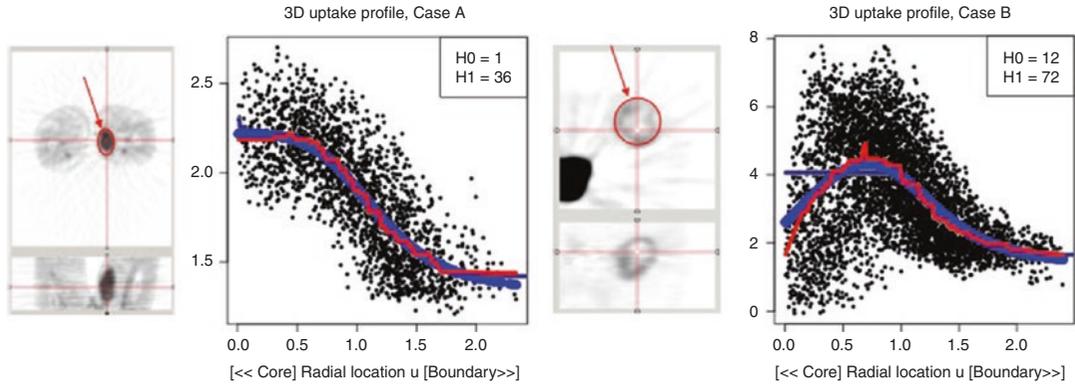
Fig. 6.2 Example of sarcoma studies from the cohort of Table 6.1, with uptake profile curve *g* and corresponding heterogeneity quantitation indicated in inset (figure adapted from [6]). Input VOIs are segmented as an ellipsoidal volume such as those outlined in red on the transverse views (top). Case A (left): 49-year-old male with upper thigh soft tissue sarcoma with an active, homogeneous core represented by a gradually decreasing uptake profile pattern. Case B (right): 48-year-old male with pelvis soft tissue sarcoma, with a heterogeneous core with low activity captured by a modal uptake profile pattern

position within the VOI. Following the mathematical definition of Wolsztynski et al. [6], a negative gradient value indicates a locally decreasing uptake profile when applied to an FDG-PET image data set (relative to peak activity and when exploring the VOI from the tumor center toward the tumor boundary). A positive gradient value corresponds to an area of increasing uptake signature. Following this representation, higher-grade sarcoma will typically exhibit increased avidity further away from its center, and tumor central uptake would be seen to decrease, resulting in local negative gradients in that area. Whether the gradient is positive or negative, its absolute value indicates the magnitude of the trend of uptake change occurring along a given tumor radius and for all radii between the tumor center and its boundary.

Since this spatial uptake gradient quantitation evaluates average radial differences in uptake throughout the VOI, one can map areas of uptake differences within the tumor, including information on the direction of change. This may be useful, for example, to locate and visualize areas with more significant local variations in uptake. The sample of gradients derived from the image can also be summarized into a single value (e.g., its median value, 95th percentile, or maximum value), to be used in multivariate prognostic

models alongside other variables. The prognostic utility of spatial uptake pattern gradient summaries in sarcoma has been demonstrated by Wolsztynski et al. [6].

Tumor Uptake Spatial Heterogeneity Characterization in Melanoma

Early-stage melanoma is associated with favorable outcome following complete resection, but challenges are much greater for effective treatment of advanced disease, including elevated risk of tumor in stage IV. Stage IV melanoma tends to present with a high level of both intratumoral and intertumoral heterogeneity in imaging agent uptake [25, 26], which can increase the mathematical uncertainty associated with baseline tumor and risk characterization. Assessment of these two forms of biologic heterogeneity would be helpful in effective treatment personalization.

Currently, biological and mutation-driven cancer therapies are treatments of choice for a number of malignancies, which explains a predominance of studies focused on tumor genomic and immune heterogeneity assessment, and assessment of metastases. To date, it is however unclear how markers of genetic heterogeneity are captured by PET imaging, and radiological

intratumoral heterogeneity assessment in melanoma is not as developed as in sarcoma. Most image quantitation is currently based on FDG PET SUV-type summaries, most often in terms of change in SUV_{max} or SUV_{peak}, SUV_{peak} interquartile range, or sum of SUV summary values measured at the predominant lesion sites [27, 28]. PET image spatial uptake heterogeneity is usually defined as a measure of variability in imaging agent uptake profiles, but with limited spatial relevance.

Artificial Intelligence and Radiomics in PET

The Role of Artificial Intelligence in Radiology

Motivation and Current Uses of Artificial Intelligence in Radiology

In order to improve patient care processes and develop patient-adaptive therapeutic pathways, modern healthcare has turned to artificial intelligence (AI). The term designates a "universal field" aiming to "build intelligent entities" [7]. It consists of integrated processes that have the ability to learn from data and take actions or formulate decisions adequately with respect to the input provided. AI solutions tend to be computationally more effective than conventional human- or computer-based techniques, allowing for efficient processing of large volumes of data at high speeds. They are built to learn continuously and aim to outperform human experts for specific tasks. Well-known universal examples include voice-powered personal assistants available in today's smartphones, driverless robotic vehicles, and the IBM DeepBlue chess-playing computer [7]. In medical environments, AI is considered for a wide range of purposes, such as the analysis of ECG data [29], computer-aided diagnosis (CAD) [30], or management of electronic health records and medical notes available in unstructured formats [31]. One of the most powerful examples of AI in healthcare is the IBM Watson system for diagnosis and adaptive treatment decision-making. However most developments in this field are currently still at the experimentation stage and face limitations in terms of regulations, logistics, and validation [31].

In radiology, current research underway is predominantly focused on AI for tumor screening, characterization, and diagnosis purposes. Applications of AI to radiological data remain experimental or exploratory, although support decision systems were reported in the 1970s [32]. Recent developments include CAD for breast cancer screening [8, 33], aiming to reduce the high false-positive rates of existing exam approaches. Screening processes perform classification of the VOI (for instance, into binary decisions such as disease presence/absence) on the basis of a number of input variables, some from routine clinical investigation, and others extracted from the radiological image. Results to date suggest that current AI practice provides improvement when used as a support tool (i.e., as a second opinion) rather than a standalone solution in this domain and needs further validation.

Image-based tumor segmentation can also benefit from AI technologies, for both tumor delineation and intratumoral parameters profiling. Current advances for these tasks rely on automatic learning techniques (which AI systems heavily rely upon) rather than implement full AI processes. Namely, "machine learning" and "deep learning" frameworks allow building of predictive models based on training data. Advanced automated processes are considered in addressing interuser variability in image segmentation and to speed up this time-consuming task. A broad range of strategies for automatic tumor delineation in images have been explored, some relying heavily on classical image processing tools such as wavelet decomposition. Others involve an elaborate panel of tumor features calculated from the VOI. Researchers report using deep learning techniques for VOI segmentation especially for MRI and CT data [34–36]. Volume delineation based on neural networks could also be applied effectively to PET data [37]. VOI segmentation remains conservative due to its critical impact on subsequent analysis. Implementation of complete AI solutions within commercialized or serialized products (either imaging devices or

software) will require considerably more development. Recent efforts have evaluated tumor image segmentation for tumor prognosis potential. This is reflected in recent major community challenges such as the Brain Tumor Segmentation Challenge (BRaTS) for glioma segmentation and prognostic assessment using AI-based techniques on multicentric preoperative MRI data [38].

Tumor characterization and prognosis is the area in which the field of radiology has enjoyed the most progress with the emergence and generalization of AI. In routine practice, translating and interpreting the wealth of information captured by 3D imaging data into precise and reliable clinical assessment and prognosis is challenging. Based on visual or semiquantitative assessment, a radiologist evaluates disease presence/absence, size, and stage and forms an expert opinion on the level of tissue heterogeneity within a tumor, areas of specific interest such as target regions for biopsy, or therapeutic response [1, 39, 40]. For PET imaging data in particular, quantitative assessment as described in Sect. 2 above is often restricted to simple SUV summaries (SUV_{max}, TLG, etc.) that do not usually include heterogeneity valuation. The imaging data, which comprises of thousands of voxels in many cases, provides an opportunity for a much finer characterization of the disease but remains underexploited in clinical settings [1]. This calls

for more elaborate image analysis approaches, which can in turn enhance prognostic modeling thanks to a more informed description of the disease process [6]. The accumulation of significant experimental results in this area of investigation has led to the rapidly emerging field of "radiomics," whose principle, described in the next section, is not unlike that of genomics [2, 14, 22, 41].

Radiomics: Machine Learning for Tumor Characterization and Prognosis

A growing number of image-derived variables have been considered in recent research to either replace or complement routine quantitative tumor assessment. Particular emphasis has been placed on capturing various aspects of intratumoral spatial uptake heterogeneity noninvasively, using a subset of dozens, sometimes hundreds, of descriptors of voxel intensity distribution, image texture, and related image processing aspects. Such a collection of metrics are often referred to as "textural parameters" or "radiomic features" and may be divided into groups of image descriptors; Table 6.2 below illustrates some of these groups with examples of well-known features for each [3]. This description is non-exhaustive. A wide range of other metrics is available, e.g., textural descriptors derived from wavelet transforms

Table 6.2 Some of the more common families of radiomic features and examples of features in each family. Figure 6.3 further below illustrates the nature of the representations of VOI uptake distributions used to define feature families (A)–(D)

(A) Morphological features	(B) Intensity features	(C) Spatial textural features	(D) Regional features
Shape and structure descriptors of the volume of interest	*First-order aspects of the distribution of voxel intensity levels (histogram-based, not spatially relevant)*	*Second-order aspects of the spatial distribution of voxel intensity levels (associations of neighboring voxels)*	*Higher-order aspects of spatial distribution of connex groups of voxels (number and size of groups, etc.)*
Examples of features (variables)			
Volume	Mean intensity	Joint average	Long-run emphasis
Surface area	Intensity variance	Joint variance	Run length variance
Sphericity	Maximum intensity	Joint entropy	Small/large zone emphasis
Major axis length	Intensity range	Correlation	Gray-level non-uniformity
Elongation	Intensity uniformity	Homogeneity	...
Flatness	Histogram entropy	Contrast	
...	...	...	

of the image. The variables of Sect. 2 measuring heterogeneity and radial variation in uptake are other candidates.

Figure 6.3 illustrates different summaries of the distribution of gray levels in the VOI in an FDG-PET sarcoma image, aligned with the columns of Table 6.2. The group of first-order metrics (second column in Table 6.2) is not spatially relevant, i.e., these metrics would remain unchanged if the image were scrambled. They are very close in nature to quantitative summaries used routinely

such as SUV_{max} or SUV_{mean}. Second-order features summarize how any two voxel intensity levels i and j may be distributed within the image—for example, if it is likely to find a very bright voxel next to a very dark one (image contrast). Higher-order features describe other aspects of the structure of intensity levels within the image after its requantization into a lower number of gray levels (usually 32 or 64), such as the size and number of groups of voxels of comparable intensities, or their homogeneity within the volume.
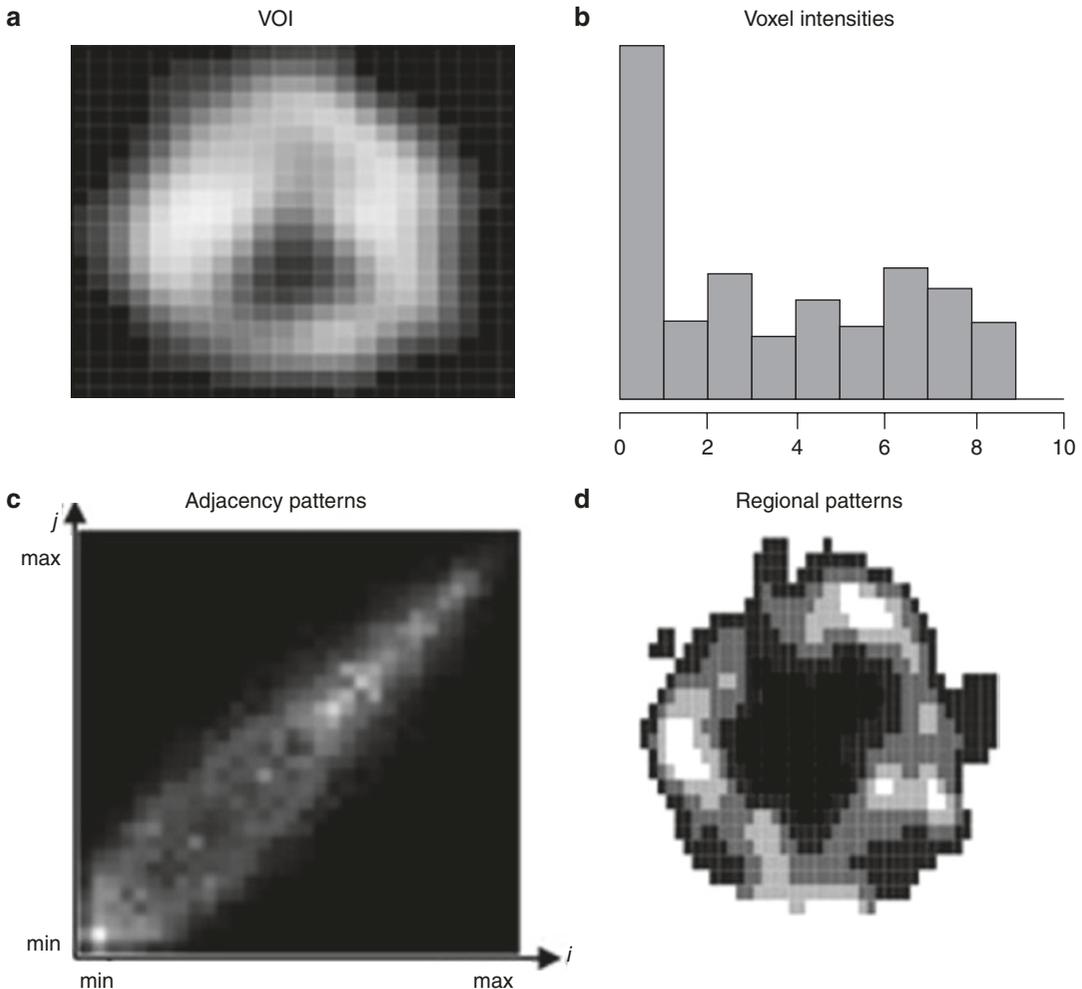
Fig. 6.3 Various distribution summaries of the voxel intensity levels found in a soft-tissue sarcoma volume of interest, of which a transverse slice is shown in (**a**) first-order probabilities of intensity levels (**b**); second-order probabilities of co-occurrences of gray level intensity pairs (i,j) (dark and bright points on this matrix, respectively, indicate low and high probabilities of finding vox-

els of intensities i and j next to each other within the volume of interest) (**c**); regional patterns in gray level intensities (**d**), obtained following image requantization into a set number of gray level bins (typically into 32 or 64 gray levels). Note this requantization is also used before computing second-order texture features (**c**), and for some of the first-order intensity features (**b**)

A very informative tumor profile can be derived from clinical assessment and patient imaging that includes clinical biomarkers, radiological markers, and other potential risk factors. Radiomics analyses rely on algorithms that have the ability to discover patterns and relationships among this wealth of clinical variables. These pattern-finding techniques belong to a sub-branch of AI called "machine learning" (ML), which greatly overlaps with the statistical learning theory framework—in fact, many data scientists and other analysts who are not involved or interested in AI use ML techniques on a routine basis. These are used to automatically build mathematical (e.g., prognostic or diagnostic) models from sample data, can be used to process very large amounts of information, and require only very mild assumptions and prior knowledge of the data. Various ML techniques can be considered, the most popular of which are named random forests, support vector machines, and artificial neural networks (ANN). The latter attempt to mimic a human brain, and allow for nonlinear interactions between variables, provides an opportunity for deeper exploration of existing patterns in the data. As a result ANN have yielded their own paradigm, termed "deep learning," within the ML framework.

The overall objective of a typical radiomics analytical pipeline is to create, from a dataset of measurable clinical information, a model that will allow predicting a target variable for future patients (patient outcome, disease type, etc., depending on the objective). For tumor characterization this target variable or label could be, e.g., tumor subtype, stage, grade, or biopsy-assessed biologic heterogeneity. For a prognostic model, the goal is usually a clinical endpoint, patient outcome at last follow-up, or duration of patient survival until last follow-up [2, 14, 22, 41].

The following key methodological aspects involved in radiomics analyses are driven by statistical considerations:

(i) Preliminary feature elimination is usually performed in order to "sieve through" the hundreds of input variables, or features, to eliminate redundant information and reduce the size of the dataset to be analyzed. This step is taken to increase pertinence and performance potential of the predictive model being built. It may be performed using a number of statistical techniques such as correlation analysis, multivariate factor analysis, or using an iterative elimination process [42–44].

(ii) Then the predictive model is "trained" on, or fit to the data, so that it learns to use the features together to characterize the target variable. This phase is usually a form of supervised learning, which means that the set of clinical and radiological features is explored retrospectively against the observed value of the target variable for previous clinical patients. It is critical that some form of statistical methodology such as cross-validation is used at this stage in order to avoid model overfitting, a case where the model would describe the known data too well but perform more poorly on independent data [42, 43].

(iii) Ideally, predictive performance of the trained model should be evaluated on independent data, to avoid any statistical bias in this assessment [44]. If this were not the case, predictive assessment of the trained model would likely be overinflated, since the latter would be used to predict values that it has already somehow "seen" during the training process. Because of the requirement to use a sound statistical validation framework, it may be unfeasible to use an independent validation sample when the clinical dataset available for analysis is relatively small (e.g., less than, say, 100 patients). In that case, significant predictive performance may still be indicative of clinical potential but should be validated with follow-on studies on larger sample sizes.

(iv) Final choice of an appropriate predictive model (e.g., a random forest, neural network, etc.) may be made via benchmarking but could also take practical aspects such as clinical interpretation into account. Conventional statistical models used extensively in medical and clinical studies, such

Fig. 6.4 Overall analysis pipeline for the construction of a predictive model from retrospective clinical data

as logistic or Cox proportional hazards models, could also be considered for such applications against more elaborate machine learning techniques, if they were obtained via sound statistical training as described above. Alternatives such as random forests or neural networks are however more appropriate when the data comprises of more variables than patients.

Figure 6.4 illustrates the arrangement of the first three aspects in a typical analytical pipeline. Phases (i)–(iii) of this process can be replicated for different models, and the best model may be selected on the basis of performance indicators that are adequate to the desired application.

Radiomics in Melanoma Imaging

The literature describes uses of radiomics for melanoma mainly for screening and to a smaller extent for classification or prognostic models. The majority of reports explore the use of AI in melanoma characterization from either dermoscopic images or digital photographs [45–49]. Study results were reported over 20 years ago [50]. A consensus is forming that AI methodologies applied to dermoscopic imaging can improve skin screening performance for melanoma.

Although PET/CT has become routine practice for staging and management of malignant melanoma, radiomics analyses of PET imaging data in melanoma have yet to be reported. Some studies have only recently considered radiomics for immunotherapy response assessment in melanoma, using CT data [51]. Currently, radiomic characterization in melanoma primarily targets the prognostic drivers in advanced forms of melanoma. As a result, recent works focused on the

analyses of metastases as opposed to the primary tumor, often using CT or MR imaging data, focusing in particular on regional lymph nodes and brain metastases [52, 53].

Radiomics in Sarcoma Imaging

Only a few studies on radiomics in human sarcoma have been reported to date. Vallières et al. [54] used radiomics to improve prediction of lung metastasis in soft-tissue sarcoma (STS) of the extremities based on joint FDG-PET and MR imaging data. MRI data were also used by Corino et al. [55] for grading of STS tumors. Survival prediction based on a radiomics analysis of baseline CT imaging of high-grade osteosarcoma was also considered by Wu et al. [56]. The only report to date of a radiomics analysis for sarcoma using exclusively FDG-PET data was contributed by Wolsztynski et al. [6], where various ML models were considered for overall survival prediction. The authors demonstrated that prognostic models could be found that yielded effective risk prediction by combining routine clinical information, tumor spatial uptake heterogeneity characterization, and radiomic features.

These scarce explorative works on relatively small clinical cohorts need to be replicated and validated at larger scales. They however have the merit to illustrate the potential of radiomics and machine learning-based methodologies for various aspects of human sarcoma risk characterization, improving upon traditional clinical risk assessment practices. They also highlight how clinical interpretation of radiomic models is more challenging than that of tumor characterization methodologies based on clinical experience. Some recent works [6] indicate that structural modeling of the PET tracer uptake distribution

within sarcoma tumors allows capturing some aspects of the PET imaging data that are not exploited by traditional radiomic analyses. Extending the spectrum of features to include these clinically relevant metrics should improve both our clinical understanding of radiomic variables and our interpretation of multivariate predictive models formed on their basis.

Limitations and Future Opportunities

Fully operative AI solutions in radiology are currently not ready for routine implementation. Current research and development efforts mainly make use of elaborate analytical methods such as ML to explore novel risk variables and pseudo-biomarkers and advance academic and clinical understanding of the vastly unexploited potential information contained in radiological imaging data. These efforts however aim at the long-term objective of integrating AI within routine practice.

Limitations of Machine Learning Methodologies in Radiology

Among the main limitations and current challenges faced by AI developers, we can list the following aspects:

- Need for upskilling: AI has matured over decades to yield a broad range of concepts and methodologies. This may put the onus on radiologists to grasp a new paradigm. Data scientists should provide support and expertise, but ideally modern radiologists should acquire a working knowledge of this particular field and what the application to radiology requires.
- Reliance: The assistance provided by AI may result in reduced vigilance and undue trust or reliance on this convenience [29].
- Interpretability: Radiomic tumor analyses usually offer limited clinical interpretability, mainly because of the involvement of textural features that do not have a direct biological or structural meaning. In the same spirit, AI-derived decision-making processes are usually not easily translatable into clinically

relevant reasoning. This is without doubt one of the major challenges facing clinical implementation of comprehensive AI decision-making solutions.

- Unbalanced performance: By construction, classification methods operate a trade-off between sensitivity (yielding very few false negatives) and specificity (yielding very few false-positive results). Current AI-based technologies for computer-aided detection (CAD) or automatic tumor segmentation tend to perform well in one of these aspects, to the detriment of the other [46]. Clinical performance is also biased toward particular medical conditions, since by construction a machine learning algorithm is usually only trained effectively to detect or classify one specific characteristic.
- Additional burden: Despite the aim to use AI screening and diagnostic solutions as a second expert opinion to the radiologist's reading of the imaging data, output from current technologies must be cross-examined by the expert, which generates additional workload. Similarly, radiomics analyses for tumor characterization require expert interpretation and clinical confirmation.
- Liability and regulation: Future clinical environments will need to clarify the distribution of legal responsibilities among radiologists, manufacturers, and other AI-related skills providers when AI technologies are involved in medical errors [7]. These technologies will also require approval and regulation by relevant bodies, such as the Food and Drug Administration for the USA, for example.

Opportunities Offered by AI

Despite its inherent challenges, AI provides notable opportunities in all areas of cancer care, including through the following pivotal radiology-based aspects:

- More effective early detection: research has already demonstrated that AI can benefit early detection of cancer from radiological data, which would in turn improve survival rates [57].

- More effective baseline prognostic models, incorporating more insightful tumor descriptors, can enhance characterization and prognostic accuracy [2, 6].
- Patient-adaptive therapy can be facilitated by enabling individualized assessment of therapeutic effectiveness [2].
- Cumbersome workloads and fatigue pertaining to repetitive routine tasks for radiologists can be alleviated by reducing reading times with computer-guided imaging data evaluation [7].

Efforts in the foreseeable future will improve our understanding of the precise nature and magnitude of these opportunities. We already know that different imaging modalities offer different capabilities, and these vary with the disease. For instance, MRI is more sensitive than CT imaging in the diagnosis of non-small cell lung cancer [58]. This is not different for AI methodologies, since they depend on and build upon the nature of the imaging data, and their potential mainly consists in enhancing this information.

The need for a multidisciplinary approach to treating most types of cancers is now universally recognized. Methodological changes in this direction naturally entail the need to exploit and recombine larger volumes of data. This will soon bring traditional statistical models to face their intrinsic limitations, and machine learning will become the only feasible methodological approach for clinical model building in many contexts. Embedding ML techniques is a computationally demanding and conceptually involved task, requiring adequate technological support and extended data manipulation. Meeting these performance requirements with reliable and scalable processes will allow more effective clinical trials support and design of more real-time utilization of radiological information in routine clinical settings. Such technical improvements may be achieved by combining many AI processes into a single treatment unit ("artificial swarm intelligence").

In any case, neither short- nor long-term AI-based innovations in radiology are likely to exclude the expert radiologist. Despite the hype surrounding the recent emergence of high throughput analytical methodologies for various aspects of cancer care (diagnosis, characterization, prognosis, and therapeutic follow-up), intelligent software that can automatically choose or adapt a sarcoma patient's treatment pathway based on their available medical record without any expert (radiologist, oncologist) input is not to be expected in the near future. The most imminent impact of AI in radiology will likely happen with its integration within modern radiological data management systems (such as PACS—picture archiving and communications systems) [29].

In melanoma and sarcoma more specifically, we may expect different directions of development of technological solutions. Based on the current state of the art, AI may more naturally emerge for melanoma with the opportunity to exploit non-radiological imaging data. As for sarcoma, therapeutic processes relying on AI solutions should not be different in nature to those developed for the characterization of more prevalent solid tumors, such as breast or non-small cell lung cancers. As such, methodological progress achieved for these diseases should be directly transferable to sarcoma care. Sarcomas however have image characteristics that may be relevant to the context of modern high throughput analytics. They often present as advanced tumors at time of diagnosis, and there is a larger proportion of tumors with relatively large sizes as a result of this fact, relative to other cancers. Other sarcoma specifics include the fact that the nature of surgical intervention (resection vs amputation) differs drastically with respect to location of the disease and the broad range of subtypes of this disease in contrast to its low prevalence in adults. These tumor specifics will affect and guide future design of intelligent processes for therapeutic decision-making in sarcoma.

Conclusion and Current Perspectives

Current Clinical Practice

Routine clinical protocols—for sarcoma, melanoma, and many other cancers—use simple statistical summaries of the PET imaging data, such

as tumor uptake volume and the average or maximum PET uptake value within the tumor VOI. These measures allow for direct image interpretation, and their prognostic potential has been well established. Such summaries however underexploit the wealth of information contained within a PET image dataset, and recent methodological developments indicate that complex tumor biological aspects can be captured noninvasively using more elaborate quantitation techniques. One key target indicator is intratumoral imaging agent spatial uptake heterogeneity (intertumoral heterogeneity is another key marker in stage IV melanoma). Many techniques are currently evaluated by the research community to capture this information, with growing evidence of their effectiveness. Quantitation methodologies relying on modern statistical frameworks can complement routine clinical assessment and provide opportunities for personalized therapy. These advances have prompted the use of machine learning algorithms toward more effective diagnostic and risk predictive models, resulting in the emergence of a whole new area of image analysis in radiology that includes radiomics. On the methodological horizon, these developments pave the way for the integration of complete artificial intelligence solutions within clinical practice.

Limitations and Need for Standardization

These advances globally remain to be validated for larger clinical cohorts and in multicenter studies. Efforts are also required for improved clinical understanding of complex prognostic models derived from radiomics, before these methodologies can be implemented in routine settings. A single reference subset of baseline radiomic variables for cancer prognosis does not yet exist for any cancer; different multivariate prognostic models are proposed with each new study. In fact, technical variations of various image processing parameters can greatly impact some of these complementary predictive variables. Recent efforts in the community, such as the Image

Biomarker Standardisation Initiative [3], clearly indicate the need for homogenization and standardization of quantitative practice, a significant challenge yet to be overcome, as the results of many studies indicate (see, for instance, multicenter studies of the American College of Radiology Imaging Network [4, 5]). Reaching colloquial agreement on future predictive models will also be guided by progress in our understanding of how PET-based tumor quantitation maps to phenotypic or genomic signatures.

Multimodality Data

The prevalence of multimodality (PET/MR and PET/CT) scanners has led to a surge in studies combining co-registered (i.e., re-aligned) image signals. For tumor delineation, a VOI mask defined from MR or CT information is often used to guide segmentation of the PET VOI. But tumor characterization can further benefit from combined imaging modality output, in order to align functional information on tumor biological features with the anatomical and physiological nature (bone, soft tissue, etc.) of corresponding areas within the VOI [59–61]. MT or CT information can complement assessment of PET spatial uptake heterogeneity and perhaps guide further understanding or interpretation of biologic descriptors derived from the PET image. A few analyses have been reported that explore various aspects of tumor assessment for sarcoma by analysis of recombined features [54, 62] and none for melanoma.

For multimodal imaging, two options are available. One is to associate scans acquired via several modalities, either by image co-registration or multimodality (e.g., PET/CT–MRI) scanners; these allow for true CT-based attenuation correction and reliably calibrated imaging data but may require higher radiation doses and longer acquisition times [59, 60]. The other option is to carry out separate scans and recombine extracted features post-acquisition, during the analytical phase, albeit using image features that may not originate from identical VOIs [54]. How to best combine these technologies and translate them to clinical practice

remains an open and challenging question, at both the operational and analytical levels. Multimodality imaging also allows integrating emerging PET technologies, such as activatable targeted nanoparticle probes, to create novel forms of imaging data, and requires the definition of adapted modeling strategies [59–61].

One practical motivation for analyses using two or more tomographs arises from patient assessment involving PET/MR and PET/CT devices (e.g., to optimize resources when both PET/MR and PET/CT scans are available on-site, or if several care centers are involved). In these contexts, direct comparison of the PET images is not guaranteed, as different calibration settings (dose, contrast, resolution, etc.) are likely and would require dedicated pre-processing to correct for statistical discrepancies in image characteristics. Homogeneous scanner characteristics would however allow for both PET images to be used for comparative analyses.

Kinetic Analysis

The methodologies described above only use static image information. They are limited in capturing other aspects of the tumor biology. For example, they do not provide insight on imaging agent vascular delivery and tissue retention, which may be important for disease assessment and treatment. Dynamic (or kinetic) analysis of the information contained in each time frame after injection of the PET radiotracer, and throughout the entire acquisition duration, allows for evaluation of key imaging parameters such as blood flow and tissue retention characteristics. Recent advances in kinetic analysis also suggest that tumor imaging agent spatial uptake heterogeneity may also be evaluated in dynamic PET information [63, 64]. Dynamic imaging data analysis also provides more informed tissue characterization and tumor delineation with reduced bias (compared to static summed PET or anatomical imaging data), which can benefit quantitation [64]. Kinetic analysis for PET dynamic image acquisition protocols remains mainly conducted in research settings, in part due to its more challenging practical requirements.

References

1. Joo Hyun O, Lodge MA, Wahl RL. Practical PERCIST: a simplified guide to PET response criteria in solid tumors 1.0. Radiology. 2016;280(2):576–84.
2. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. Radiology. 2016;278(2):563–77.
3. Zwanenburg A, Abdalah MA, Apte A, Ashrafinia S, Beukinga J, Bogowicz M, et al. Results from the image biomarker standardisation initiative. Radiother Oncol. 2018;127:S543–S4.
4. Gerstner ER, Zhang Z, Fink JR, Muzi M, Hanna L, Greco E, et al. ACRIN 6684: assessment of tumor hypoxia in newly diagnosed Glioblastoma Using 18F-FMISO PET and MRI. Clin Cancer Res. 2016;22(20):5079–86.
5. Ratai EM, Zhang Z, Fink J, Muzi M, Hanna L, Greco E, et al. ACRIN 6684: multicenter, phase II assessment of tumor hypoxia in newly diagnosed glioblastoma using magnetic resonance spectroscopy. PLoS One. 2018;13(6):e0198548.
6. Wolsztynski E, O'Sullivan F, Keyes E, O'Sullivan J, Eary JF. Positron emission tomography-based assessment of metabolic gradient and other prognostic features in sarcoma. J Med Imaging (Bellingham). 2018;5(2):024502.
7. Russell SJ, Norvig P. Artificial intelligence: a modern approach. 3rd ed. Prentice Hall: Upper Saddle River; 2010.
8. Rodriguez-Ruiz A, Krupinski E, Mordang JJ, Schilling K, Heywang-Kobrunner SH, Sechopoulos I, et al. Detection of breast Cancer with mammography: effect of an artificial intelligence support system. Radiology. 2019;290(2):305–14.
9. Al-Nabhani KZ, Syed R, Michopoulou S, Alkalbani J, Afaq A, Panagiotidis E, et al. Qualitative and quantitative comparison of PET/CT and PET/MR imaging in clinical practice. J Nucl Med. 2014;55(1):88–94.
10. Moses WW. Fundamental Limits of Spatial Resolution in PET. Nucl Instrum Methods Phys Res A. 2011;648(Supplement 1):S236–S40.
11. Mou T, Huang J, O'Sullivan F. The gamma characteristic of reconstructed PET images: implications for ROI analysis. IEEE Trans Med Imaging. 2018;37(5):1092–102.
12. Soret M, Bacharach SL, Buvat I. Partial-volume effect in PET tumor imaging. J Nucl Med. 2007;48(6):932–45.
13. Hatt M, Majdoub M, Vallieres M, Tixier F, Le Rest CC, Groheux D, et al. 18F-FDG PET uptake characterization through texture analysis: investigating the complementary nature of heterogeneity and func-

tional tumor volume in a multi-cancer site patient cohort. J Nucl Med. 2015;56(1):38–44.

14. Hatt M, Tixier F, Pierce L, Kinahan PE, Le Rest CC, Visvikis D. Characterization of PET/CT images using texture analysis: the past, the present... any future? Eur J Nucl Med Mol Imaging. 2017;44(1):151–65.

15. Yan J, Lim JC, Townsend DW. MRI-guided brain PET image filtering and partial volume correction. Phys Med Biol. 2015;60(3):961–76.

16. Kratochwil C, Flechsig P, Lindner T, Abderrahim L, Altmann A, Mier W, et al. (68)Ga-FAPI PET/CT: tracer uptake in 28 different kinds of cancer. J Nucl Med. 2019;60(6):801–5.

17. Hasebe M, Yoshikawa K, Nishii R, Kawaguchi K, Kamada T, Hamada Y. Usefulness of (11) C-methionine-PET for predicting the efficacy of carbon ion radiation therapy for head and neck mucosal malignant melanoma. Int J Oral Maxillofac Surg. 2017;46(10):1220–8.

18. Bedard PL, Hansen AR, Ratain MJ, Siu LL. Tumour heterogeneity in the clinic. Nature. 2013;501(7467):355–64.

19. Eary JF, Conrad EU, O'Sullivan J, Hawkins DS, Schuetze SM, O'Sullivan F. Sarcoma mid-therapy [F-18]fluorodeoxyglucose positron emission tomography (FDG PET) and patient outcome. J Bone Joint Surg Am. 2014;96(2):152–8.

20. Grzywa TM, Paskal W, Wlodarski PK. Intratumor and intertumor heterogeneity in melanoma. Transl Oncol. 2017;10(6):956–75.

21. Davnall F, Yip CS, Ljungqvist G, Selmi M, Ng F, Sanghera B, et al. Assessment of tumor heterogeneity: an emerging imaging tool for clinical practice? Insights Imaging. 2012;3(6):573–89.

22. Buvat I, Orlhac F, Soussan M. Tumor texture analysis in PET: where do we stand? J Nucl Med. 2015;56(11):1642–4.

23. Soussan M, Orlhac F, Boubaya M, Zelek L, Ziol M, Eder V, et al. Relationship between tumor heterogeneity measured on FDG-PET/CT and pathological prognostic factors in invasive breast cancer. PLoS One. 2014;9(4):e94017.

24. O'Sullivan F, Wolsztynski E, O'Sullivan J, Richards T, Conrad EU, Eary JF. A statistical modeling approach to the analysis of spatial patterns of FDG-PET uptake in human sarcoma. IEEE Trans Med Imaging. 2011;30(12):2059–71.

25. Yancovitz M, Finelt N, Warycha MA, Christos PJ, Mazumdar M, Shapiro RL, et al. Role of radiologic imaging at the time of initial diagnosis of stage T1b-T3b melanoma. Cancer. 2007;110(5):1107–14.

26. Rajkumar S, Watson IR. Molecular characterisation of cutaneous melanoma: creating a framework for targeted and immune therapies. Br J Cancer. 2016;115(2):145–55.

27. Carlino MS, Saunders CA, Haydu LE, Menzies AM, Martin Curtis C Jr, Lebowitz PF, et al. (18)F-labelled fluorodeoxyglucose-positron emission tomography (FDG-PET) heterogeneity of response is prognostic

in dabrafenib treated BRAF mutant metastatic melanoma. Eur J Cancer. 2013;49(2):395–402.

28. Mena E, Sanli Y, Marcus C, Subramaniam RM. Precision medicine and PET/computed tomography in melanoma. PET Clin. 2017;12(4):449–58.

29. Fazal MI, Patel ME, Tye J, Gupta Y. The past, present and future role of artificial intelligence in imaging. Eur J Radiol. 2018;105:246–50.

30. Zhang Z, Chen P, McGough M, Xing F, Wang C, Bui M, et al. Pathologist-level interpretable whole-slide cancer diagnosis with deep learning. Nat Mach Intell. 2019;1(5):236–45.

31. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: past, present and future. Stroke Vasc Neurol. 2017;2(4):230–43.

32. Kahn CE Jr. Artificial intelligence in radiology: decision support systems. Radiographics. 1994;14(4):849–61.

33. Rodriguez-Ruiz A, Lang K, Gubern-Merida A, Broeders M, Gennaro G, Clauser P, et al. Stand-alone artificial intelligence for breast cancer detection in mammography: comparison with 101 radiologists. J Natl Cancer Inst. 2019;111(9):916–22.

34. Christ PF, Elshaer MEA, Ettlinger F, Tatavarty S, Bickel M, Bilic P, et al. Automatic liver and lesion segmentation in CT using cascaded fully convolutional neural networks and 3D conditional random fields. arXiv e-prints [Internet]. October 1, 2016. Available from: https://ui.adsabs.harvard.edu/abs/2016arXiv161002177C.

35. Havaei M, Davy A, Warde-Farley D, Biard A, Courville A, Bengio Y, et al. Brain tumor segmentation with deep neural networks. Med Image Anal. 2017;35:18–31.

36. Savadjiev P, Chong J, Dohan A, Agnus V, Forghani R, Reinhold C, et al. Image-based biomarkers for solid tumor quantification. Eur Radiol. 2019;29(10):5431–40.

37. Berthon B, Evans M, Marshall C, Palaniappan N, Cole N, Jayaprakasam V, et al. Head and neck target delineation using a novel PET automatic segmentation algorithm. Radiother Oncol. 2017;122(2):242–7.

38. Bakas S, Reyes M, Jakab A, Bauer S, Rempfler M, Crimi A, et al. Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge. arXiv e-prints [Internet]. November 1, 2018. Available from: https://ui.adsabs.harvard.edu/abs/2018arXiv181102629B.

39. Pirotte B, Goldman S, Massager N, David P, Wikler D, Vandesteene A, et al. Comparison of 18F-FDG and 11C-methionine for PET-guided stereotactic brain biopsy of gliomas. J Nucl Med. 2004;45(8):1293–8.

40. Schwartz LH, Litiere S, de Vries E, Ford R, Gwyther S, Mandrekar S, et al. RECIST 1.1-update and clarification: from the RECIST committee. Eur J Cancer. 2016;62:132–7.

41. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, et al. Radiomics:

extracting more information from medical images using advanced feature analysis. Eur J Cancer. 2012;48(4):441–6.

42. Guyon I, Elisseeff A. An introduction to variable and feature selection. J Mach Learn Res. 2003;3(7–8):1157–82.

43. 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(10):6562–6.

44. Hastie T, Tibshirani R, Friedman JH. The elements of statistical learning: data mining, inference, and prediction. 2nd ed. New York: Springer; 2009. xxii, 745 p.

45. Magalhaes C, Mendes J, Vardasca R. The role of AI classifiers in skin cancer images. Skin Res Technol. 2019;25(5):750–7.

46. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts H. Artificial intelligence in radiology. Nat Rev Cancer. 2018;18(8):500–10.

47. Bhattacharya A, Young A, Wong A, Stalling S, Wei M, Hadley D. Precision diagnosis of melanoma and other skin lesions from digital images. AMIA Jt Summits Transl Sci Proc. 2017;2017:220–6.

48. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Corrigendum: dermatologist-level classification of skin cancer with deep neural networks. Nature. 2017;546(7660):686.

49. Gareau DS, Correa da Rosa J, Yagerman S, Carucci JA, Gulati N, Hueto F, et al. Digital imaging biomarkers feed machine learning for melanoma screening. Exp Dermatol. 2017;26(7):615–8.

50. Dreiseitl S, Ohno-Machado L, Kittler H, Vinterbo S, Billhardt H, Binder M. A comparison of machine learning methods for the diagnosis of pigmented skin lesions. J Biomed Inform. 2001;34(1):28–36.

51. Sun R, Limkin EJ, Vakalopoulou M, Dercle L, Champiat S, Han SR, et al. A radiomics approach to assess tumour-infiltrating CD8 cells and response to anti-PD-1 or anti-PD-L1 immunotherapy: an imaging biomarker, retrospective multicohort study. Lancet Oncol. 2018;19(9):1180–91.

52. Giesel FL, Schneider F, Kratochwil C, Rath D, Moltz J, Holland-Letz T, et al. Correlation between SUVmax and CT radiomic analysis using lymph node density in PET/CT-based lymph node staging. J Nucl Med. 2017;58(2):282–7.

53. Kniep HC, Madesta F, Schneider T, Hanning U, Schonfeld MH, Schon G, et al. Radiomics of brain MRI: utility in prediction of metastatic tumor type. Radiology. 2019;290(2):479–87.

54. Vallieres 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. Phys Med Biol. 2015;60(14):5471–96.

55. Corino VDA, Montin E, Messina A, Casali PG, Gronchi A, Marchiano A, et al. Radiomic analysis of soft tissues sarcomas can distinguish intermediate from high-grade lesions. J Magn Reson Imaging. 2018;47(3):829–40.

56. Wu Y, Xu L, Yang P, Lin N, Huang X, Pan W, et al. Survival prediction in high-grade osteosarcoma using radiomics of diagnostic computed tomography. EBioMedicine. 2018;34:27–34.

57. Ardila D, Kiraly AP, Bharadwaj S, Choi B, Reicher JJ, Peng L, et al. Author correction: end-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. Nat Med. 2019;25(8):1319.

58. Planchard D, Popat S, Kerr K, Novello S, Smit EF, Faivre-Finn C, et al. Metastatic non-small cell lung cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol. 2018;29(Supplement_4):iv192–237.

59. Chetty IJ, Martel MK, Jaffray DA, Benedict SH, Hahn SM, Berbeco R, et al. Technology for innovation in radiation oncology. Int J Radiat Oncol Biol Phys. 2015;93(3):485–92.

60. Veit-Haibach P, Kuhn FP, Wiesinger F, Delso G, von Schulthess G. PET–MR imaging using a tri-modality PET/CT–MR system with a dedicated shuttle in clinical routine. MAGMA. 2013;26(1):25–35.

61. Kircher MF, Hricak H, Larson SM. Molecular imaging for personalized cancer care. Mol Oncol. 2012;6(2):182–95.

62. Vallieres M, Zwanenburg A, Badic B, Cheze Le Rest C, Visvikis D, Hatt M. Responsible radiomics research for faster clinical translation. J Nucl Med. 2018;59(2):189–93.

63. Asselin M-C, O'Connor JPB, Boellaard R, Thacker NA, Jackson A. Quantifying heterogeneity in human tumours using MRI and PET. Eur J Cancer. 2012;48(4):447–55.

64. Muzi M, O'Sullivan F, Mankoff DA, Doot RK, Pierce LA, Kurland BF, et al. Quantitative assessment of dynamic PET imaging data in cancer imaging. Magn Reson Imaging. 2012;30(9):1203–15.