

Normal model construction for statistical image analysis of torso FDG-PET images based on anatomical standardization by CT images from FDG-PET/CT devices

Kenshiro Takeda¹ · Takeshi Hara¹  · Xiangrong Zhou¹ · Tetsuro Katafuchi² · Masaya Kato³ · Satoshi Ito³ · Keiichi Ishihara⁴ · Shinichiro Kumita⁴ · Hiroshi Fujita¹

Received: 17 August 2016 / Accepted: 14 January 2017 / Published online: 6 February 2017
© CARS 2017

Abstract

Purpose A better understanding of the standardized uptake value (SUV) ranges of fludeoxyglucose positron emission tomography (FDG-PET) is crucial for radiologists. We have developed a statistical image analysis method for FDG-PET imaging of the torso, based on comparisons with normal data. The purpose of this study was to verify the accuracy of the normal model and usefulness of the statistical image analysis method by using typical cancer cases in the liver, lungs, and abdomen.

Methods Our study and the data collection (49 normal and 34 abnormal cases, in terms of PET/CT findings) were approved by the institutional review board. Our scheme consisted of the following steps: (1) normal model construction, (2) anatomical standardization of patient images, and (3) Z-score calculation to show the results of the statistical image analysis. To validate the Z-score index, we sampled 3603 and 1270 voxels in normal organs and abnormal regions, respectively, from the liver, lungs, and the abdomen. We then obtained the SUV and Z-score for each region. A receiver operating characteristics (ROC) analysis-based method was performed to evaluate the discrimination performances of the SUV and Z-score.

Results The discrimination performances of the SUV and Z-score for the objective regions of interest (ROIs) were evaluated by the areas under the ROC curves (AUCs). As a result of the ROC analysis and statistical tests, all AUCs were found to be larger than 0.98. When the ROIs in the objective regions were combined, the mean AUCs of the Z-score and SUV were 0.99 and 0.98, respectively, the difference being statistically significant ($p < 0.001$).

Conclusions The results suggested the possibility of applying a quantitative image reading method for torso FDG-PET imaging. Furthermore, a combination of the SUV and Z-score may provide increased accuracy of the determination methods, such as computer-aided detection and diagnosis.

Keywords SUV · Z-score · Torso FDG-PET · ROC · PET/CT

Introduction

Globally, cancer mortality rates are on the rise. To reduce the number of deaths caused by cancer, it is important to detect and appropriately treat cancer in its early stages. Molecular imaging examination using fludeoxyglucose positron emission tomography (FDG-PET) is often employed for the detection, diagnosis and staging of cancer, and for prediction of the prognosis of cancer. For the imaging-based diagnosis, the standardized uptake value (SUV) of FDG-PET is used for evaluating the regional activity of glucose metabolism. SUV can detect regions with high glucose metabolism, and these regions can be considered as areas of abnormal accumulation [1].

In analysis of brain function, statistical image analysis during brain function analysis has been widely accepted in the field of nuclear medicine. The basis of the statis-

✉ Takeshi Hara
takeshi.hara@mac.com

¹ Department of Intelligent Image Information, Graduate School of Medicine, Gifu University, 1-1 Yanagido, Gifu, Gifu 501-1194, Japan

² Faculty of Health Science, Gifu University of Medical Science, Seki, Japan

³ Department of Radiology, Daiyukai General Hospital, Ichinomiya, Japan

⁴ Department of Radiology, Nippon Medical School, Tokyo, Japan

tical image analysis is the voxel-by-voxel comparison of regional activities, with a normalized database obtained after standardization of the shapes of the normal cases into a normal model. The comparison is performed using computerized tools to obtain a quantitative index for the diagnosis. For brain function analysis, computerized packages such as 3D-SSP and SPM are used to quantify the regional blood flow in brain single-photon emission computed tomography (SPECT) images of the brain. Their usefulness in the diagnosis of dementia has been recognized [2–7]. However, their application in the imaging of the torso has not yet been established. Therefore, it is natural to extend the idea of statistical image analysis to this region. We have embodied this idea by using FDG-PET images in an earlier study [8]. Use of appropriate statistical analysis in interpretation of PET/CT images will enhance the detection and discrimination performances of human readers and computerized schemes [9]. The normal model was constructed using FDG-PET, and not computed tomography (CT) images in the previous studies. In addition, the studies did not establish the accuracy of the normal model following anatomical standardization methods and did not utilize anatomical information from CT images.

The construction of a normal model is essential in statistical image analysis. The construction process consisted of location of an organ on PET images followed by a deformation process to fit images to the standard body shape. These two processes were performed for anatomical standardization of the statistical image analysis.

Normal model accuracy depends on the landmark location accuracy of the organs of the torso on PET images. We focused on the construction method of the normal model in this study because CT images could be with the PET images almost simultaneously, by using a PET/CT scanner at very low radiation doses. As a result, automated organ recognition methods using CT images were available for model construction.

The purpose of this study was to establish the accuracy of the organ localization method using CT images at a low radiation dose and resolutions and to verify the usefulness of the statistical index based on the statistical analysis method for torso FDG-PET images.

Materials and methods

Figure 1 shows an overview of our scheme. The scheme consisted of (1) normal model construction, (2) standardization of patient images, and (3) presentation of *Z*-score images resulting from the statistical comparison of patient images with the normal model. The normal model construction and standardization of patient images employed the same procedures of anatomical standardization to fit a standard body shape.

Anatomical standardization

The anatomical standardization process was used for the normal model construction from normal cases, and calculation of *Z*-score images from patients' cases. Patient cases were also standardized to compare with the normal model. In the anatomical standardization process, images were deformed to fit a standard model selected from the database based on the average weight of a standard Japanese body in this study. The process of anatomical standardization included organ detection using CT images and image deformation to fit the standard body shape.

Before organ detection, the coordinates of the CT images were converted to match the pixel size and field of view in PET images, by three-dimensional affine transformation using a linear interpolation method. Initial alignment of PET and CT images was performed by PET/CT scanner during the image analysis. In the organ detection step, the whole torso region was extracted by thresholding at -204HU in CT number and by using the connected component analysis technique to limit the detection area, because the background regions tend to have low CT values. After the connected component analysis process, the region with the maximum volume was considered as the torso region. Based on the volume of the torso region, three deformation steps of physique, organ, and body surface area were applied to the region by using three-dimensional affine transformation and non-rigid image deformation using thin plate spline (TPS) [10] processes.

First, the physique registration was performed based on a previously described method [8]. We applied the method to CT images by changing the process parameters experimentally. Six planes, i.e., the neck–shoulder plane, left arm–chest plane, right arm–chest plane, thigh–hip plane, anterior body plane, and posterior body plane were determined from the CT images of the torso region. Figure 2 shows examples of the six planes determined using CT images. The regions of the left and right arms and legs, and head were excluded during this process. Eight vertices determined as the crossing points of the planes were used to identify regions of the physique as rectangular shapes.

The same region of the physique marked by the six planes was also determined using PET images, using the previously described method [8]. The eight vertices with yellow points in Fig. 3 correspond to the eight vertices created by the six planes in Fig. 2 from the CT images.

Second, organ regions in the CT images were extracted as rectangular regions (bounding boxes) based on an automated organ extraction method [14, 15]. The smallest rectangular area that surrounded the organ region was determined from the CT images. Figure 4 shows an example of the organ extraction results. Nine rectangular regions (one each from the heart, left and right lungs, liver, pancreas, stomach, left

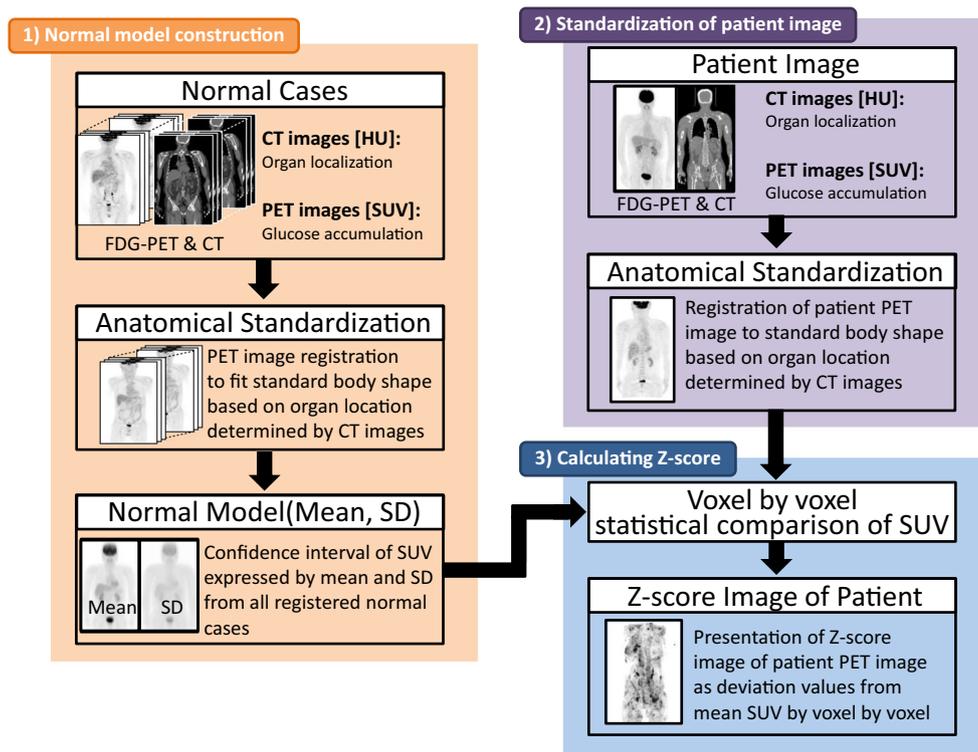


Fig. 1 Three steps of the statistical image analysis for torso FDG-PET/CT images

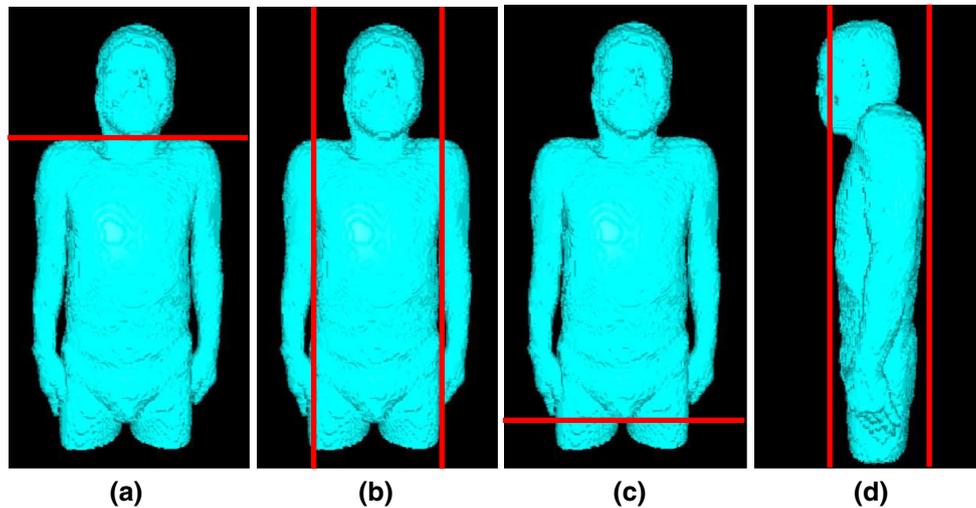


Fig. 2 The six determined planes (red lines) used for the physique registration in CT images. **a** Neck–shoulder plane, **b** left and right arms–chest planes, **c** thigh–hip plane, and **d** anterior and posterior planes

and right kidneys, and bladder) were extracted. Each region had eight vertices to include the organ regions. The locations were transferred to the PET images to determine the organ regions. Every side of the rectangular regions was divided to set landmarks deforming the regions. Table 1 shows the number of divisions and landmarks. Each region was divided into lattice regions with vertices that were landmarks for TPS deformation. Every region of the organs was deformed using

the TPS method to fit the standard body shape based on the landmarks.

Finally, the body surface area for the standard body shape was recorded, conforming to the region of the physique determined during the first step. The three-dimensional labeled region of the physique on CT images was converted into PET images using the nearest-neighbor interpolation and affine transformation techniques. The labeled image was divided

Fig. 3 Example of physique landmarks in a PET image

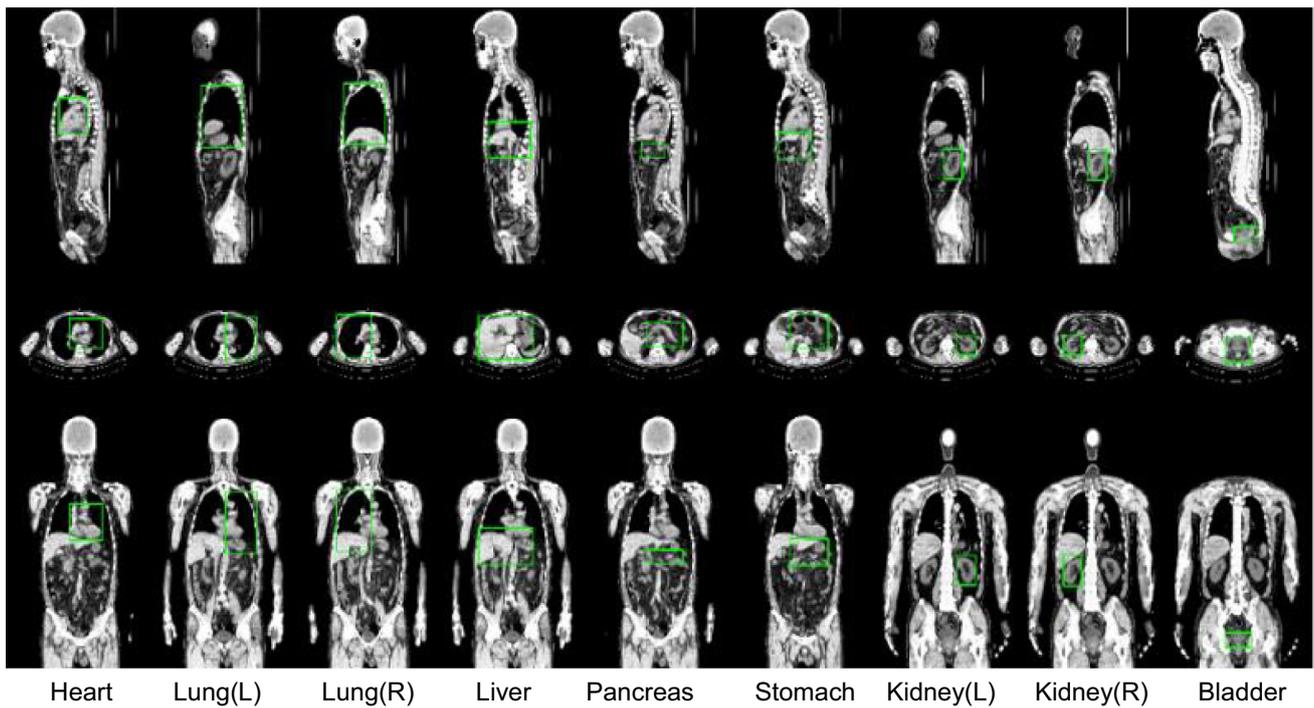
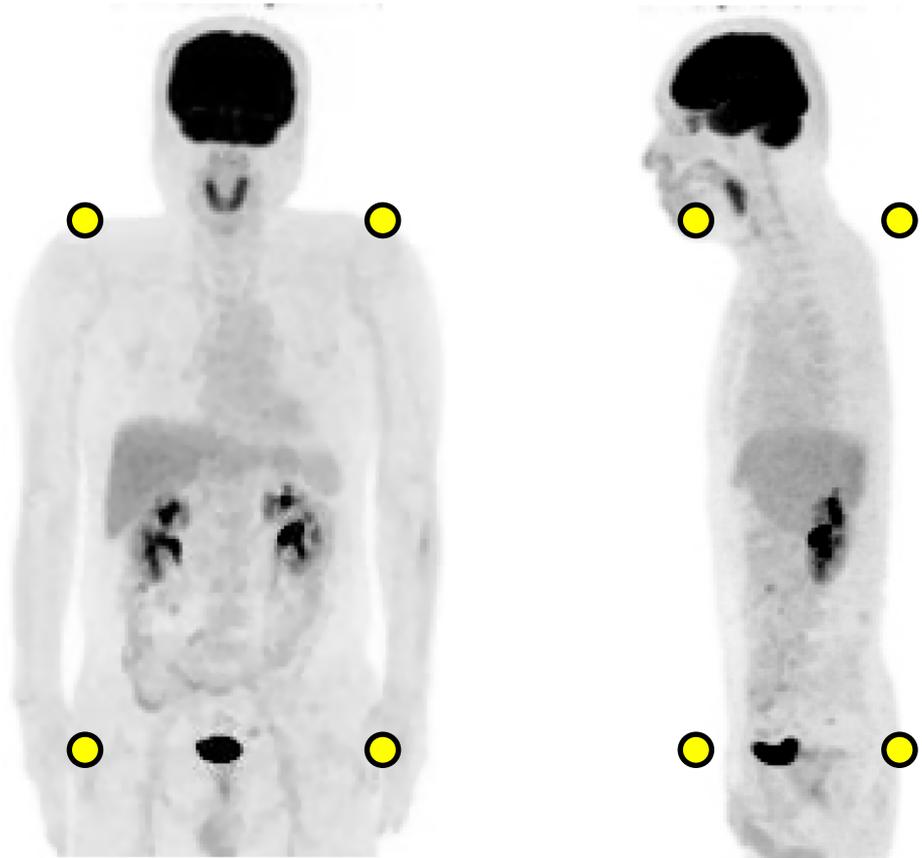


Fig. 4 Example of extraction results (*green boxes*) from the automatic organ detection

Table 1 The number of landmarks in each organ

Organs	# of Div. ^a	# of LMs ^b
Heart	4	64 (=4 × 4 × 4)
Liver	2	8 (=2 × 2 × 2)
Stomach	2	8 (=2 × 2 × 2)
Right kidney	3	27 (=3 × 3 × 3)
Left kidney	3	27 (=3 × 3 × 3)
Pancreas	3	27 (=3 × 3 × 3)
Bladder	3	27 (=3 × 3 × 3)
Right lung	3	27 (=3 × 3 × 3)
Left lung	3	27 (=3 × 3 × 3)

^a # of Div: # of division to set LMs on each side of rectangular region
^b LM landmark

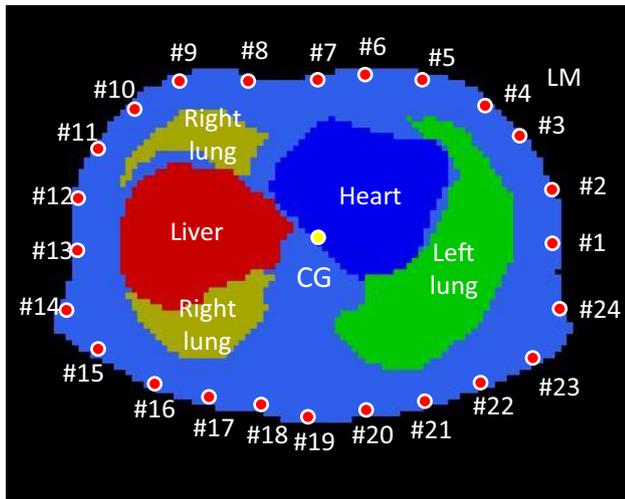


Fig. 5 Example of center of gravity (CG) and landmarks (LM) on body surface and organ names on a slice

into 28 slices axially. The center of gravity was obtained for each slice. The landmarks were set on the body surface at an angle of 15 degrees around the center of gravity on all of the slices. A total of 672 points were set as landmarks for the whole torso region. Figure 5 shows an example of the center of gravity and landmarks on the body surface. Image deformation using the TPS method was applied to fit the standard body shape.

Normal model construction

The normal model consisted of two distributions of means and standard deviations (SDs). The means and SDs were obtained by summarizing the normal cases that were deformed to fit the standard body based on the anatomical standardization process, so that each voxel in the normal model contained means and SDs, and that the voxel values at location (x, y, z) could be expressed by $\text{Mean}(x, y, z)$

and $\text{SD}(x, y, z)$. $\text{Mean}(x, y, z)$ and $\text{SD}(x, y, z)$ statistically expressed ranges of SUV at location (x, y, z) . The normal cases were selected from individuals undergoing a medical checkup with cancer screening using FDG. Figure 6 shows an example of the normal model. The normal model was constructed without accommodating for sex because no reproductive organs were extracted from CT images in this study.

Z-score calculation

The Z-score is a general index to show deviation from the mean and the SD that were determined in the normal model. Regional activities in the patient images were expressed as the Z-score after the patient’s images were anatomically standardized. The Z-score is often employed to show results of statistical image analysis.

The Z-score in this study was obtained by comparing the normal model voxel-by-voxel with the standardized target case. The Z-score at location (x, y, z) was defined as Eq. (1) when the patient images were deformed as $PDef(x, y, z)$ to fit the standard body based on the anatomical standardization process. $PDef(x, y, z)$ is defined as the SUV value at location (x, y, z) after image deformation. Figure 7 shows two examples of Z-score images in abnormal cases.

$$Z\text{-score}(x, y, z) = \frac{PDef(x, y, z) - \text{Mean}(x, y, z)}{\text{SD}(x, y, z)} \quad (1)$$

Database

We included 49 (M: 39, F: 10) and 34 (M: 5, F: 29) patients with normal and abnormal PET/CT findings, respectively, after approval by the institutional review board (#23-131, #28-114). The cases were classified as normal or abnormal based on the interpretation of the radiologists. All images were obtained using the same PET/CT scanner. Table 2 shows the image specifications of the PET/CT device employed. Table 3 shows data regarding the patient’s weights and ages. Data regarding the patient’s heights were not collected.

Statistical analysis

SUVs and Z-scores were compared statistically based on the receiver operating characteristics (ROC) analysis. R (ver. 3.11) and DBM-MRMC (ver. 2.1) were employed for the statistical tests, and p values less than 0.05 were considered to be statistically significant. Power analysis was also conducted [11]. G*Power (ver. 3.192) was employed for the power analysis [12, 13].

Fig. 6 Example of the normal model: **a** distributions of the mean, and **b** distributions of the standard deviation

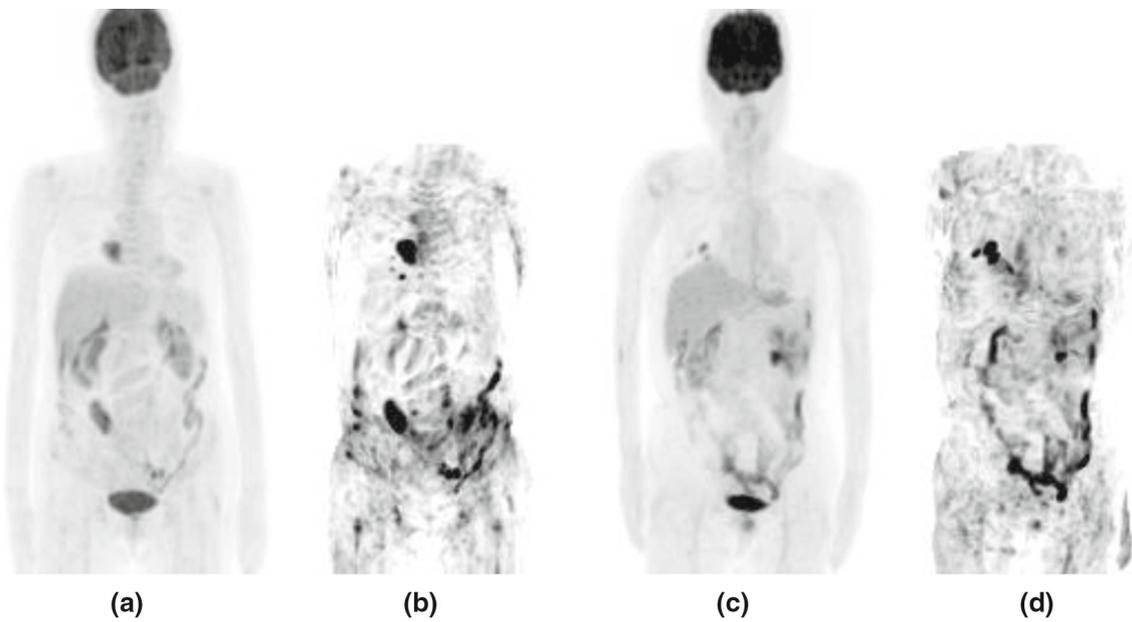
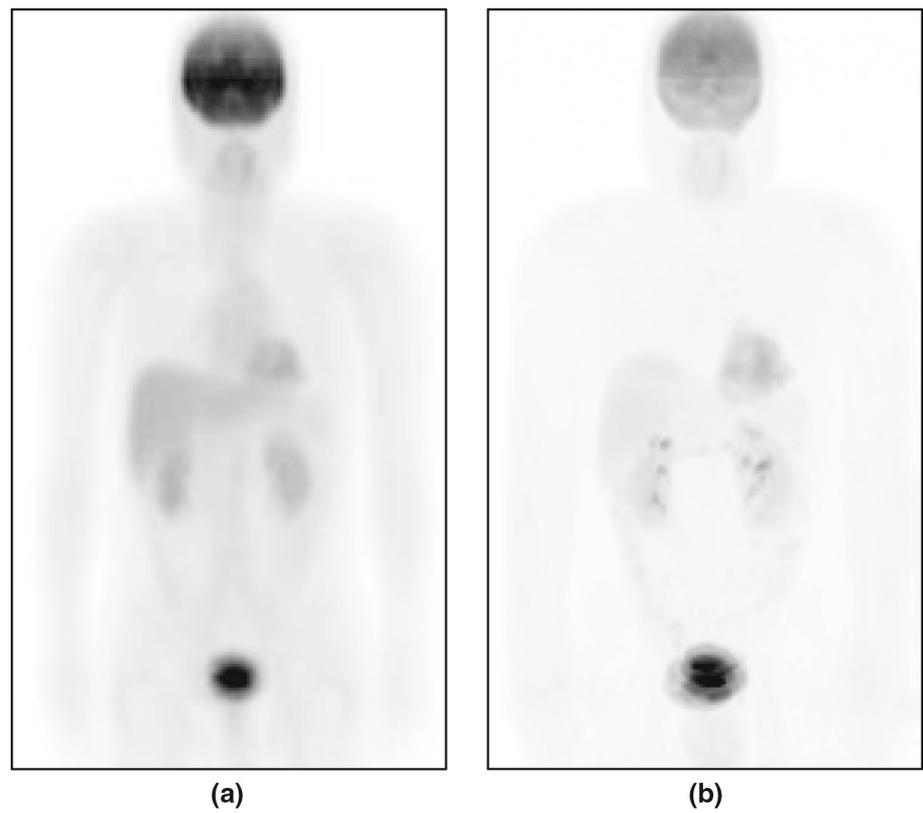


Fig. 7 Two examples of the Z-score image of FDG-PET scans: **a, c** original FDG-PET image, and **b, d** Z-score image

Table 2 Image specifications of the PET/CT imaging device employed

	PET	CT
Image matrix size	144 × 144 × (234–276)	512 × 512 × (188–221)
Axial image pixel size (mm)	4.00 × 4.00	1.15 × 1.15
Slice thickness (mm)	4.00	5.00

Table 3 The statistics of patient weights and ages

	Accumulation type	Mean	max	Min	SD	# of unknown
Weight	Normal	67.8	88.4	47.5	10.5	0
	Abnormal	51.7	77.5	36.7	9.3	0
Age	Normal	55.2	80	29	28.4	29
	Abnormal	63.3	85	17	13.1	0

SD standard deviation

Results

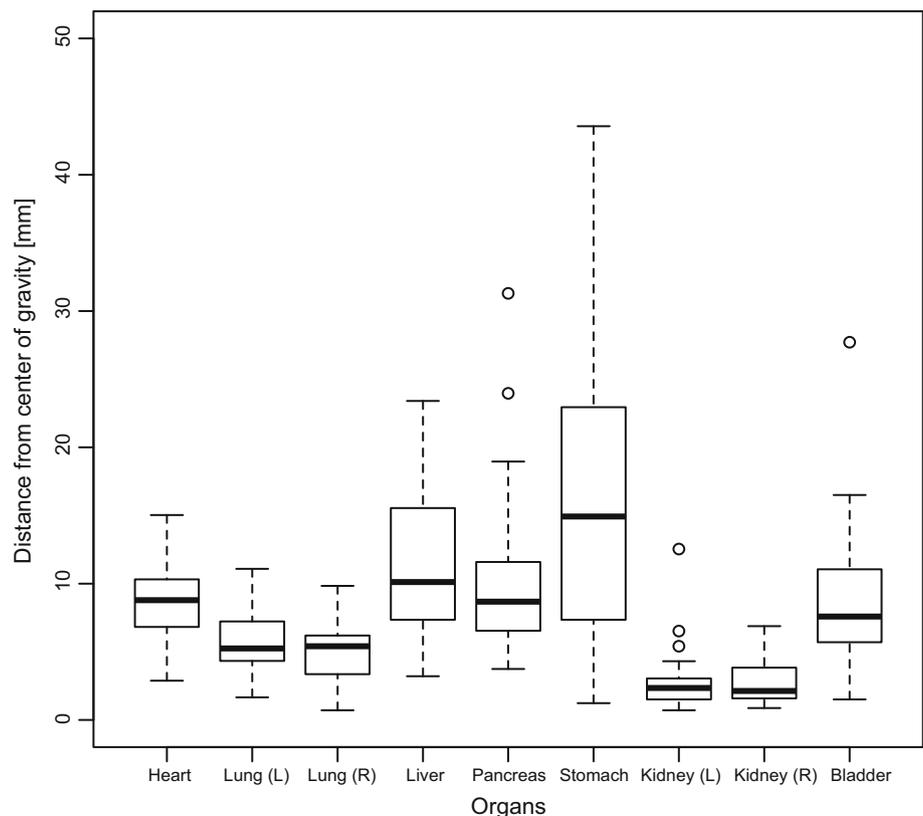
Performance of organ extraction

The centers of gravity (CG) of organs, obtained by automated extraction, were evaluated by comparing with those obtained by manual extraction. All normal cases ($n = 49$) were used for the evaluation. Figure 8 shows box plots of the distances between the CGs of automated results and gold standards obtained by manually segmented organ regions. The outliers are plotted separately as circles. The median in each distribution is shown as a thick bar in the box. The outliers are defined by three parameters: first quartile (Q1), third quartile (Q3), and interquartile range (IQR), which is defined as the difference between Q3 and Q1. The lower and upper limits to determine the outliers were defined as $Q1 - 1.5 IQR$ and $Q3 + 1.5 IQR$, respectively.

Performance of Z-score imaging

We evaluated the discrimination performance of the SUVs and Z-scores in normal and abnormal accumulations based on the ROC analysis and t -test. Figure 9 shows the target regions. Discrimination performance was also evaluated when the regions were combined. We sampled 3603 and 1270 voxels in normal organs and abnormal regions, respectively. Voxel sampling within each ROI was performed systematically at fixed intervals of 3 or 4 voxels in the three dimensions. The organs and regions were manually determined by the graph-cut method. While 3603 voxels were sampled from 25 normal cases, 1270 voxels were sampled from 34 abnormal cases. The sample sizes of each normal and abnormal region are presented in Table 4. The discrimination performance was evaluated by using the area under the ROC curve (AUC). Comparisons of the AUCs are shown in Table 5. All

Fig. 8 Results of automated organ extractions evaluated by the distance from the center of gravity (CG) in the gold standard to that in the determined organ region



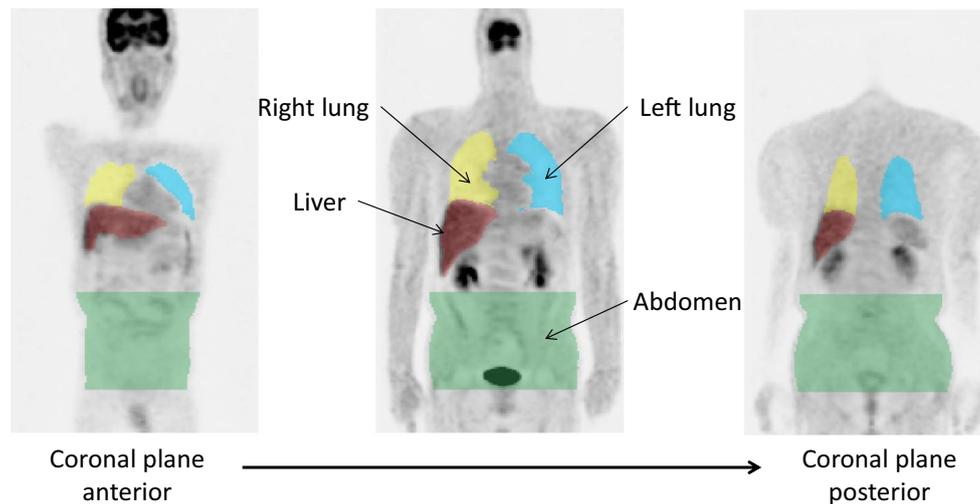


Fig. 9 Target regions in a PET image Right lung (yellow region), left lung (blue region), liver (red region), abdomen (green region)

Table 4 Means and SDs, and results of power analysis of the SUV and Z-score in normal and abnormal accumulations for regions of the liver, lung, and abdomen

Regions	Sample size	Accumulation type	SUV (mean, SD)	Power	ES	Z-score (mean, SD)	Power	ES
Liver	1201	Normal	1.90, 0.50	1.000	1.711	0.06, 0.99	1.000	1.704
	251	Abnormal	3.34, 1.08			6.51, 5.26		
Lung	1011	Normal	0.66, 0.45	1.000	2.488	-0.21, 1.18	1.000	2.558
	144	Abnormal	3.13, 1.33			9.86, 5.44		
Abdomen	1391	Normal	0.61, 0.54	1.000	2.648	-0.01, 1.68	1.000	1.876
	875	Abnormal	4.58, 2.05			14.61, 10.89		
Pooled	3603	Normal	1.06, 0.78	1.000	2.113	-0.04, 1.35	1.000	1.751
	1270	Abnormal	4.17, 1.93			12.47, 10.07		

Power: $1-\beta$ error probability of two-tailed t -test at α error probability is 0.05.

ES effect size

Table 5 Comparison of the area under the ROC curves and p values

Regions	# of case	Interval	SUV	Z-score	p value
Liver	7	4	0.987	0.997	<0.001
Lung	11	4	0.995	0.999	0.019
Abdomen	26	3	1.000	0.998	<0.001
Combined region	34	3 or 4 ^a	0.981	0.993	<0.001

^a Depends on regions of Liver, Lung, or Abdomen

the AUCs showed high scores (over 0.98). This implied that the discrimination performances of both SUVs and Z-scores were very good. In the liver and lung results, the AUCs of the Z-scores were slightly greater than those of the SUVs. There were statistically significant differences between all AUCs. Furthermore, the Z-scores were slightly higher than the SUVs and there was a statistically significant difference when the regions were combined.

Figure 10 shows the results of statistical comparison of normal accumulations with abnormal ones in terms of SUV and Z-score. Each mean was compared by t -test. There were statistically significant differences between all normal and abnormal accumulations. The mean and SD of the SUV and Z-score in normal and abnormal accumulations in each region are shown in Table 4. The results of power analysis are also shown in the table. The statistical power and effect size were obtained by using G*Power [12, 13].

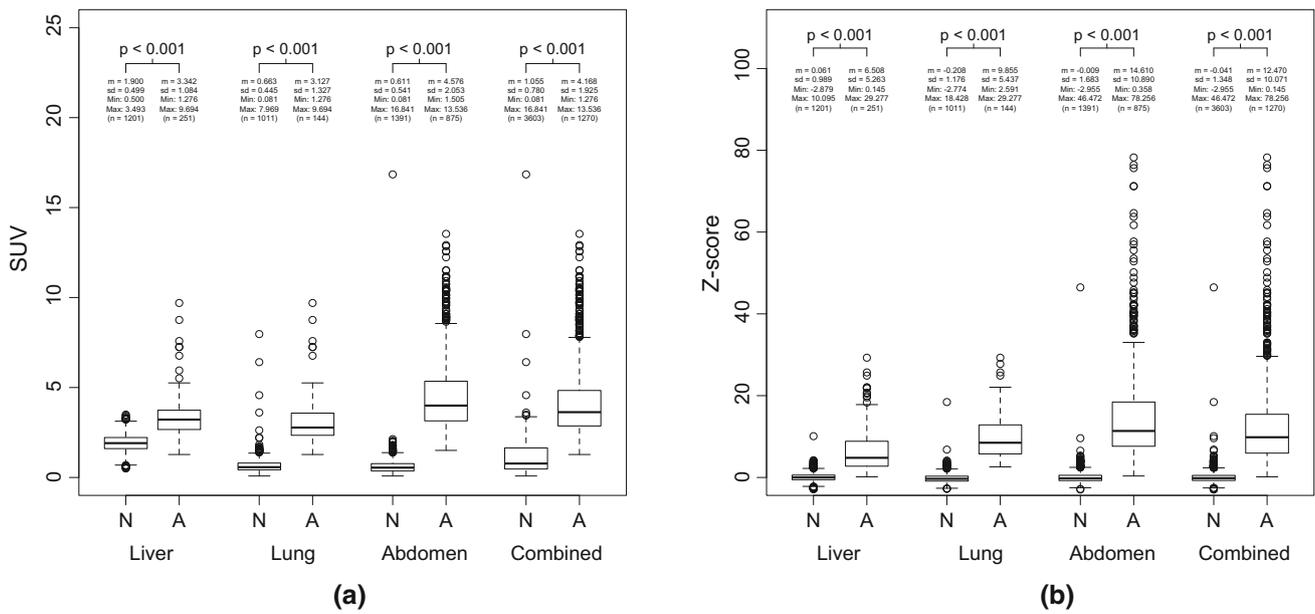


Fig. 10 Results of statistical comparisons of normal (N) accumulations with abnormal (A) ones in terms of SUV (a) and Z-score (b)

Table 6 The statistics of volume variance ratios in each organ

Organs	Abs. Mean (%)	Mean (%)	SD	Max (%)	Min (%)
Heart	16.80	-13.21	18.31	23.71	-82.40
Liver	18.27	-8.37	21.50	45.13	-52.83
Stomach	120.06	-118.27	121.89	35.41	-646.18
Right kidney	29.60	-28.87	20.79	17.05	-78.69
Left kidney	26.71	-26.71	17.72	-1.40	-75.10
Pancreas	105.07	-97.85	123.10	50.42	-745.07
Bladder	162.50	-160.30	96.94	53.77	-425.09
Right lung	35.03	-34.65	15.51	5.69	-64.12
Left lung	34.80	-34.04	18.78	13.24	-79.20

Abs. mean averaged absolute value of volume variance ratio, SD standard deviation

Discussions

Organ extraction

We extracted organ regions using an automated organ extraction method that was developed for CT images with high resolutions and regular radiation doses [14, 15]. By applying this method to the present study, we evaluated the performance of the automated organ extraction method when the CT images were acquired using PET/CT scanners with a lower resolution and lower radiation dose. To evaluate the results of the organ extraction, we used the distance between the CGs of the extracted results and gold standards. As shown in Fig. 8, most of the distances were between 0 and 20 mm except distances for large and shape-complex organs such as the liver, pancreas, and stomach. In comparison, in previous studies of extraction performance, errors were found to be

distributed from 5 to 20 mm [14, 15]. Table 6 shows the volume variance ratio for in each organ. These values express the ratio of misalignment between the gold standard and the results obtained by the automated method [14]. Lower mean values indicate good agreements between the values. Mean value in organs with convex shapes, such as the heart, liver, kidneys, pancreas, and lungs showed good agreement, while organs with complex shapes showed bad agreement. We found that it was difficult to estimate the agreements of elastic organs with the gold standard when bounding box regions were provided. Jaccard similarity coefficients of actual organ regions defined by the organ surface are required for precise estimation of agreement of organ boundaries. Further analysis of detailed performance related to the organ shape is required, after precisely segmenting the organ shapes from the images.

Table 7 Results of the Jarque–Bera tests

Organs	Jarque–Bera (%)
Heart	49.10
Liver	85.86
Stomach	83.51
Right kidney	56.77
Left kidney	64.33
Pancreas	90.01
Bladder	13.71
Right lung	64.97
Left lung	74.12

Normal model construction

We used our automated method based on a three-dimensional image deformation technique to obtain the normal model used for analyzing the scan images. However, the Z -score calculated in our method is highly dependent on the normal model. Therefore, we evaluated the normal model using the Jarque–Bera test to evaluate normality. Usefulness of a normal model can be considered to rise with an increase in the number of areas demonstrating a normal distribution. The result of Jarque–Bera test in each region is shown in Table 7. This table shows the ratio of pixels that meet the significance level of 5% in each organ. Although the ratio in the bladder is low because of individual differences, the ratios in other organs are relatively high. These tests imply that the normal model is appropriate and has good normality.

Z -score imaging

The discrimination performance of the Z -score was lower than that of the SUV in the abdomen, and there were statistically significant differences from the results of the t -test. As shown in Table 4, the SD of the Z -score in abnormal accumulations in the abdomen was higher than the other regions. The Z -scores in normal accumulations were around zero, which means that the regional activity was within the statistically normal ranges despite the presence of the organs.

The discrimination performance of the Z -score showed high accuracy in all regions, although the performances of the SUV were also quite high. Furthermore, the discrimination performance of the Z -score was slightly higher than that of the SUV in the combined regions. This can be attributed to the fact that the average SUV of each region is different from that of the other regions. The Z -score is likely to excel in the evaluation of normal and abnormal accumulations in the entire torso region.

Limitations

The normal model in this study was constructed without consideration of sex, because we could not recruit sufficient number of members of each sex. The Z -score in the abdominal regions may be affected when a gender-specific normal model is constructed using the extraction results of reproductive organs such as the uterus, ovaries, or prostate. The determination of the number of normal cases and the case selection from normal cases are significant in the results of the statistical image analysis.

Image deformation, mainly accomplished by the TPS technique, often affects voxel values in normal and abnormal cases. The Z -score of patient image includes the effect of changes in voxel values caused by interpolation methods during image deformation. High voxel values sometimes result in underestimation of accumulations.

The precision of the normal model mainly depends on the accuracy of the landmark locations. To evaluate the accuracy of the locations, we have to establish a large database of organ shapes in various normal and abnormal cases. The procedure includes graph-cut method to determine the organ surface. It is a lengthy procedure as it involves performing manual segmentations and will require complex verifications by physicians or radiologists. We are designing another database of PET/CT images with over 300 normal and 100 abnormal cases for our future investigations.

Conclusions

We developed an automated image registration scheme for torso FDG-PET/CT. The results of this study suggest the possibility of a quantitative method to interpret SUVs as indicators of FDG accumulation in each target region. Statistical image analysis of the torso FDG-PET/CT images might provide a new index not only for evaluating the accuracy of the interpretation, but also for image features in computer-aided detection and diagnosis.

Acknowledgements This research work was funded in part by Grant-in-Aid for Scientific Research on Innovative Areas (26108005) and in part by Grant-in-Aid for Scientific Research (C26330134), MEXT, Japan.

Compliance with ethical standards

Conflict of interest Takeshi Hara has received research grants from Nihon Medi-Physics Co. Ltd. Kenshiro Takeda, Xiangrong Zhou, Tetsuro Katafuchi, Masaya Kato, Satoshi Ito, Keiichi Ishihara, Shinichiro Kumita and Hiroshi Fujita declare that they have no conflict of interest.

Ethical approval All procedures performed in the studies involving human participants were in accordance with the ethical standards of the institutional and national research committees, and with the 1964 Helsinki declaration and its later amendments or comparable ethical

standards. The Institutional Review Board at Gifu University approved this study (#23-131, #28-114). For this type of study, formal consent is not required.

References

1. Keyes JW Jr (1995) SUV: standardized uptake or silly uptake value. *J Nucl Med* 36:1836–1839
2. Huang SC (2000) Anatomy of SUV. Standardized uptake value. *Nucl Med Biol* 27:643–646
3. Ishii K, Kono AK, Sasaki H, Miyamoto N, Fukuda T, Sakamoto S, Mori E (2006) Fully automatic diagnostic system for early- and late-onset mild Alzheimer's disease using FDG PET and 3D-SSP. *Eur J Nucl Med Mol Imaging* 33:575–583
4. Mito Y, Yoshida K, Yabe I, Makino K, Hirotsu M, Tashiro K, Kikuchi S, Sasaki H (2005) Brain 3D-SSP SPECT analysis in dementia with Lewy bodies, Parkinson's disease with and without dementia, and Alzheimer's disease. *Clin Neurol Neurosurg* 107:396–400
5. Friston KJ, Ashburner JT, Kiebel SJ, Nichols TE, Penny WD (2007) *Statistical parametric mapping: the analysis of functional brain images*. Academic Press, London
6. McNally KA, Paige AL, Varghese G, Zhang H, Novotny EJ, Spencer SS, Zubal IG, Blumenfeld H (2005) Localizing value of ictal-interictal SPECT analyzed by SPM (ISAS). *Epilepsia* 46:1450–1464
7. Minoshima S, Giordani B, Berent S, Frey K, Foster N, Kuhl D (1997) Metabolic reduction in the posterior cingulate cortex in very early Alzheimer's disease. *Ann Neurol* 42:85–94
8. Hara T, Kobayashi T, Ito S, Zhou X, Katafuchi T, Fujita H (2015) Quantitative analysis of torso FDG-PET scans by using anatomical standardization of normal cases from thorough physical examinations. *PLoS ONE* 10(5):e0125713
9. Shimizu Y, Hara T, Fukuoka D, Zhou X, Muramatsu C, Ito S, Hakozaaki K, Kumita S, Ishihara K, Katafuchi T, Fujita H (2013) Temporal subtraction system on torso FDG-PET scans based on statistical image analysis. *Proc SPIE Med Imaging* 8670:86703F-1–86703F-6
10. Bookstein FL (1989) Principal warps: thin-plate splines and the decomposition of deformation. *IEEE Trans Pattern Anal Mach Intell* 11:567–585
11. Cohen J (1992) A power primer. *Psychol Bull* 112(1):155–159
12. Faul F, Erdfelder E, Lang AG, Buchner A (2007) G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods* 39(2):175–191
13. Faul F, Erdfelder E, Buchner A, Lang AG (2009) Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. *Behav Res Methods* 41(4):1149–1160
14. Zhou X, Yamaguchi S, Zhou X, Chen H, Hara T, Yokoyama R, Kanematsu M, Fujita H (2013) Automatic organ localizations on 3D CT images by using majority-voting of multiple 2D detections based on local binary patterns and Haar-like features. *Proc SPIE Med Imaging* 8670:86703A-1–86703A-7
15. Zhou X, Xu R, Hara T, Hirano Y, Yokoyama R, Kanematsu M, Hoshi H, Kido S, Fujita H (2014) Development and evaluation of statistical shape modeling for principal inner organs on torso CT images. *Radiol Phys Technol* 7:277–283