



Semi-automatic segmentation from intrinsically-registered 18F-FDG–PET/MRI for treatment response assessment in a breast cancer cohort: comparison to manual DCE–MRI

Maren Marie Sjaastad Andreassen¹ · Pål Erik Goa^{2,3} · Torill Eidhammer Sjøbakk¹ · Roja Hedayati^{4,5} · Hans Petter Eikesdal^{6,7} · Callie Deng¹ · Agnes Østlie³ · Steinar Lundgren^{4,5} · Tone Frost Bathen^{1,3} · Neil Peter Jerome^{1,3}

Received: 13 July 2019 / Revised: 27 August 2019 / Accepted: 16 September 2019 / Published online: 27 September 2019
© The Author(s) 2019

Abstract

Objectives To investigate the reliability of simultaneous positron emission tomography and magnetic resonance imaging (PET/MRI)-derived biomarkers using semi-automated Gaussian mixture model (GMM) segmentation on PET images, against conventional manual tumor segmentation on dynamic contrast-enhanced (DCE) images.

Materials and methods Twenty-four breast cancer patients underwent PET/MRI (following 18F-fluorodeoxyglucose (18F-FDG) injection) at baseline and during neoadjuvant treatment, yielding 53 data sets (24 untreated, 29 treated). Two-dimensional tumor segmentation was performed manually on DCE–MRI images (manual DCE) and using GMM with corresponding PET images (GMM–PET). Tumor area and mean apparent diffusion coefficient (ADC) derived from both segmentation methods were compared, and spatial overlap between the segmentations was assessed with Dice similarity coefficient and center-of-gravity displacement.

Results No significant differences were observed between mean ADC and tumor area derived from manual DCE segmentation and GMM–PET. There were strong positive correlations for tumor area and ADC derived from manual DCE and GMM–PET for untreated and treated lesions. The mean Dice score for GMM–PET was 0.770 and 0.649 for untreated and treated lesions, respectively.

Discussion Using PET/MRI, tumor area and mean ADC value estimated with a GMM–PET can replicate manual DCE tumor definition from MRI for monitoring neoadjuvant treatment response in breast cancer.

Keywords Breast cancer · Diffusion imaging · Mixture modelling · PET/MRI · Segmentation

Introduction

Breast cancer is the most frequent type of cancer in women worldwide [1], with a mean 5-year survival of 90.4% in Norway [2]. Patients diagnosed with locally advanced breast

cancer (LABC, stage 3), have a worse survival outcome (78.3%) [2]. They receive neoadjuvant chemotherapy treatment before surgery with the goal of complete pathological tumor regression, which correlates with improved survival and a reduced chance of breast cancer recurrence [3].

✉ Neil Peter Jerome
neil.p.jerome@ntnu.no

¹ Department of Circulation and Medical Imaging, NTNU, Norwegian University of Science and Technology, Trondheim, Norway

² Department of Physics, NTNU, Norwegian University of Science and Technology, Trondheim, Norway

³ Department of Radiology and Nuclear Medicine, St. Olav's University Hospital, Trondheim, Norway

⁴ Department of Clinical and Molecular Medicine, NTNU, Norwegian University of Science and Technology, Trondheim, Norway

⁵ Department of Oncology, St. Olav's University Hospital, Trondheim, Norway

⁶ Section of Oncology, Department of Clinical Science, University of Bergen, Bergen, Norway

⁷ Department of Oncology, Haukeland University Hospital, Bergen, Norway

Objective response evaluation during neoadjuvant therapy is important to assess treatment efficacy and to avoid unnecessary toxic side effects [4]. Radiologically, response evaluation has traditionally focused on measurements of tumor size [5], but several recent studies [6–10] have established functional imaging modalities as useful indicators of early response during neoadjuvant chemotherapy.

Diffusion-weighted magnetic resonance imaging (DWI) is a functional imaging modality with contrast arising from water molecule motion, and is, therefore, sensitized to tissue microstructure characteristics. DWI is most commonly utilized to assess tissue cellularity, where highly cellular tissues such as malignant tumors exhibit decreased diffusivity [11], quantified by calculation of an apparent diffusion coefficient (ADC). A robust empirical biomarker that is reduced in malignant tumors [12], ADC has shown higher specificity than conventional anatomical MRI for discriminating malignant and benign breast tumors [13].

Tumor ADC is commonly measured by the mean value of manually placed regions-of-interest (ROIs). There is no standard protocol for this tumor segmentation, and different approaches can significantly influence resulting ADC values [14]. Given that direct tumor segmentation of DWI may be confounded by noise and lack of conspicuity, tumor ROIs are commonly delineated on dynamic contrast-enhanced (DCE) images before being transferred to DWI. The definition of tumor on DCE images is thus governed by leakage of gadolinium contrast through pathological vessels and, therefore, linked to vascularity, whereas diffusion changes, reflecting cellularity, do not necessarily coincide [15].

Simultaneous positron emission tomography and magnetic resonance imaging (PET/MRI) is a recent technology with a significant potential in many aspects of breast cancer practice, including diagnostics, staging, and neoadjuvant response evaluation [16]. PET/MRI examinations allow simultaneous collection of structural, functional, and metabolic imaging properties in the same spatial and temporal domain. ¹⁸F-fluorodeoxyglucose (FDG)–PET visualizes upregulated glucose metabolism, while MRI reflects other hallmarks of cancer [17] including invasion and metastatic propensity (by ADC) and increased angiogenesis (DCE). Several studies report correlations between standardized uptake values (SUV) from FDG–PET and ADC in malignant tissue [18–20], indicating that intrinsically-registered ¹⁸F-FDG uptake may provide an alternative approach to manually drawn DCE–ROI delineation for use in DWI analysis [19]. FDG–PET is also known to outperform MRI tumor volume measurements in some cancers [21]. In this study, a simple, semi-automated Gaussian mixture model (GMM) segmentation algorithm was selected, to allow for heterogeneous FDG uptake across tumors and expected decline through treatment [22, 23].

The aim of the current study is to investigate the reliability of deriving lesion diffusion imaging characteristics from ¹⁸F-FDG uptake in invasive breast cancers > 4 cm or LABC (i.e., cT2-4N0-3) during neoadjuvant treatment. Specifically, we tested the reliability of deriving functional tumor area and ADC values in diffusion-weighted images from intrinsically-registered ¹⁸F-FDG–PET uptake using a semi-automated GMM segmentation algorithm in comparison with metrics derived from manually drawn DCE–ROIs.

Materials

Participants

This prospective study was approved by the Regional Committee for Medical and Health Research Ethics (REC) in western Norway (identifier 2015/1493). Informed consent was obtained from all individual participants included in the study. A total of 24 patients (median age 53 years, range 37–74) with biopsy-proven, invasive breast cancers > 4 cm or LABC (i.e., cT2-4N0-3). This minimum size was an inclusion criterion for recruitment to the phase II PETREMAC trial (Clinicaltrials.gov #NCT02624973), where lesions of this size are targets for neoadjuvant chemotherapy. The patients underwent individualized neoadjuvant therapy, based on tumor characteristics: estrogen (ER)/progesterone receptor (PgR), human epidermal growth factor-2 (HER2), and TP53 mutation status. The therapy used was primarily endocrine treatment (full details given in Table 1).

Patients were examined with ¹⁸F-FDG–PET/MRI scans, at baseline and up to four scanning sessions during neoadjuvant treatment (depending on trial progression and individual response). Two patients received neoadjuvant therapy 2 days prior to the baseline scan. The cohort scans are summarized in Fig. 1. All except three patients have undergone breast surgery (one patient dropped out, two scheduled after time of analysis). Tumor categorization was done by histopathologic analysis of core needle and open incisional biopsies.

Lesions with no remaining enhancement on DCE were excluded (8 data sets), resulting in 53 data sets overall: 24 untreated lesions, and 29 of treated lesions. Median number of scans was 2 (range 1–5), with mean intervals from baseline of 6, 12, 16, and 24 weeks. The mean time from study entry to surgery was 23 weeks. The two patients receiving neoadjuvant therapy 2 days prior to the baseline scan were considered untreated, as the lesions at this timepoint had undergone minimal treatment effect. This study did not explicitly consider clinical outcome or treatment effects, and therefore, this classification is predominantly to distinguish between lesions that have had the opportunity to undergo significant response.

Table 1 Clinical characteristics of patient cohort

Characteristic	All patients ($n=24$)
Age (median, range), years	53 (37–74)
Height (median, range), m	1.65 (1.54–1.79)
Weight (median, range), kg	67 (50–100)
Tumor volume (median, range), cm ³	9.91 (2.88–60.56)
Histological type	
IDC	18
ILC	2
Other	4
Histological grade	
1	0
2	9
3	13
Unknown	2
Estrogen receptor (ER) status (%)	
Negative	8
≥ 1 –10	1
≥ 10 –50	0
> 50	15
Progesteron receptor (PgR) status (%)	
Negative	9
≥ 10 –50	1
≥ 50	4
HER2 status	
Negative	15
Positive	7
Not applicable	1
Ki67 (%)	
$< 30\%$	9
$\geq 30\%$	15
Treatment	
Endocrine	11
Docetaxel and cyclophosphamid	1
Pertuzumab, trastuzumab and docetaxel/ cyclophosphamid	3
Pertuzumab, trastuzumab and docetaxel	5
Olaparib and carboplatin	4
Pathological response	
Complete response	8
Non-response	13
Not operated	3

Pathological characteristics are determined based on histopathologic analysis of pre-treatment core needle biopsy; for $n=3$ patients, histological grade was determined from surgical specimen

Others (ICD and ILC ($n=1$), poorly differentiated carcinoma ($n=1$), carcinoma with medullary features ($n=2$))

IDC invasive ductal carcinoma, ILC invasive lobular carcinoma

Methods

PET/MRI acquisition

All patients underwent simultaneous PET/MRI on a 3 T Biograph mMR scanner (Siemens Healthcare, Erlangen, Germany), 75 min after 18F-FDG injection (4 MBq/kg dose following 6 h fasting). The 18F-FDG was produced by the Norwegian Medical Cyclotron Center (Oslo, Norway). MRI acquisition utilized a designated 4-channel breast coil and included Dixon, T2-weighted, DWI, and DCE. DCE parameters included: 3D FLASH sequence, transverse orientation, TR/TE 5.88/2.21 ms, resolution $0.7 \times 0.7 \times 2.5$ mm, 72 slices, flip angle 15° , 1 baseline, and 7 contrast sequences, time resolution 1 min. Multiple b value DWI parameters were: axial bilateral single-shot echo planar imaging, TR/TE 9000/77 ms, fat suppressed, b values = 0, 50, 120, 200, 400, 700 mm² s⁻¹, resolution $2 \times 2 \times 2.5$ mm, 60 slices, and FoV 380×190 mm, with additional phase-reversed $b=0$ mm² s⁻¹ (hereafter ‘b0’) image. Concurrent PET data were acquired at a bed position giving full breast region coverage, and reconstructed using a manufacturer-supplied algorithm (OSEM-PSF, 21 subsets, 3 iterations, and a 4 mm full-width-half-maximum Gaussian filter; Siemens, Erlangen). PET data concurrent with DCE acquisition were available for $n=31$ data sets (16 untreated and 15 treated lesions).

Image preparation and manual segmentation (Manual DCE)

The lesion segmentation process is summarized in Fig. 2. DW images were distortion-corrected using phase-reversed b0 images [24]; dixon-based PET attenuation correction was performed on the scanner [25], and standardized uptake values (SUV) were normalized against body weight and dose. DCE and PET images were resampled using Elastix [26] to exactly match the resolution of the DWI images, and thus give direct voxel-to-voxel correspondence. Perfusion-insensitive ADC maps were calculated from DWI data ($b \geq 200$ mm² s⁻¹) using a mono-exponential model:

$$S_b = S_{b=0} \cdot e^{-b \cdot ADC}$$

Manual DCE: manual segmentation of a single tumor region was performed by researcher (M.M.S.A.) on a single central slice of the enhancing solid tumor on DCE, ignoring satellite regions, with resulting ROIs supervised and approved by an expert radiologist (A.Ø.).

Fig. 1 Breast cancer patients ($n=24$) received one pretreatment baseline scan, and additional scans during neoadjuvant treatment prior to surgery. On average, scans during treatment were 6, 12, 16, and 24 weeks after baseline, and surgery was 23 weeks after baseline scan. Total data set included 53 scans: 24 from untreated, and 29 from treated lesions

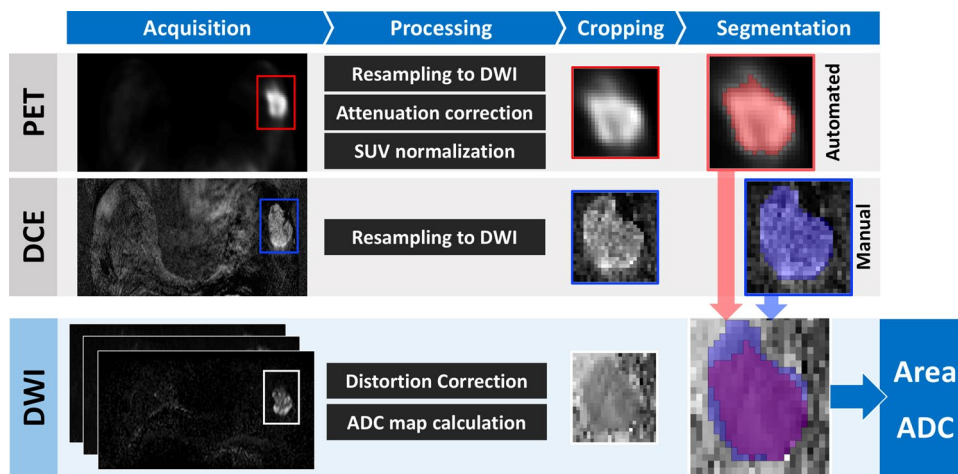
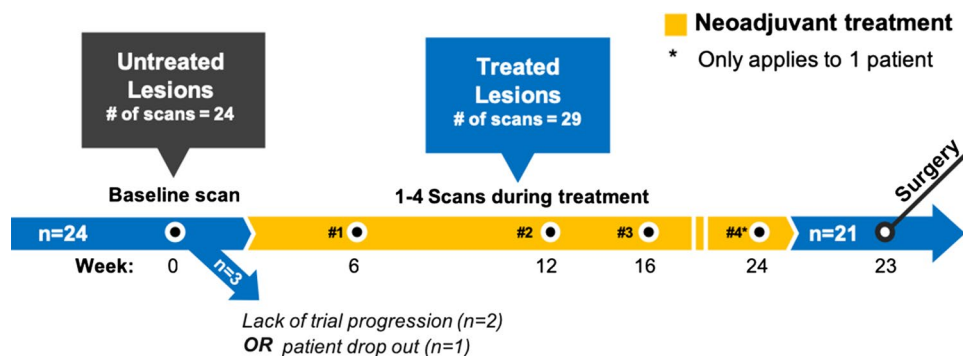


Fig. 2 All images were resampled to diffusion-weighted imaging (DWI) resolution. PET images were attenuation-corrected and SUV normalized. Apparent diffusion coefficient (ADC) maps were calculated from distortion-corrected diffusion images. Tumor segmenta-

tion was performed by semi-automated Gaussian mixture modelling (GMM) segmentation on cropped PET images, and manually on DCE images. Resulting regions-of-interest (ROIs) were transferred to the ADC maps for derivation of tumor area and mean ADC

Lesion cropping and Gaussian mixture modelling (GMM–PET)

A rectangular region containing the visible lesion was manually cropped from the SUV map (corresponding to the single central slice of enhancing tumor on DCE) for Gaussian mixture modelling (GMM–PET). An algorithm using default k-means++ [27] initialization (MATLAB; Mathworks, Natick, MA, USA) and an assumption of three Gaussian distribution classes were used, returning an assignment for each voxel based on highest probability [28] of belonging to each class: tumor (highest intensity), ‘non-tumor’ background (lowest intensity) and unknown (intermediate intensity). To compromise between accuracy and avoiding overestimation from partial volume effects, voxels classed as ‘unknown’ were considered non-tumor, defining the tumor class threshold as the intersection of tumor and unknown class distributions (Fig. 7 in Appendix). User input is thus limited to initial region cropping.

Two common simple thresholding-based PET segmentation methods, a fixed threshold of 2.5 ($SUV_{2.5}$) and 42% of the maximum SUV ($SUV_{42\%}$) [22], were also performed to provide comparison with GMM–PET (Fig. 7 in Appendix).

Derivation of DWI metrics from manual DCE and GMM–PET

ROIs, for the whole lesion within the chosen slice, derived from both manual DCE and from GMM–PET were transferred to ADC maps and used to calculate the tumor ROI area and the mean ADC value for the whole ROI.

Statistical analysis

Performance of the three PET segmentation techniques in reference to manual DCE–ROIs was measured using the Dice similarity coefficient, varying between 0 and 1

indicating degree of spatial overlap [29], and center-of-gravity displacement (CoG). CoG was normalized based on corresponding area DCE. Tumor area and ADC values from the different segmentation methods were compared using a paired *t* test, and Pearson's test for correlation; relationship of these segmentation metrics with $SUV_{2.5}$ and $SUV_{42\%}$ was also assessed. A two-sample *t* test was used to assess the difference between the untreated ($n=24$) and treated cohort ($n=29$) for all metrics. All *p* values were corrected for multiple testing with the Benjamini and Hochberg [30] approach, with values <0.05 considered statistically significant.

Results

Derived diffusion parameters (area, ADC) in GMM–PET versus manual DCE

DWI metrics for untreated, treated, and all lesions are given in Table 2, and show no significant differences. There were strong positive correlations between GMM–PET and manual

DCE for area and ADC for untreated and treated lesions, as shown for longitudinal scans of two patients receiving neo-adjuvant therapy in Fig. 3. GMM–PET successfully tracks the same changes in ADC and tumor area observed using the manual DCE, even when performance parameters to manual DCE are poor.

Spatial agreement of GMM–PET with manual DCE

Dice score [29] for GMM–PET was significantly higher, indicating better performance, than $SUV_{42\%}$ for untreated lesions ($p=0.012$) and higher than $SUV_{2.5}$ for both untreated ($p=0.024$) and treated lesions ($p<0.001$) (Fig. 4a). CoG measurements were significantly lower for GMM–PET compared to $SUV_{2.5}$ for treated lesions ($p=0.002$) (Fig. 4b). GMM–PET is able to successfully identify tumor tissue in untreated lesions where uptake is heterogeneous across the cohort, where $SUV_{42\%}$ and $SUV_{2.5}$ over- and underestimate tumor areas, respectively, compared to the DCE definition. In 16 cases (3 untreated, 13 treated lesions), $SUV_{2.5}$ could not define any tumor area, meaning that CoG measurements

Table 2 Values given as mean (range) and *p* values

	ADC mean ($\times 10^{-3} \text{ mm}^2 \text{ s}^{-1}$)			Tumor area (cm^2)		
	Manual DCE	GMM–PET	<i>p</i> value	Manual DCE	GMM–PET	<i>p</i> value
Untreated ($n=24$)	0.957 (0.3796)	0.964 (0.411)	0.930	6.189 (4.277)	5.923 (3.944)	0.899
Treated ($n=29$)	1.167 (0.392)	1.1701 (0.445)	0.930	4.076 (2.857)	4.147 (2.745)	0.930
All ($n=53$)	1.073 (0.397)	1.079 (0.439)	0.930	5.015 (3.677)	4.936 (3.416)	0.930

No significant difference between resulting parameters from manual DCE and GMM–PET

Fig. 3 Patient in **a** demonstrated good spatial overlay and excellent agreement of response parameters over time. The patient in **b** shows a patient with excellent spatial agreement for the untreated lesion, but for the treated lesion, the segmentation is offset, with corresponding poor performance parameters (low Dice score and high CoG). However, GMM–PET was still able to accurately assess parameter changes over time

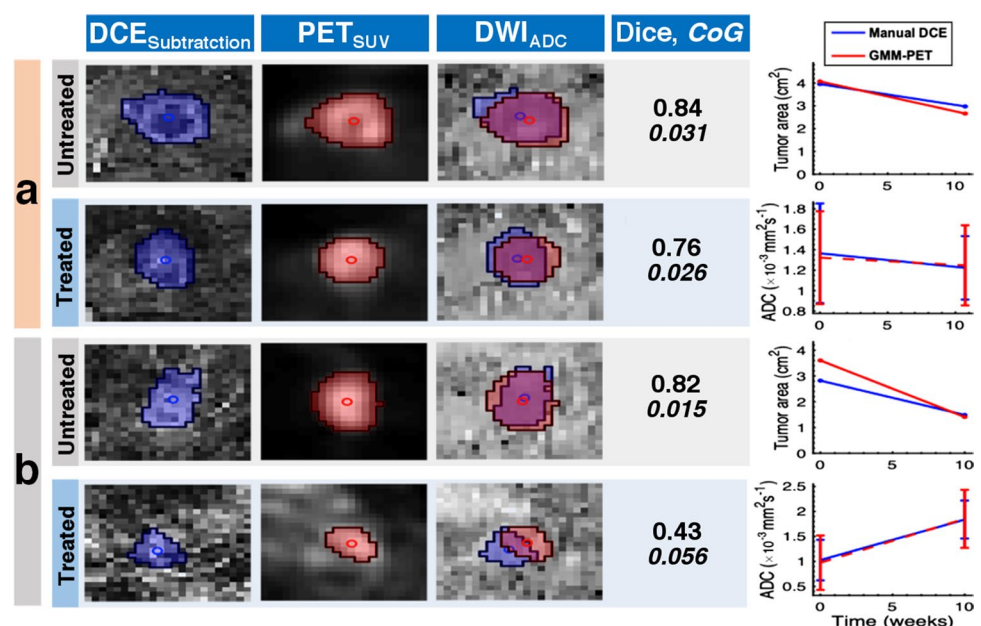
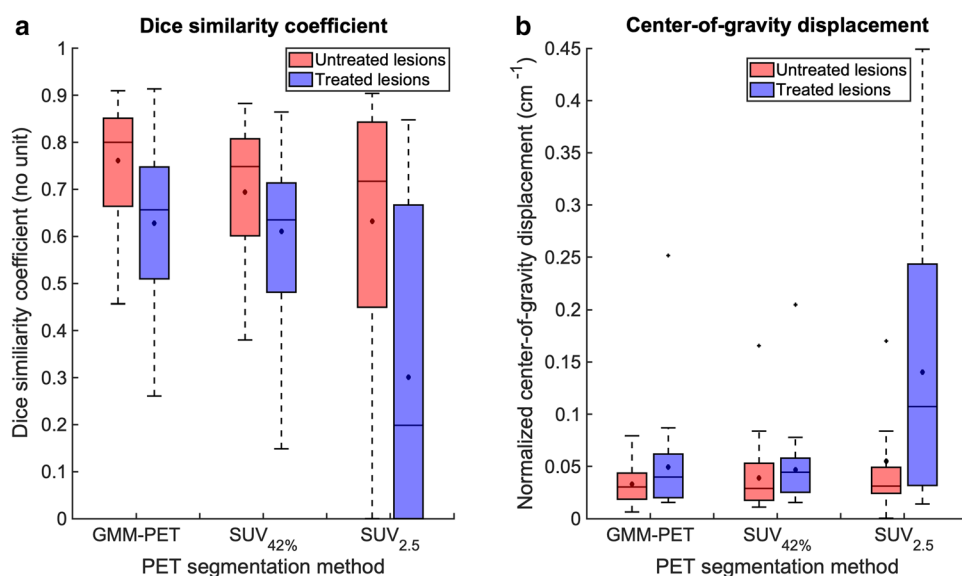


Fig. 4 **a** Dice similarity coefficient and **b** center-of-gravity displacement, normalized to manual DCE tumor area, from GMM, SUV_{42%}, and SUV_{2.5}. Median and mean values indicated by lines and asterisks; boxes show interquartile range, and whiskers show data range



were not applicable for these cases. GMM-PET and SUV_{2.5} performed significantly better in the treated lesions group compared to untreated lesions for both Dice score ($p=0.005$ and $p=0.002$) and CoG ($p=0.025$ and $p=0.005$), while SUV_{42%} only had significantly higher CoG ($p=0.002$).

Discussion

The main finding of the current breast cancer study is that functional tumor area and corresponding mean ADC values from GMM-PET ROIs matched those derived from manual DCE. As a superficial interpretation, these findings indicate that GMM-PET is a reliable technique to efficiently derive functional diffusion parameters for monitoring neoadjuvant treatment response in breast cancer. The segmentation is data driven, requiring minimal user input, and obviates the requirement for gadolinium contrast administration and, therefore, could have significant potential as an alternative objective evaluation method among the increasing number of breast cancer patients undergoing neoadjuvant treatment. At a deeper level, the results illuminate the concordance—and discordance—between ROIs derived from different imaging modalities, and as such allow interrogation of the spatial relationship existing between functional information arising from PET, diffusion, and DCE imaging, and ultimately the tissue characteristics these modalities are sensitized to.

Our study demonstrates a strong correlation between tumor ADC values derived from GMM-PET and manual DCE segmentation, in line with a previous study by Byun et al. [19] using a similar approach in breast carcinomas; our study utilizes the intrinsic voxel correspondence of simultaneous PET/MRI, thus avoiding the additional registration

required by sequential FDG-PET/CT and DWI and conferring greater confidence in the results. Notably, the calculated mean ADC from GMM-PET was not significantly different from mean ADC from manual DCE, despite ADC metrics having been shown to be significantly influenced by segmentation method [14, 31]. This suggests that GMM-PET may have value even while accurate assessment of ADC metrics is considered increasingly important in a neoadjuvant treatment response setting [6–8].

Conventional manual DCE segmentation means that diffusion measurements, reflecting cellularity [13, 32, 33], are drawn from areas defined by gadolinium contrast enhancement, which is not necessarily optimal and may introduce bias to functional biomarker measurements [15]. It can thus be argued that tumor definition for diffusion studies is better performed on another MR modality more closely related to cellularity. Several studies have described an underlying link between metabolism and cellularity, such as correlation between FDG uptake to cellularity [34, 35]. Consequently, a negative correlation should be expected between SUV and ADC, and it could be argued that GMM-PET would coincide better with changes in cellularity. However, the previous reports are contradictory with either negative [18–20] or no [36, 37] correlations between SUV and ADC, indicating that imaging metrics from DCE, DWI, and PET do not capture all relevant physiological properties, even when GMM-PET is able to localize tumors equivalent to DCE.

GMM-PET segmentation gives good spatial concordance with manual DCE for untreated breast cancer lesions, while the segmentation performance was significantly poorer for treated lesions, with lower Dice score and higher CoG. It is well known that therapy affects tumor vascularity [38], which may have influenced both manual DCE and

GMM–PET segmentation, as both modalities are dependent on sufficient blood flow for contrast and tracer uptake. However, DCE and PET reflect different physiological properties, and therefore, it would be of interest to observe if these have been altered differently during treatment. In addition, it should be noted that several treated lesions with poor overlap are cases where manual DCE segmentation was difficult due to low enhancement on DCE.

In this study, GMM–PET segmentation performed significantly better than the commonly used $SUV_{42\%}$ threshold in untreated lesions, and $SUV_{2.5}$ for both treated and untreated lesions in recapitulating manual DCE. Using GMM–PET, the tumor area is not underestimated in the cases of heterogeneous uptake (Fig. 5) or for tumors with high-intensity relative to their immediate surroundings, which is a well-known issue [39, 40]. However, as FDG uptake in the tumor decreases during treatment [23], GMM–PET did not perform better than $SUV_{42\%}$, which is known to give larger estimates of tumor size as SUV_{max} approaches background levels (Fig. 5), which is a limit of adaptive and data-driven algorithms that are sensitive to the FDG-uptake range [41, 42]. Other algorithm-based approaches such as gradient methods [40, 43] have also been recommended [22], although these require increased user input and were not investigated in this study (Fig. 6).

While the current findings suggest that GMM–PET segmentation can work as a proxy for manual DCE, we do not suggest that PET might replace manual DCE in today’s clinical context. Manual DCE is currently the most sensitive test for both breast cancer detection [44]

and monitoring treatment response in a neoadjuvant setting [45], although concerns regarding contrast allergy and potential brain deposition [46] of gadolinium create a setting for exploration of complementary techniques. The use of PET tracer comes with its own challenges, in handling and cost, and is neither available nor suitable in all contexts. Our study indicates that the use of PET data for tumor segmentation is more reliable in pre-treatment lesions; in cases where FDG uptake is substantially reduced by treatment the GMM–PET method becomes less effective, where the tumor may become more diffuse. In these cases, automated segmentation procedures will be more prone to variation. It is worth noting that this is not unique to the technique in this study; the reduction of DCE contrast in successfully treated tumors also makes tumor definition more challenging for the conventional approach. Thus, in a simple sense, PET data are able to act as a proxy for tumor definition using DCE, but are also more likely to become useful as an adjunct to DCE for deeper investigations of tumor characteristics in multimodality examinations. It would also be possible to examine other modality combinations, such as taking a PET ROI definition to derive quantitative DCE markers from a suitable protocol.

Benefits of a data-driven, semi-automated GMM–PET approach include reduced radiologist workload, faster segmentation processes, and reduced interobserver variability [47]. The major advantage of PET/MRI systems, exploited here, is intrinsic registration of simultaneously acquired PET and MR data, giving voxel-level correspondence, without additional registration processes that generally involve user

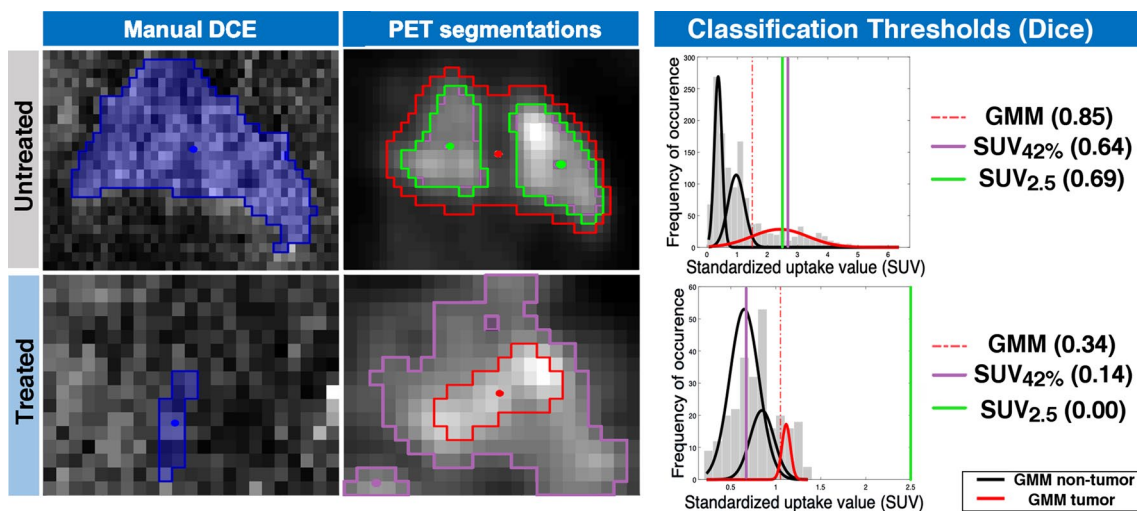
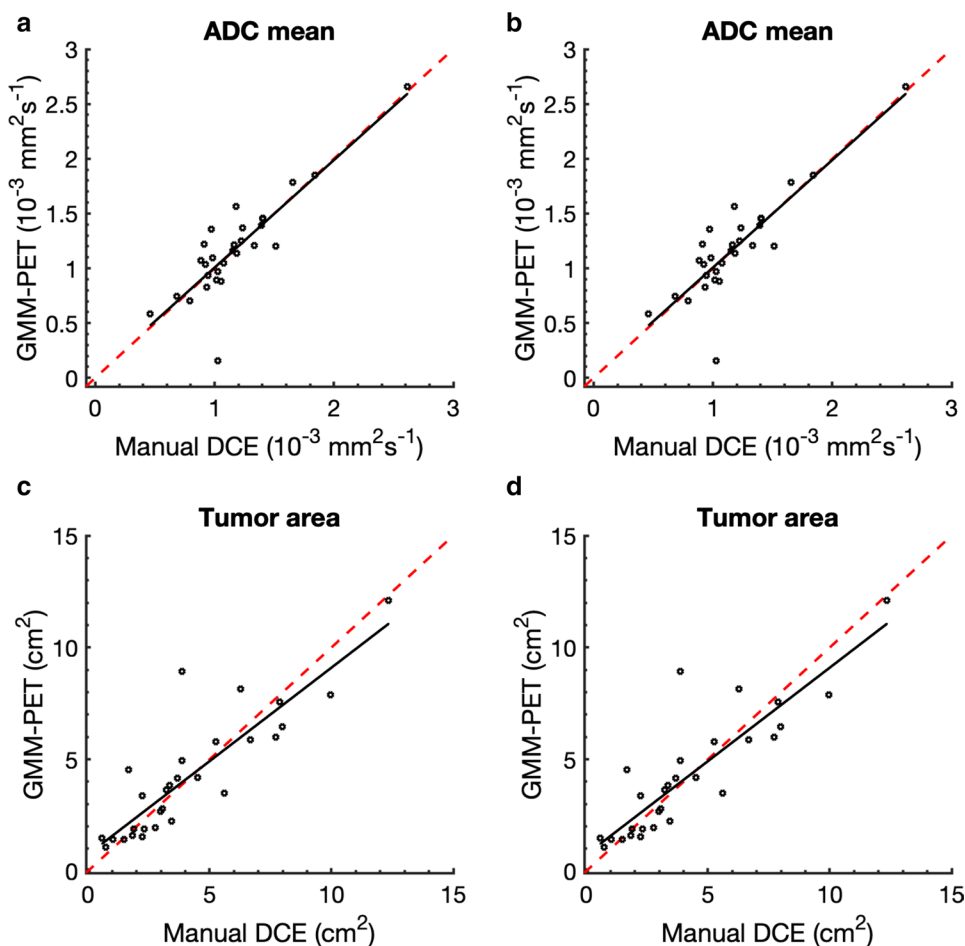


Fig. 5 Change in GMM–PET, $SUV_{42\%}$, and $SUV_{2.5}$ for an untreated (top row) and treated (bottom row) lesion from an illustrative patient, with corresponding histogram displays of SUV signal intensity and Dice similarity coefficient. GMM–PET is able to properly identify the whole tumor tissue of the heterogeneous untreated lesion, while

$SUV_{2.5}$ and $SUV_{42\%}$ give lower estimates. As SUV is reduced through treatment, $SUV_{2.5}$ cannot classify any tumor tissue, $SUV_{42\%}$ overestimates tumor area relative to DCE, while GMM–PET remains stable. Using DCE as a tumor definition standard becomes problematic when contrast leakage is reduced through treatment

Fig. 6 Relationship between the resulting metrics from manual DCE and GMM–PET for **a** ADC mean for untreated lesions ($r=0.866$, $p<0.001$) and **b** treated lesions ($r=0.895$, $p<0.001$) and **c** tumor area from untreated ($r=0.870$, $p<0.0001$) and **d** treated ($r=0.928$, $p<0.001$) lesions. Red identity lines included show that area from GMM–PET is slightly smaller than from manual DCE



input or additional processing [48]. Further optimization of GMM–PET might include automated data-driven initialization and number of classes, with the goal of entirely eliminating manual input [49]. We illustrate the influence of lesion-region initialization, for which automation could be developed, in Appendix (Fig. 8). The number of voxels within the lesion is also a factor that will affect the performance of the GMM method, and is thus intrinsically tied to image resolution, and will ultimately limit use in smaller lesions.

One limitation to this methodology is that increased tracer uptake is less specific to disease than manual DCE [50], potentially introducing errors in the GMM–PET classification if not accounted for by appropriate selection of tumor-containing region. In addition, while the analysis in this study explicitly did not consider patient response from the clinical viewpoint, as well as other clinically relevant data such as hormone status, the variety of treatments and responses provided a suitably large range of situations in which to test the GMM–PET methodology, including cases where response included an almost total loss of detectable disease.

Conclusion

The potential implications of improved imaging technology in breast cancer are large, and PET/MRI is a unique tool to investigate links between increased metabolism (PET), perfusion (DCE), and decreased diffusion (DWI), without additional scan time or registration errors. The current findings show that PET/MRI, using a semi-automated GMM segmentation strategy, yields tumor area and mean ADC value estimates that can replicate today's gold standard of tumor definition of manual DCE from MRI. Furthermore, the GMM–PET method also captures tumor changes associated with response to neoadjuvant chemotherapy, which supplements today's gold standard which is manual DCE in the neoadjuvant setting. The potential benefits include a broader assessment of morphological and metabolic changes to guide clinical decisions regarding tumor operability, and thus to ensure a high probability of complete tumor regression, and subsequent cancer cure.

Acknowledgements We acknowledge support from the liaison Committee between the Central Norway Regional Health Authority and

the Norwegian University of Science and Technology (Project nr. 90065000).

Author contributions MMAA: acquisition of data, analysis and interpretation of data, drafting of manuscript, and critical revision. PEG: study conception and design, acquisition of data, and critical revision. TES: study conception and design, acquisition of data, and critical revision. RH: acquisition of data and critical revision. HPE: study conception and design, acquisition of data, and critical revision. CD: analysis and interpretation of data and critical revision. AØ: analysis and interpretation of data, critical revision. SL: study conception and design, acquisition of data, and critical revision. TFB: study conception and design, analysis and interpretation of data, and critical revision. NPJ: study conception and design, acquisition of data, analysis and interpretation of data, drafting of manuscript, and critical revision.

Funding We acknowledge support from the liaison Committee between the Central Norway Regional Health Authority and the Norwegian University of Science and Technology.

Compliance with ethical standards

Conflict of interest None of the authors of this work have any conflict of interest to declare

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee (Regional Committee for Medical and Health Research Ethics (REC) in western Norway, identifier 2015/1493) and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Appendix

See Figs. 7 and 8.

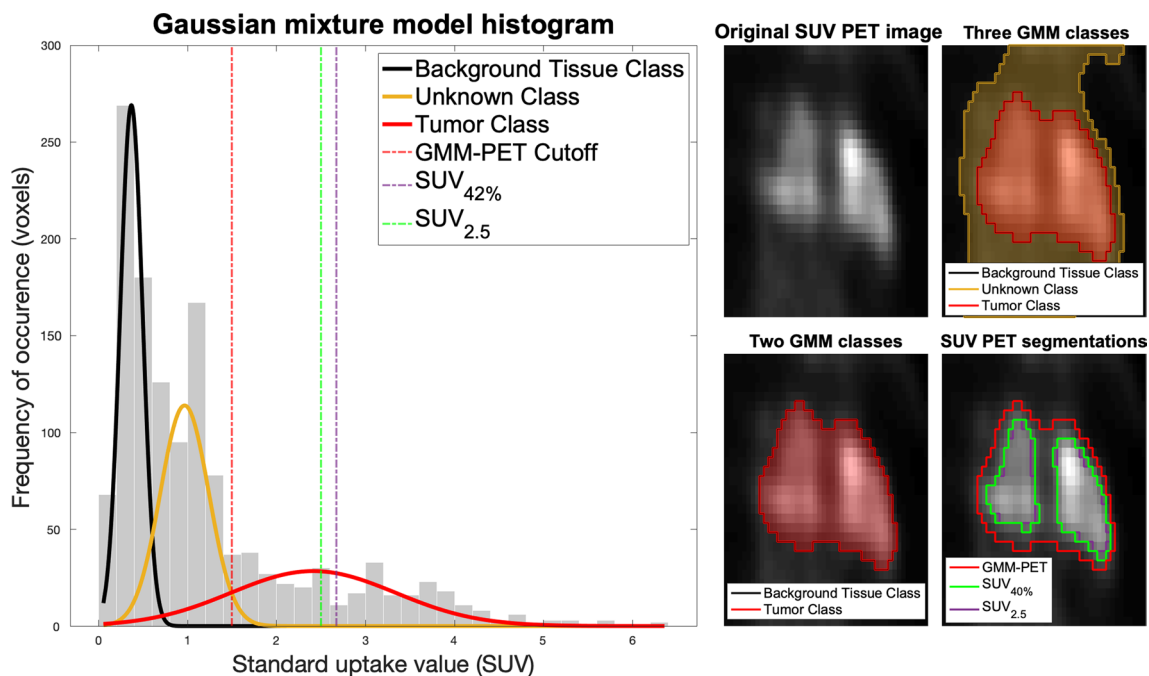
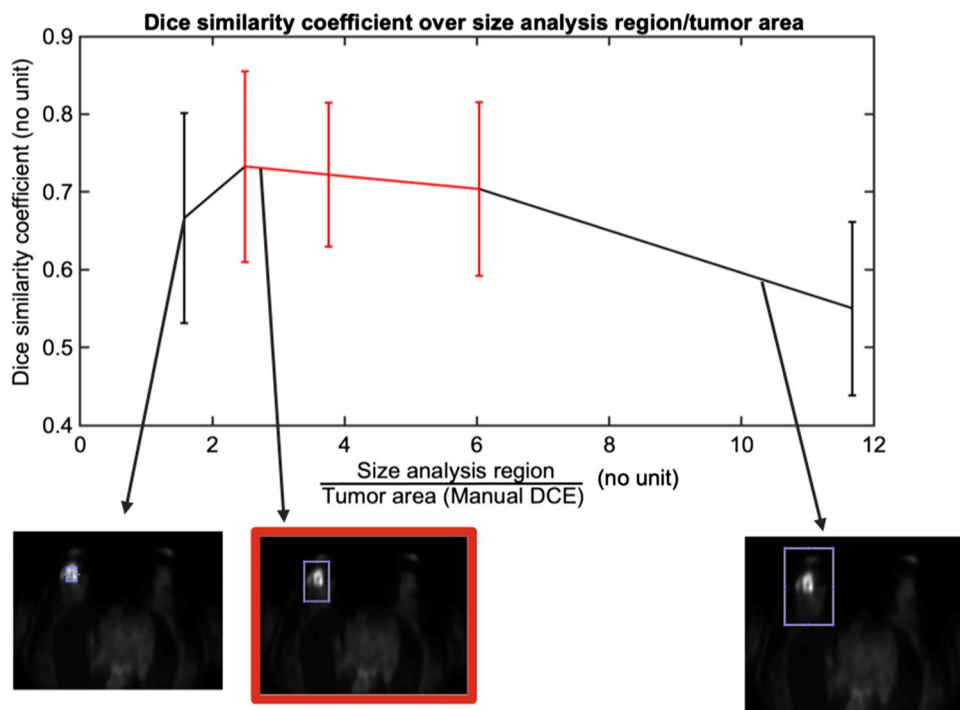


Fig. 7 Schematic illustration of the semi-automated Gaussian mixture model (GMM) segmentation performance, and the thresholding segmentation methods $SUV_{42\%}$ and $SUV_{2.5}$. **a** Histogram display of the three Gaussian distributions, together with $SUV_{42\%}$ and $SUV_{2.5}$ thresholds. **b** Cropped original PET image (upper), and resulting areas from GMM segmentation algorithm (lower) with three classes:

tumor (red), unknown (yellow), and background tissue (black) class. **c** GMM segmentation (red), $SUV_{42\%}$ (blue), and $SUV_{2.5}$ (green) ROIs overlaid on original PET image. For this example, $SUV_{42\%}$ is similar to GMM, but $SUV_{2.5}$ provides a much lower estimate of tumor volume

Fig. 8 Dice similarity coefficient (cohort mean \pm std) for Gaussian mixture model ROIs compared against manual dynamic contrast-enhanced (DCE) ROI. Indicated in red, the best results were given when the analysis regions were cropped at approximately 2–6 times the tumor area, and this was used in the study. This sensitivity would need to be accounted for in an automated cropping routine, as well as some sort of quality metric on the resulting ROI



References

1. Ferlay J EM, Lam F, Colombet M, Mery L, Piñeros M, Znaor A, Soerjomataram I, Bray F (2018) Global cancer observatory: cancer today. <https://gco.iarc.fr/today>. Accessed 29 Dec 2018
2. Cancer Registry of Norway (2018) Cancer in Norway 2017—cancer incidence, mortality, survival and prevalence in Norway. Accessed 10 Dec 2018
3. Cortazar P, Zhang L, Untch M, Mehta K, Costantino JP, Wolmark N, Bonnefoi H, Cameron D, Gianni L, Valagussa P, Swain SM, Prowell T, Loibl S, Wickerham DL, Bogaerts J, Baselga J, Perou C, Blumenthal G, Blohmer J, Mamounas EP, Bergh J, Semiglazov V, Justice R, Eidtmann H, Paik S, Piccart M, Sridhara R, Fasching PA, Slaets L, Tang S, Gerber B, Geyer CE Jr, Pazdur R, Ditsch N, Rastogi P, Eiermann W, von Minckwitz G (2014) Pathological complete response and long-term clinical benefit in breast cancer: the CTNeoBC pooled analysis. *Lancet* 384(9938):164–172
4. Thomas E, Holmes FA, Smith TL, Buzdar AU, Frye DK, Fraschini G, Singletary SE, Theriault RL, McNeese MD, Ames F, Walters R, Hortobagyi GN (2004) The use of alternate, non-cross-resistant adjuvant chemotherapy on the basis of pathologic response to a neoadjuvant doxorubicin-based regimen in women with operable breast cancer: long-term results from a prospective randomized trial. *J Clin Oncol* 22(12):2294–2302
5. Graham LJ, Shupe MP, Schneble EJ, Flynt FL, Clemenshaw MN, Kirkpatrick AD, Gallagher C, Nissan A, Henry L, Stojadinovic A, Peoples GE, Shumway NM (2014) Current approaches and challenges in monitoring treatment responses in breast cancer. *J Cancer* 5(1):58–68
6. Chu W, Jin W, Liu D, Wang J, Geng C, Chen L, Huang X (2018) Diffusion-weighted imaging in identifying breast cancer pathological response to neoadjuvant chemotherapy: a meta-analysis. *Oncotarget* 9(6):7088–7100
7. Partridge SC, Zhang Z, Newitt DC, Gibbs JE, Chenevert TL, Rosen MA, Bolan PJ, Marques HS, Romanoff J, Cimino L, Joe BN, Umphrey HR, Ojeda-Fournier H, Dogan B, Oh K, Abe H, Drukeinis JS, Esserman LJ, Hylton NM (2018) Diffusion-weighted MRI findings predict pathologic response in neoadjuvant treatment of breast cancer: the ACRIN 6698 multi-center trial. *Radiology* 289(3):618–627
8. Pickles MD, Gibbs P, Lowry M, Turnbull LW (2006) Diffusion changes precede size reduction in neoadjuvant treatment of breast cancer. *Magn Reson Imaging* 24(7):843–847
9. Nilsen L, Fangberget A, Geier O, Olsen DR, Seierstad T (2010) Diffusion-weighted magnetic resonance imaging for pretreatment prediction and monitoring of treatment response of patients with locally advanced breast cancer undergoing neoadjuvant chemotherapy. *Acta Oncol (Stockholm, Sweden)* 49(3):354–360
10. Sharma U, Danishad KK, Seenu V, Jagannathan NR (2009) Longitudinal study of the assessment by MRI and diffusion-weighted imaging of tumor response in patients with locally advanced breast cancer undergoing neoadjuvant chemotherapy. *NMR Biomed* 22(1):104–113
11. Charles-Edwards EM, deSouza NM (2006) Diffusion-weighted magnetic resonance imaging and its application to cancer. *Cancer Imaging* 6(1):135–143
12. Sinha S, Lucas-Quesada FA, Sinha U, DeBruhl N, Bassett LW (2002) In vivo diffusion-weighted MRI of the breast: potential for lesion characterization. *J Magn Reson Imaging* 15(6):693–704
13. Guo Y, Cai Y-Q, Cai Z-L, Gao Y-G, An N-Y, Ma L, Mahankali S, Gao J-H (2002) Differentiation of clinically benign and malignant breast lesions using diffusion-weighted imaging. *J Magn Reson Imaging* 16(2):172–178
14. Bickel H, Pinker K, Polanec S, Magometschnigg H, Wengert G, Spick C, Bogner W, Bago-Horvath Z, Helbich TH, Baltzer P (2017) Diffusion-weighted imaging of breast lesions: region-of-interest placement and different ADC parameters influence apparent diffusion coefficient values. *Eur Radiol* 27(5):1883–1892
15. Kim TH, Yoon JK, Kang DK, Lee SJ, Jung YS, Yim H, An YS (2015) Correlation between F-18 fluorodeoxyglucose positron emission tomography metabolic parameters and dynamic contrast-enhanced MRI-derived perfusion data in patients with invasive ductal breast carcinoma. *Ann Surg Oncol* 22(12):3866–3872

16. Plecha DM, Faulhaber P (2017) PET/MRI of the breast. *Eur J Radiol* 94:A26–A34
17. Hanahan D, Weinberg RA (2011) Hallmarks of cancer: the next generation. *Cell* 144(5):646–674
18. Nakajo M, Kajiyama Y, Kaneko T, Kaneko Y, Takasaki T, Tani A, Ueno M, Koriyama C, Nakajo M (2010) FDG PET/CT and diffusion-weighted imaging for breast cancer: prognostic value of maximum standardized uptake values and apparent diffusion coefficient values of the primary lesion. *Eur J Nucl Med Mol Imaging* 37(11):2011–2020
19. Byun BH, Noh WC, Lim I, Lee SS, Cho AR, Park JA, Kim KM, Kim HA, Kim EK, Kim BI, Choi CW, Lim SM (2013) A new method for apparent diffusion coefficient measurement using sequential (18)F-FDG PET and MRI: correlation with histological grade of invasive ductal carcinoma of the breast. *Ann Nucl Med* 27(8):720–728
20. Kitajima K, Yamano T, Fukushima K, Miyoshi Y, Hirota S, Kawana Y, Miya M, Doi H, Yamakado K, Hirota S (2016) Correlation of the SUVmax of FDG-PET and ADC values of diffusion-weighted MR imaging with pathologic prognostic factors in breast carcinoma. *Eur J Radiol* 85(5):943–949
21. Daisne JF, Duprez T, Weynand B, Lonneux M, Hamoir M, Reyckler H, Gregoire V (2004) Tumor volume in pharyngolaryngeal squamous cell carcinoma: comparison at CT, MR imaging, and FDG PET and validation with surgical specimen. *Radiology* 233(1):93–100
22. Im HJ, Bradshaw T, Solaiyappan M, Cho SY (2018) Current methods to define metabolic tumor volume in positron emission tomography: which one is better? *Nucl Med Mol Imaging* 52(1):5–15
23. Schelling M, Avril N, Nahrig J, Kuhn W, Romer W, Sattler D, Werner M, Dose J, Janicke F, Graeff H, Schwaiger M (2000) Positron emission tomography using [(18)F]Fluorodeoxyglucose for monitoring primary chemotherapy in breast cancer. *J Clin Oncol* 18(8):1689–1695
24. Holland D, Kuperman JM, Dale AM (2010) Efficient correction of inhomogeneous static magnetic field-induced distortion in Echo Planar Imaging. *Neuroimage* 50(1):175–183
25. Martinez-Moller A, Souvatzoglou M, Delso G, Bunschuh RA, Chef d'hotel C, Ziegler SI, Navab N, Schwaiger M, Nekolla SG (2009) Tissue classification as a potential approach for attenuation correction in whole-body PET/MRI: evaluation with PET/CT data. *J Nucl Med* 50(4):520–526
26. Klein S, Staring M, Murphy K, Viergever MA, Pluim JP (2010) Elastix: a toolbox for intensity-based medical image registration. *IEEE Trans Med Imaging* 29(1):196–205
27. Drzezga A, Souvatzoglou M, Eiber M, Beer AJ, Furst S, Martinez-Moller A, Nekolla SG, Ziegler S, Ganter C, Rummeny EJ, Schwaiger M (2012) First clinical experience with integrated whole-body PET/MR: comparison to PET/CT in patients with oncologic diagnoses. *J Nucl Med* 53(6):845–855
28. Geoff McLachlan DP (2000) Finite mixture models. Wiley series in probability and statistics. Applied probability and statistics. Wiley, New York
29. Dice LR (1945) Measures of the amount of ecologic association between species. *Ecology* 26(3):297–302
30. Benjamini Y, Hochberg Y (2000) On the adaptive control of the false discovery rate in multiple testing with independent statistics. *J Educ Behav Stat* 25(1):60–83
31. Nogueira L, Brandao S, Matos E, Nunes RG, Ferreira HA, Loureiro J, Ramos I (2015) Region of interest demarcation for quantification of the apparent diffusion coefficient in breast lesions and its interobserver variability. *Diagn Interv Radiol* 21(2):123–127
32. Nonomura Y, Yasumoto M, Yoshimura R, Haraguchi K, Ito S, Akashi T, Ohashi I (2001) Relationship between bone marrow cellularity and apparent diffusion coefficient. *J Magn Reson Imaging* 13(5):757–760
33. Yoshikawa MI, Ohsumi S, Sugata S, Kataoka M, Takashima S, Mochizuki T, Ikura H, Imai Y (2008) Relation between cancer cellularity and apparent diffusion coefficient values using diffusion-weighted magnetic resonance imaging in breast cancer. *Radiat Med* 26(4):222–226
34. Ito K, Kato T, Ohta T, Tadokoro M, Yamada T, Ikeda M, Nishino M, Ishigaki T, Gambhir S (1996) Fluorine-18 fluoro-2-deoxyglucose positron emission tomography in recurrent rectal cancer: relation to tumour size and cellularity. *Eur J Nucl Med* 23(10):1372–1377
35. Higashi T, Tamaki N, Torizuka T, Nakamoto Y, Sakahara H, Kimura T, Honda T, Inokuma T, Katsushima S, Ohshio G, Imamura M, Konishi J (1998) FDG uptake, GLUT-1 glucose transporter and cellularity in human pancreatic tumors. *J Nucl Med* 39(10):1727–1735
36. Choi BB, Kim SH, Kang BJ, Lee JH, Song BJ, Jeong SH, Yim HW (2012) Diffusion-weighted imaging and FDG PET/CT: predicting the prognoses with apparent diffusion coefficient values and maximum standardized uptake values in patients with invasive ductal carcinoma. *World J Surg Oncol* 10:126
37. Inglese M, Cavaliere C, Monti S, Forte E, Incoronato M, Nicolai E, Salvatore M, Aiello M (2019) A multi-parametric PET/MRI study of breast cancer: evaluation of DCE-MRI pharmacokinetic models and correlation with diffusion and functional parameters. *NMR Biomed* 32(1):e4026
38. Muz B, de la Puente P, Azab F, Azab AK (2015) The role of hypoxia in cancer progression, angiogenesis, metastasis, and resistance to therapy. *Hypoxia (Auckl)* 3:83–92
39. Daisne JF, Duprez T, Weynand B (2004) Tumor volume in pharyngolaryngeal squamous cell carcinoma: comparison at CT, MR imaging, and FDG PET and validation with surgical specimen. *Radiology* 233:93
40. Sridhar P, Mercier G, Tan J, Truong MT, Daly B, Subramaniam RM (2014) FDG PET metabolic tumor volume segmentation and pathologic volume of primary human solid tumors. *Am J Roentgenol* 202(5):1114–1119
41. Burger IA, Vargas HA, Apte A, Beattie BJ, Humm JL, Gonen M, Larson SM, Ross Schmidlein C (2014) PET quantification with a histogram derived total activity metric: superior quantitative consistency compared to total lesion glycolysis with absolute or relative SUV thresholds in phantoms and lung cancer patients. *Nucl Med Biol* 41(5):410–418
42. Burger IA, Casanova R, Steiger S, Husmann L, Stolzmann P, Huellner MW, Curioni A, Hillinger S, Schmidlein CR, Soltermann A (2016) 18F-FDG PET/CT of non-small cell lung carcinoma under neoadjuvant chemotherapy: background-based adaptive-volume metrics outperform TLG and MTV in predicting histopathologic response. *J Nucl Med* 57(6):849–854
43. Werner-Wasik M, Nelson AD, Choi W (2012) What is the best way to contour lung tumors on PET scans? Multiobserver validation of a gradient-based method using a NSCLC digital PET phantom. *Int J Radiat Oncol Biol Phys* 82:1164
44. Pinker K, Helbich TH, Morris EA (2017) The potential of multiparametric MRI of the breast. *Br J Radiol* 90(1069):20160715
45. Chen JH, Feig B, Agrawal G, Yu H, Carpenter PM, Mehta RS, Nalcioglu O, Su MY (2008) MRI evaluation of pathologically complete response and residual tumors in breast cancer after neoadjuvant chemotherapy. *Cancer* 112(1):17–26
46. Gulani V, Calamante F, Shellock FG, Kanal E, Reeder SB (2017) Gadolinium deposition in the brain: summary of evidence and recommendations. *Lancet Neurol* 16(7):564–570
47. Kim HR, Jung HK, Ko KH, Kim SJ, Lee KS (2016) Mammography, US, and MRI for preoperative prediction of extensive

- intraductal component of invasive breast cancer: interobserver variability and performances. *Clin Breast Cancer* 16(4):305–311
48. Riola-Parada C, Garcia-Canamaque L, Perez-Duenas V, Garcera-Tafur M, Carreras-Delgado JL (2016) Simultaneous PET/MRI vs PET/CT in oncology. A systematic review. *Rev Esp Med Nucl Imagen Mol* 35(5):306–312
49. Aristophanous M, Penney BC, Martel MK, Pelizzari CA (2007) A Gaussian mixture model for definition of lung tumor volumes in positron emission tomography. *Med Phys* 34(11):4223–4235
50. Adejolu M, Huo L, Rohren E, Santiago L, Yang WT (2012) False-positive lesions mimicking breast cancer on FDG PET and PET/CT. *Am J Roentgenol* 198(3):W304–W314

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.