Artificial Intelligence-Driven Single-Shot PET Image Artifact Detection and Disentanglement: Toward Routine Clinical Image Quality Assurance

Shiri Lord, Isaac; Salimi, Yazdan; Hervier, Elsa; Pezzoni, Agathe; Sanaat, Amirhossein; Mostafaei, Shayan; Rahmim, Arman; Mainta, Ismini Charis; Zaidi, Habib

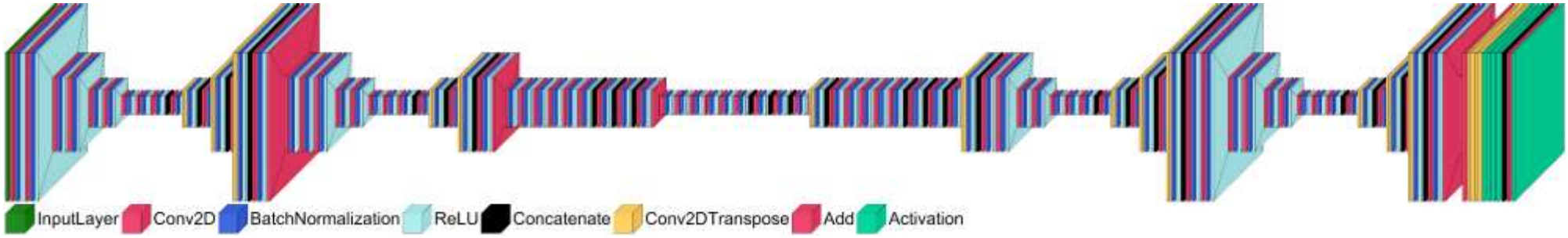
Image preprocessing

Different image preprocessing steps were taken for practical deep neural network training and to reproduce the quantitative PET values[1-6]. First, to reduce the PET intensity dynamic range due to varying injected activity, decay factors, and patient weights, both CT-ASC and non-ASC PET images were converted to standardized uptake values (SUVs) [1-6]. Then, based on 90% of the histogram of all data for further dynamic range reduction, CT-ASC and non-ASC PET images were divided by empiric factors of 5 and 2, respectively[1-6]. Hence, most values were in the range [0-1], and we did not apply any clipping on the values to ensure reproducibility in a clinical context[1-6].

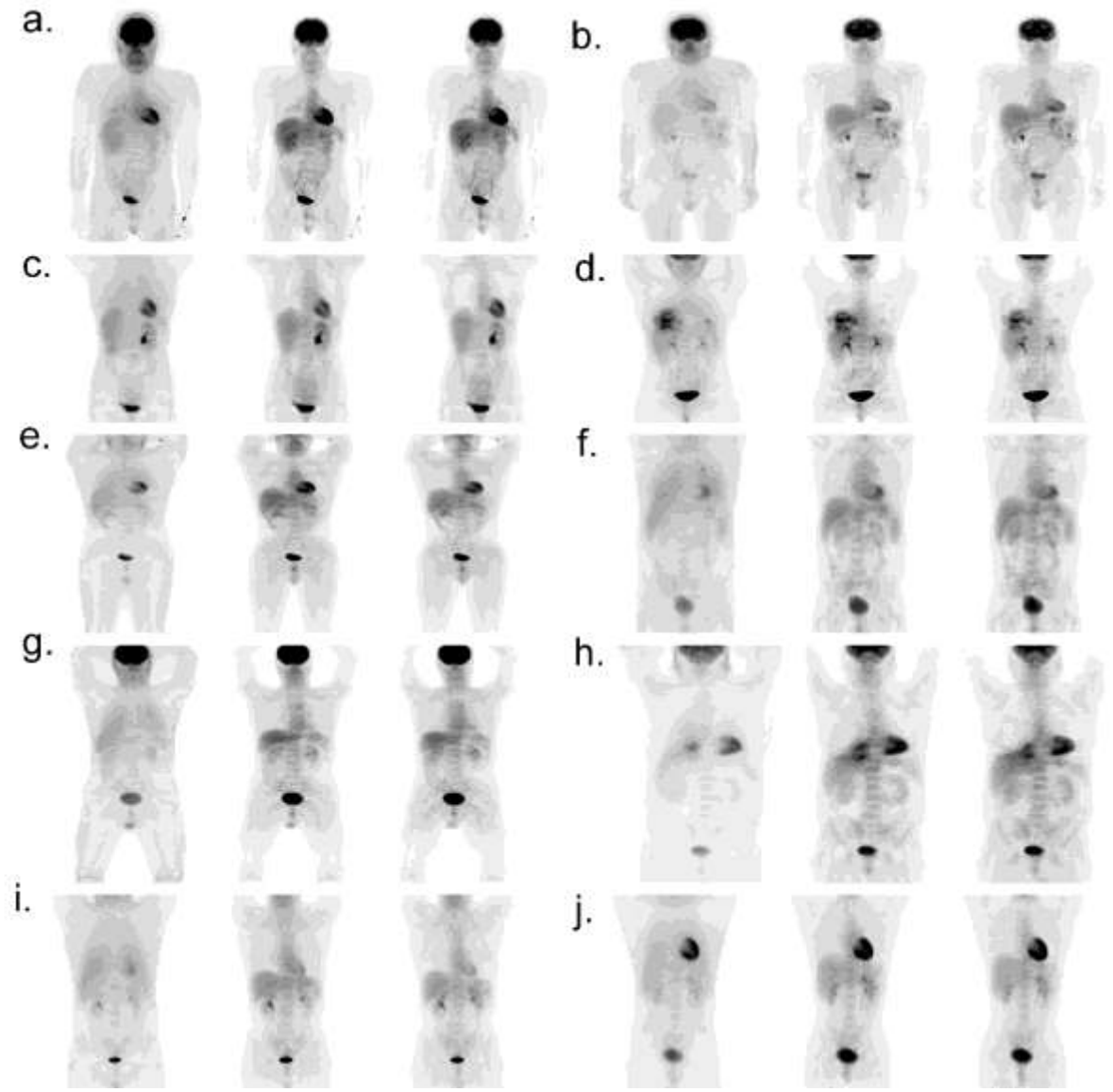
PET-QA-NET Deep Neural Network

We employed the modified U^2-Net architecture[1-7] that utilizes residual U-blocks in a U-shaped architecture[1-7]. In U^2-Net, unlike the U-Net-based network, the image resolution is not sacrificed with successive downsampling. Instead, the general U-structure of the U-Net was kept, and each block used another U-Net with a symmetric U-Shaped form known as ReSidual U-block (RSU) [1-6]. This RSU block allows multiscale feature extraction using a mixture of receptive fields. U^2-Net[7] uses deep supervision to include training loss in all scales for further local and global contextual information extraction[7]. The modified U^2-Net was trained in a 2D manner with an Adam optimizer, a learning rate of 0.001, an L2-norm loss, and a weight decay of 0.0001[1-6]. During model development and evaluation, we followed various steps adopted from best practice guidelines in nuclear medicine for AI algorithmic development [8, 9]. Quantitative metrics at the image and region level were calculated on clean test sets for model evaluation[1-6]. Furthermore, a subset of images (100 patients) with artifacts was used for qualitative analysis by physicians.

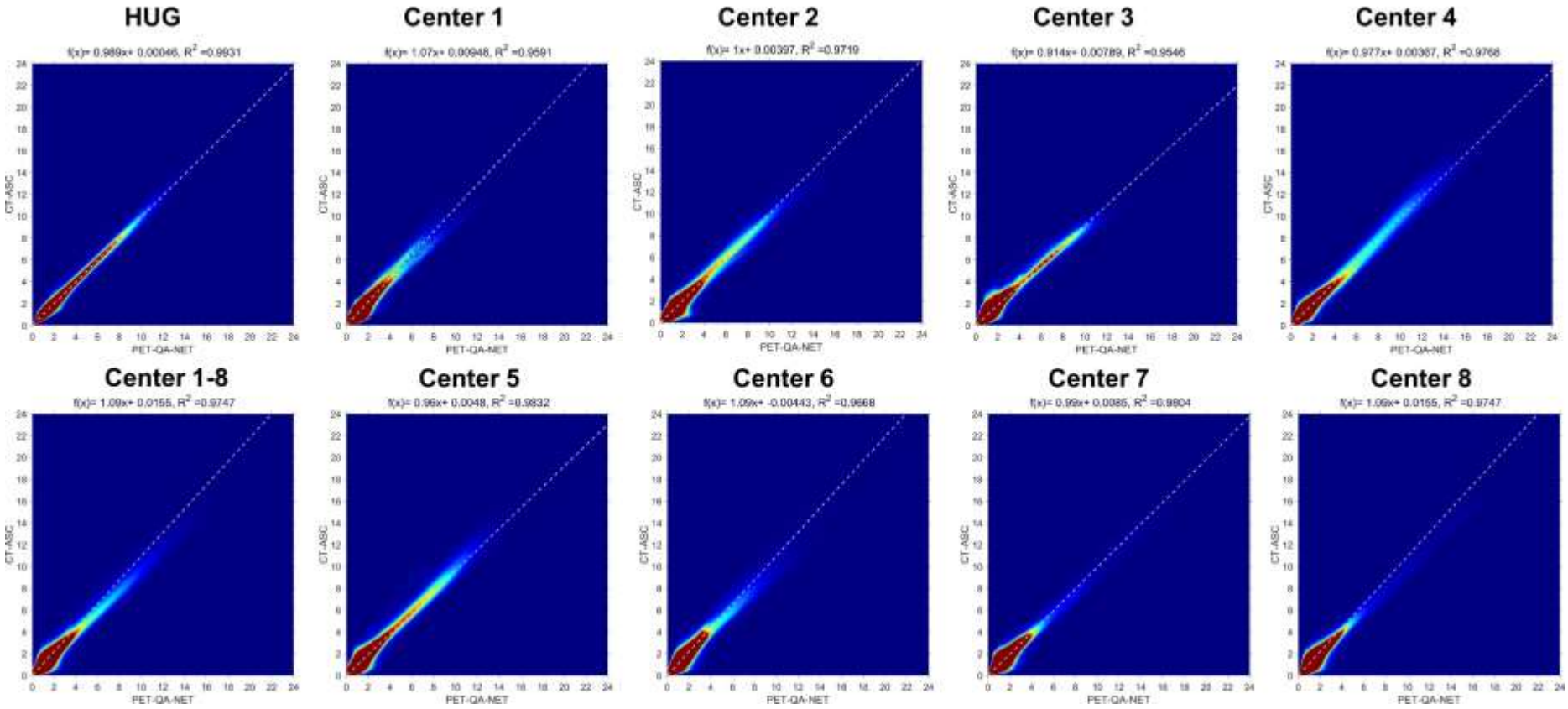
Supplemental Figures



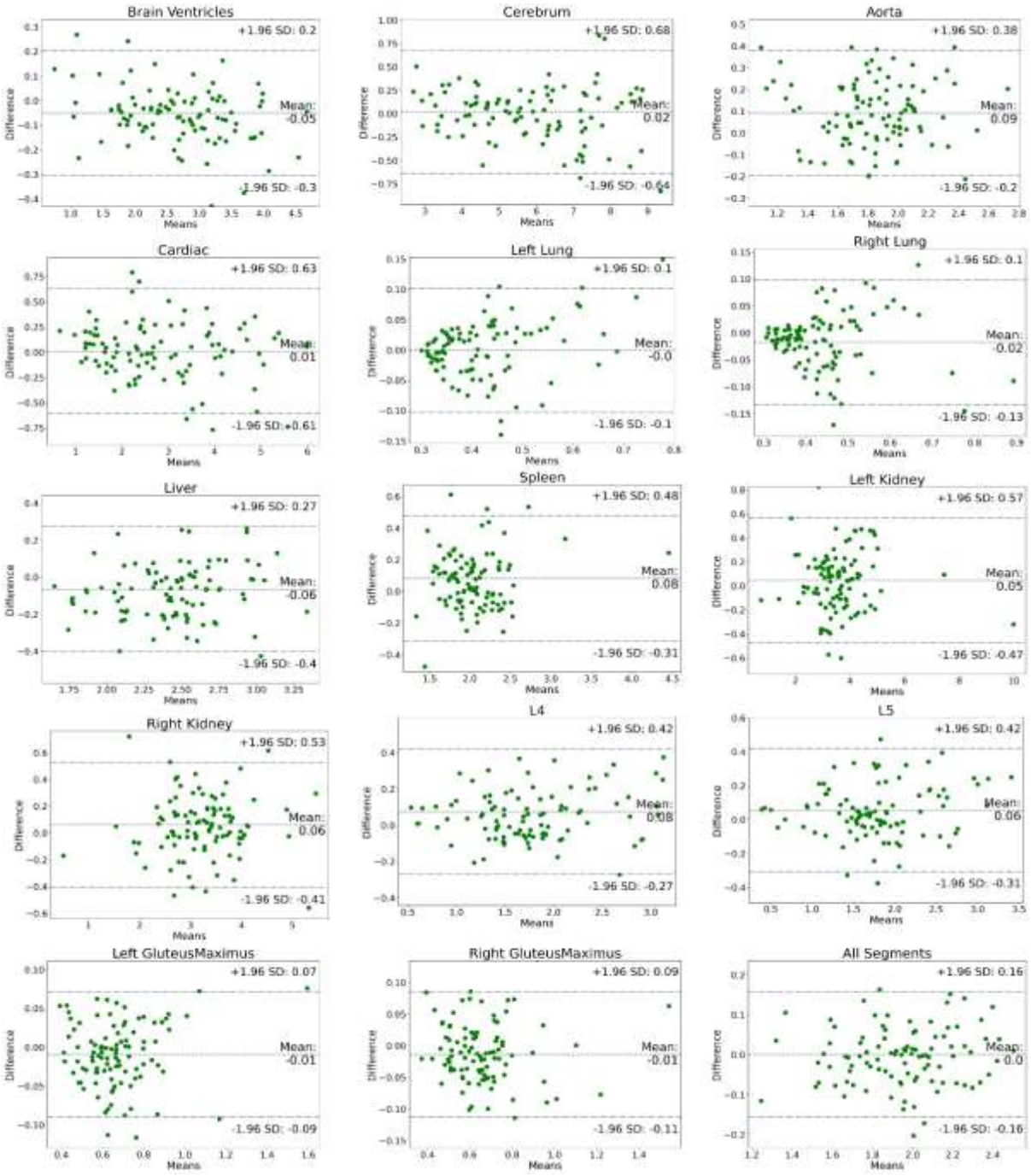
Supplemental Figure 1. U2Net architecture implemented in this study.



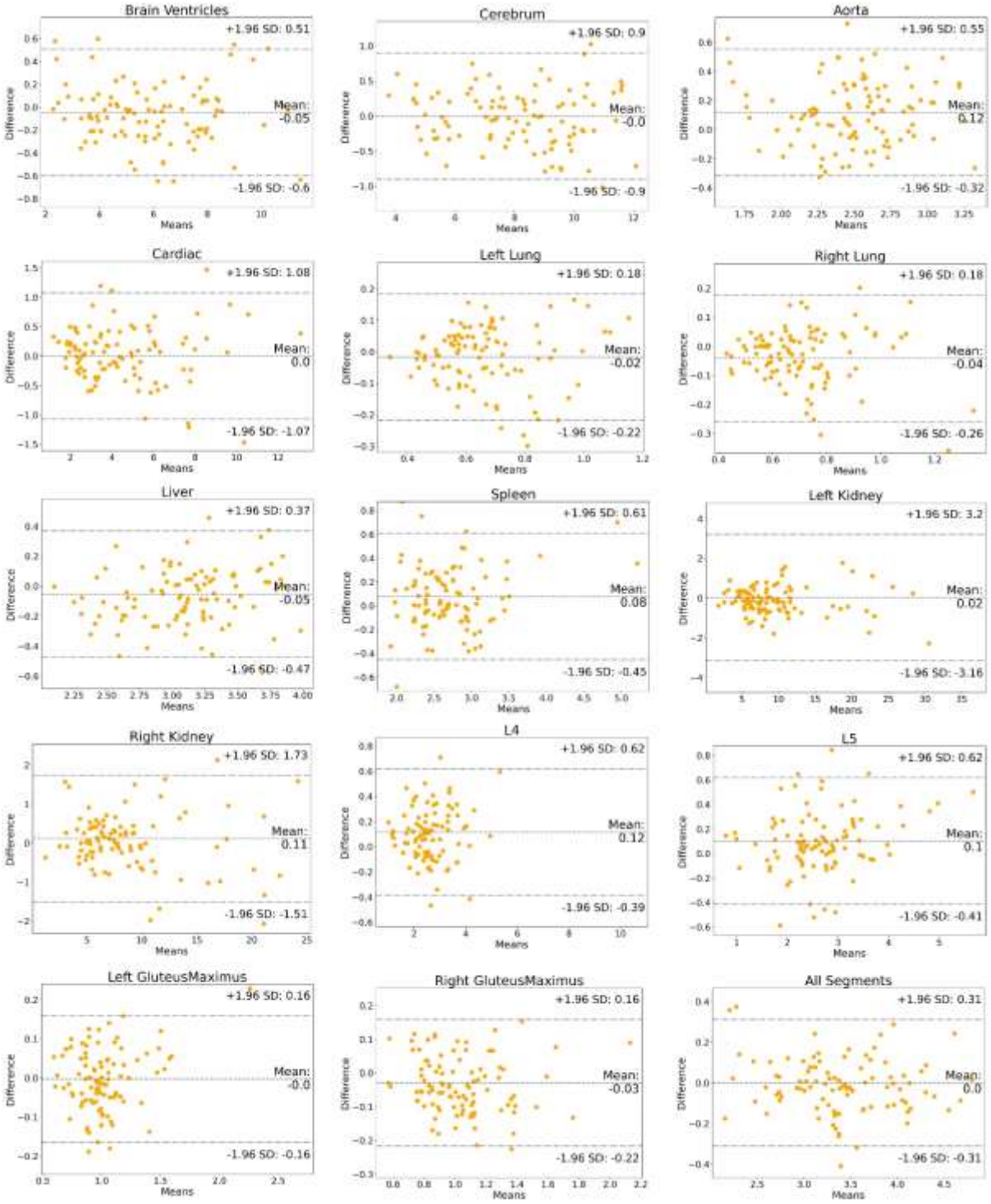
Supplemental Figure 2. MIPs of 8 different centers Non-ASC, CT-ASC, PET-QA-NET from left to right.



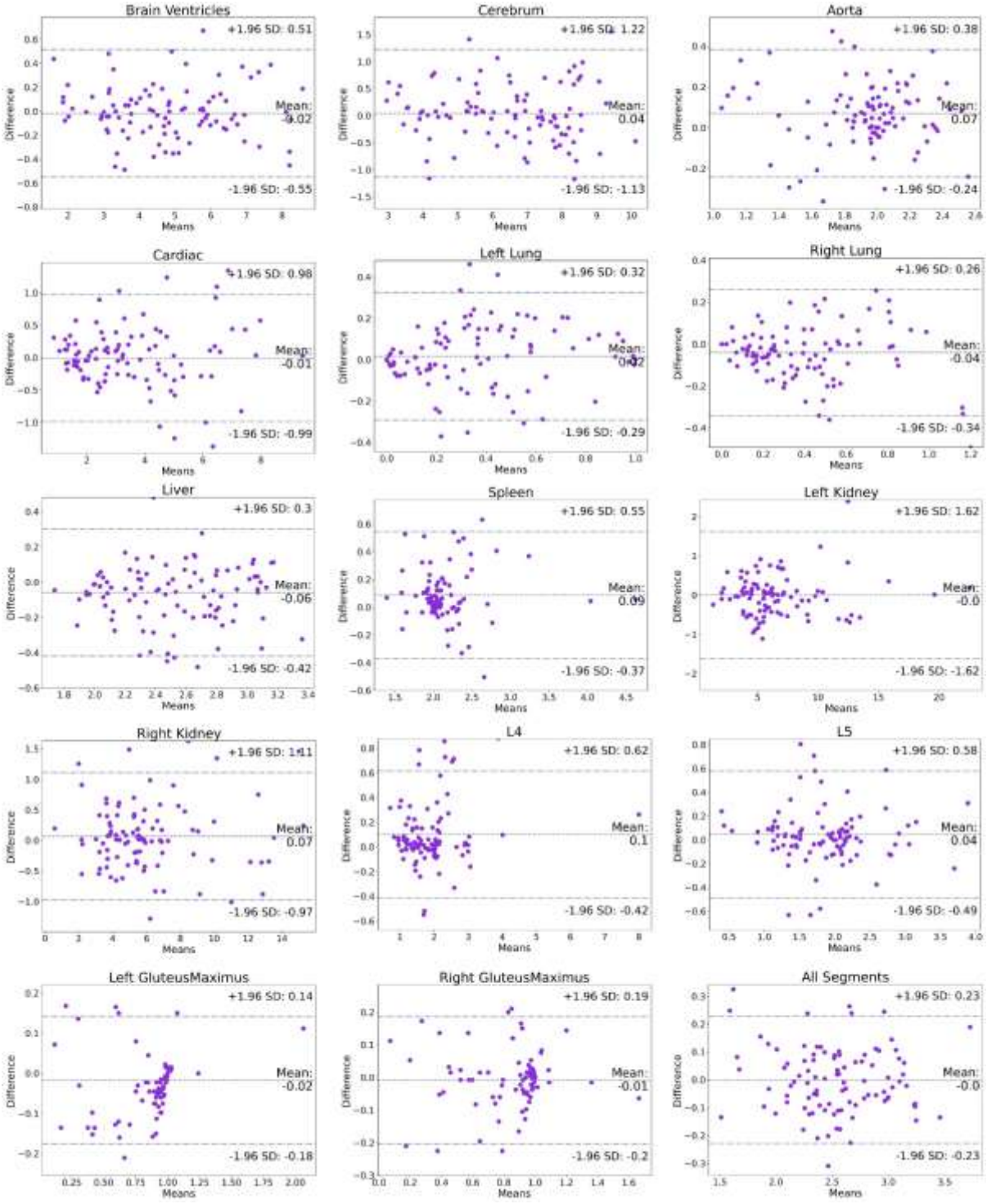
Supplemental Figure 3. Joint histogram of different centers.



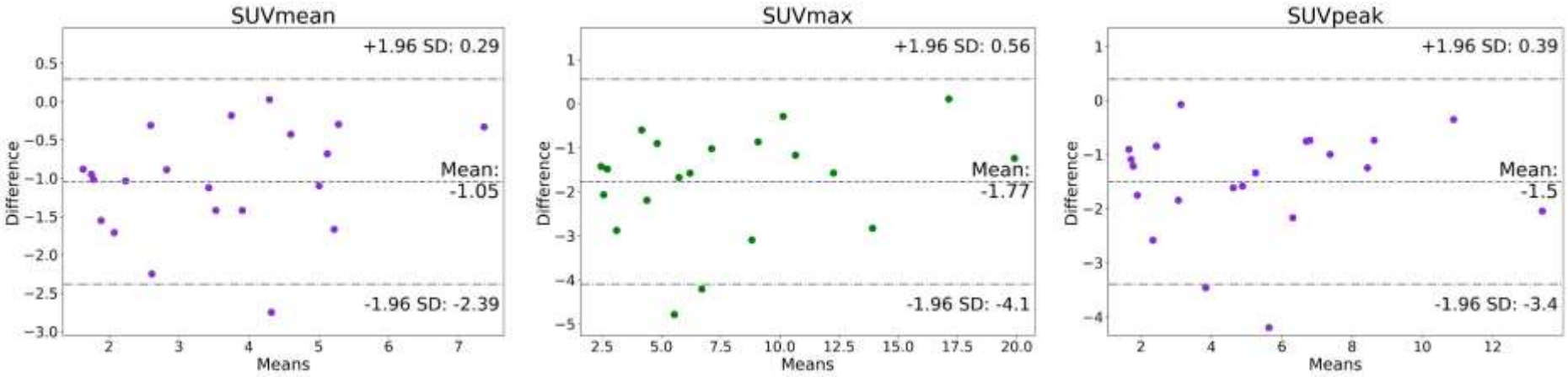
Supplemental Figure 4. Bland Altman of SUV_{mean} for different regions.



Supplemental Figure 5. Bland Altman of SUV_{max} for different regions.



Supplemental Figure 6. Bland Altman of SUV_{peak} for different regions.



Supplemental Figure 7. Bland Altman for SUVs for Motion artifact comparison between CT-ASC and PET-QA-NET**.**

Supplemental Tables

Supplemental Table 1. Patient demographics and PET/CT image acquisition parameters in the different centers for transfer learning and an external test set [1-6, 10-22].

		Centre #1	Centre #2	Centre #3	Centre #4	Centre #5	Centre #6	Centre #7	Centre 8
Demographic	**Sex (F/M/NA)**	29/44/0	15/31/0	12/39/2	2/23/50	27/37/0	40/57/0	18/121/0	27/120/0
	Age	64.58 ± 12.12	63.83 ± 8.63	61.65 ± 7.25	56.43 ± 7.22	49.9 ± 21.58	67.31 ± 10.58	56.58 ± 9.62	58.88 ± 10.06
Scanner	**PET**	GE-Discovery ST	GE-Discovery LS	GE-Discovery STE	GE-Discovery STE	Siemens Bigraph 6	GE-Discovery690	GE-Discovery STE	GE-Discovery RX
PET Acquisition and Reconstruction Parameters	**Time to Scan**	76.04 ± 18.45	82.22 ± 22.69	100.49 ± 30.71	99.44 ± 30.93	96.85 ± 14.04	71.98 ± 18.24	86.86 ± 20.53	82.22 ± 16.31
	Time per Bed	3.08 ± 0.99	3.48 ± 1.35	3.21 ± 0.52	3.21 ± 0.51	3.09 ± 0.35	2.49 ± 0.84	3.03 ± 0.17	2.98 ± 0.23
	Scatter Correction	Model-based, convolution subtraction	Model-based, convolution subtraction	Model-based, convolution subtraction	Model-based, convolution subtraction	Model-based	Model-based, convolution subtraction	Model-based	Model-based
	CT-ASC Reconstruction	OSEM	OSEM	OSEM	OSEM	OSEM	3D FORE+FBP	3D IR	OSEM
	Matrix Size	128×128	128×128	256×256	128×128	168×168	192×192	128×128	128×128
	Slice Thickness	3.27	4.25	3.4	3.27	3	3.27	3.27	3.27

Supplemental Table 2. Summary of radiation dose (mGy) to different organs and effective dose (mSv) for HUG dataset from low-dose CT acquisition for PET attenuation/scatter correction calculated by Impact Dosimetry Tool.

Gender	Male						Female					
	Mean	STD	Q1	Median	Q3	Max	Mean	STD	Q1	Median	Q3	Max
Effective dose (mSv)	2.91	2.39	1.43	2.33	3.72	17.30	3.68	2.79	2.02	3.01	4.52	24.53
Year (from background)	1.21	0.99	0.60	0.97	1.55	7.21	1.53	1.16	0.84	1.25	1.88	10.22
$CTDI_W$ (mGy)	3.53	2.41	2.06	3.25	4.45	18.72	3.69	2.39	2.33	3.37	4.58	25.46
$CTDI_{vol}$ (mGy)	4.41	3.01	2.57	4.06	5.56	23.40	4.61	2.99	2.92	4.21	5.73	31.82
DLP (mGy×cm)	790	540	461	728	996	4193	825	535	522	754	1026	570
SSDE (mGy)	3.42	2.53	1.89	2.89	4.50	17.72	3.57	2.35	2.22	3.06	4.38	18.74
Adrenals (mGy)	3.40	2.81	1.66	2.70	4.36	20.38	3.83	2.93	2.09	3.13	4.71	26.15
Bladder (mGy)	3.77	3.10	1.85	3.01	4.82	22.46	4.24	3.21	2.32	3.46	5.20	28.28
Brain (mGy)	3.77	3.10	1.85	3.01	4.82	22.45	4.18	3.16	2.30	3.42	5.13	27.82
Breast (mGy)	-	-	-	-	-	-	3.42	2.56	1.89	2.81	4.18	22.41
Colon (mGy)	3.10	2.57	1.50	2.45	3.94	18.65	3.69	2.83	2.00	3.00	4.54	25.19
Esophagus (mGy)	3.07	2.56	1.47	2.41	3.90	18.57	3.56	2.75	1.92	2.88	4.42	24.64
Extrathoracic Region (mGy)	4.34	3.56	2.15	3.49	5.54	25.79	4.83	3.63	2.66	3.95	5.89	31.79
Eye (mGy)	3.97	3.23	1.99	3.20	5.11	23.72	4.20	3.11	2.35	3.47	5.17	26.91
Gall Bladder (mGy)	3.50	2.89	1.70	2.77	4.47	20.94	4.05	3.10	2.20	3.30	4.99	27.46
Gonads (mGy)	3.95	3.21	1.97	3.18	5.07	23.52	3.59	2.77	1.93	2.90	4.45	24.86
Heart (mGy)	3.88	3.19	1.90	3.09	4.97	23.14	4.22	3.21	2.31	3.46	5.21	28.39
Kidneys (mGy)	3.67	3.01	1.81	2.94	4.68	21.81	4.20	3.17	2.31	3.44	5.14	27.91
Liver (mGy)	3.34	2.75	1.63	2.65	4.27	19.95	3.92	2.98	2.14	3.20	4.83	26.35
Lower Large Intestine (mGy)	2.94	2.45	1.42	2.32	3.74	17.75	3.52	2.70	1.90	2.85	4.34	24.16
Lung (mGy)	3.72	3.06	1.83	2.97	4.76	22.18	4.27	3.23	2.34	3.49	5.24	28.50
Lymphatic Nodes (mGy)	3.48	2.85	1.72	2.79	4.44	20.64	3.99	3.00	2.20	3.27	4.86	26.24
Muscle (mGy)	3.48	2.85	1.72	2.79	4.44	20.64	3.99	3.00	2.20	3.27	4.86	26.24
Oral Mucosa (mGy)	3.94	3.24	1.93	3.14	5.05	23.53	4.47	3.38	2.45	3.66	5.49	29.68
Pancreas (mGy)	3.12	2.60	1.50	2.45	3.96	18.83	3.66	2.83	1.98	2.97	4.54	25.27
Prostate (mGy)	3.74	3.08	1.84	2.98	4.79	22.33	-	-	-	-	-	-
Red Bone Marrow (mGy)	0.82	0.66	0.41	0.66	1.05	4.90	0.95	0.70	0.53	0.78	1.17	5.92
Salivary Glands (mGy)	4.49	3.65	2.24	3.61	5.75	26.70	4.90	3.65	2.72	4.04	6.01	31.80
Skeleton (mGy)	7.76	6.26	3.92	6.27	10.04	46.76	9.18	6.73	5.17	7.59	11.31	56.94
Skin (mGy)	3.53	2.86	1.77	2.85	4.53	21.08	3.99	2.96	2.23	3.29	4.89	25.60
Small Intestine (mGy)	3.24	2.69	1.57	2.55	4.12	19.48	3.80	2.92	2.06	3.09	4.69	26.00
Spleen (mGy)	3.12	2.58	1.52	2.47	3.99	18.67	3.73	2.84	2.03	3.05	4.59	25.20
Stomach (mGy)	3.41	2.81	1.67	2.71	4.37	20.38	4.02	3.05	2.20	3.29	4.95	26.99
Thymus (mGy)	4.31	3.52	2.14	3.45	5.51	25.53	4.67	3.50	2.58	3.82	5.69	30.55
Thyroid (mGy)	5.73	4.62	2.89	4.63	7.44	34.68	6.19	4.56	3.48	5.11	7.62	39.10
Upper Large Intestine (mGy)	3.25	2.70	1.58	2.57	4.15	19.55	3.86	2.95	2.10	3.14	4.74	26.22
Uterus (mGy)	-	-	-	-	-	-	3.92	3.01	2.12	3.19	4.83	26.77

Supplemental Table 3. Summary statistics of image quality metrics reflecting the quantitative accuracy of the estimated tracer uptake for the clean, untouched test set in the HUG dataset, and 80% of each of the 8 other centers' test sets (transfer learning on 20% of each center for initial fine tuning). MAE: mean absolute error, MSE: mean square error, RE (%): percent relative error, ARE (%): percent absolute relative error, SSIM: Structural Similarity Index, PSNR: Peak signal-to-noise ratio.

		MAE	MSE	RE (%)	ARE (%)	SSIM	PSNR
HUG (n=100)	**Mean**	0.09±0.02	0.03±0.01	-1.19±3.60	9.91±1.53	0.99±0.00	36.31±1.16
	95% CI	0.09 to 0.10	0.02 to 0.03	-1.90 to -0.47	9.60 to 10.22	0.99 to 0.99	36.08 to 36.54
Centre #1 (n=4)	**Mean**	0.16±0.02	0.06±0.02	10.19±5.28	18.07±2.54	0.94±0.02	34.33±0.89
	95% CI	0.12 to 0.19	0.03 to 0.08	1.79 to 18.59	14.02 to 22.11	0.90 to 0.98	32.91 to 35.74
Centre #2 (n=8)	**Mean**	0.16±0.01	0.07±0.02	-3.85±6.96	20.08±1.04	0.92±0.04	33.64±0.90
	95% CI	0.145 to 0.17	0.06 to 0.09	-9.67 to 1.97	19.21 to 20.95	0.89 to 0.95	32.89 to 34.39
Centre #3 (n=11)	**Mean**	0.16±0.02	0.07±0.02	-1.65±2.77	18.47±1.17	0.81±0.02	34.13±1.08
	95% CI	0.15 to 0.17	0.05 to 0.08	-3.51 to 0.21	17.68 to 19.25	0.80 to 0.82	33.41 to 34.86
Centre #4 (n=15)	**Mean**	0.14±0.02	0.05±0.01	-2.14±4.98	16.53±1.91	0.94±0.02	34.31±0.86
	95% CI	0.13 to 0.15	0.04 to 0.062	-4.89 to 0.61	15.48 to 17.59	0.93 to 0.95	33.83 to 34.79
Centre #5 (n=31)	**Mean**	0.15±0.03	0.08±0.09	-2.55±3.63	17.15±2.09	0.90±0.03	33.86±1.54
	95% CI	0.14 to 0.16	0.05 to 0.11	-3.88 to -1.22	16.38 to 17.91	0.89 to 0.92	33.30 to 34.43
Centre #6 (n=37)	**Mean**	0.16±0.03	0.08±0.05	0.39±5.87	17.82±2.41	0.93±0.03	33.43±1.30
	95% CI	0.14 to 0.16	0.06 to 0.10	-1.56 to 2.35	17.01 to 18.62	0.92 to 0.94	32.99 to 33.87
Centre #7 (n=40)	**Mean**	0.13±	0.05±0.01	-3.06±6.91	15.91±2.12	0.95±0.01	34.60±1.05
	95% CI	0.12 to 0.13	0.040 to 0.05	-5.27 to -0.85	15.23 to 16.59	0.95 to 0.95	34.26 to 34.93
Centre #8 (n=51)	**Mean**	0.12±0.02	0.06±0.09	-0.15±4.82	14.87±1.83	0.02	34.84±1.55
	95% CI	0.12 to 0.13	0.032 to 0.08	-1.50 to 1.21	14.36 to 15.38	0.95 to 0.96	34.41 to 35.28
All Centers 1-8 (n=197)	**Mean**	0.14±0.03	0.06±0.06	-1.19±5.73	16.60±2.43	0.93±0.04	34.23±1.39
	95% CI	0.13 to 0.14	0.054 to 0.07	-1.99 to -0.38	16.26 to 16.94	0.92 to 0.93	34.04 to 34.43

Supplemental Table 4. Mean error and absolute mean error of SUV_{peak}, SUV_{max}, and SUV_{mean} across the different VOIs in a different region for 99 clean test sets. CT-ASC was defined as standard of reference for SUV comparison.

		Mean Error (SUV)		Absolute Mean Error (SUV)		Mann Witney
	Organ	**Mean±SD**	**CI95%**	**Mean±SD**	**CI95%**	**P-value**
SUV_{mean}	**Brain Ventricles**	0.05 ± 0.13	0 to 0.02	0.11 ± 0.09	0.1 to 0.09	0.682
	Cerebrum	-0.02 ± 0.34	-0.1 to -0.09	0.24 ± 0.24	0.2 to 0.19	0.96
	Aorta	-0.09 ± 0.15	-0.1 to -0.12	0.14 ± 0.1	0.1 to 0.12	**<0.05**
	Cardiac	-0.01 ± 0.32	-0.1 to -0.07	0.25 ± 0.21	0.2 to 0.21	0.886
	Right Lung	0.02 ± 0.06	0 to 0.01	0.05 ± 0.05	0 to 0.04	0.124
	Left Lung	0 ± 0.05	0 to -0.01	0.05 ± 0.03	0 to 0.04	0.708
	Liver	0.06 ± 0.17	0 to 0.03	0.16 ± 0.11	0.1 to 0.14	0.206
	Spleen	-0.08 ± 0.2	-0.1 to -0.12	0.16 ± 0.15	0.1 to 0.13	0.144
	Left Kidney	-0.05 ± 0.26	-0.1 to -0.1	0.21 ± 0.17	0.2 to 0.17	0.827
	Right Kidney	-0.06 ± 0.24	-0.1 to -0.11	0.18 ± 0.17	0.1 to 0.15	0.584
	L4	-0.08 ± 0.18	-0.1 to -0.11	0.13 ± 0.14	0.1 to 0.1	0.452
	L5	-0.06 ± 0.19	-0.1 to -0.09	0.14 ± 0.14	0.1 to 0.11	0.562
	Right Gluteus Maximus	0.01 ± 0.05	0 to 0	0.04 ± 0.03	0 to 0.03	0.466
	Left Gluteus Maximus	0.01 ± 0.04	0 to 0	0.03 ± 0.03	0 to 0.03	0.524
	All Segments	0 ± 0.08	0 to -0.02	0.13 ± 0.04	0.1 to 0.12	0.854
SUV_{max}	**Brain Ventricles**	0.05 ± 0.28	0 to -0.01	0.24 ± 0.18	0.2 to 0.2	0.835
	Cerebrum	0 ± 0.46	-0.1 to -0.09	0.35 ± 0.3	0.3 to 0.29	0.956
	Aorta	-0.12 ± 0.22	-0.2 to -0.16	0.2 ± 0.15	0.2 to 0.17	0.046
	Cardiac	0 ± 0.55	-0.1 to -0.11	0.43 ± 0.37	0.4 to 0.36	0.921
	Right Lung	0.04 ± 0.11	0 to 0.02	0.11 ± 0.09	0.1 to 0.09	0.13
	Left Lung	0.02 ± 0.1	0 to 0	0.1 ± 0.06	0.1 to 0.09	0.712
	Liver	0.05 ± 0.22	0 to 0.01	0.2 ± 0.13	0.2 to 0.18	0.436
	Spleen	-0.08 ± 0.27	-0.1 to -0.13	0.22 ± 0.19	0.2 to 0.18	0.312
	Left Kidney	-0.02 ± 1.62	-0.3 to -0.34	0.78 ± 1.43	0.5 to 0.5	0.829
	Right Kidney	-0.11 ± 0.83	-0.3 to -0.28	0.58 ± 0.6	0.5 to 0.46	0.891
	L4	-0.12 ± 0.26	-0.2 to -0.17	0.2 ± 0.19	0.2 to 0.16	0.348
	L5	-0.1 ± 0.26	-0.2 to -0.15	0.2 ± 0.2	0.2 to 0.16	0.416
	Right Gluteus Maximus	0.03 ± 0.1	0 to 0.01	0.08 ± 0.06	0.1 to 0.07	0.246
	Left GluteusMaximus	0 ± 0.08	0 to -0.01	0.06 ± 0.05	0.1 to 0.05	0.469
	All Segments	0 ± 0.16	0 to -0.03	0.24 ± 0.1	0.2 to 0.22	0.774
SUV_{peak}	**Brain Ventricles**	0.02 ± 0.27	0 to -0.04	0.23 ± 0.21	0.2 to 0.19	0.919
	Cerebrum	-0.04 ± 0.6	-0.2 to -0.16	0.43 ± 0.42	0.4 to 0.35	0.933
	Aorta	-0.07 ± 0.16	-0.1 to -0.1	0.13 ± 0.12	0.1 to 0.11	0.073
	Cardiac	0.01 ± 0.5	-0.1 to -0.09	0.39 ± 0.35	0.3 to 0.32	0.968
	Right Lung	0.04 ± 0.15	0 to 0.01	0.15 ± 0.11	0.1 to 0.13	0.286
	Left Lung	-0.02 ± 0.16	0 to -0.05	0.15 ± 0.11	0.1 to 0.13	0.787
	Liver	0.06 ± 0.18	0 to 0.02	0.17 ± 0.13	0.1 to 0.15	0.321
	Spleen	-0.09 ± 0.23	-0.1 to -0.13	0.16 ± 0.19	0.1 to 0.12	0.035
	Left Kidney	0 ± 0.83	-0.2 to -0.16	0.48 ± 0.68	0.3 to 0.34	0.897
	Right Kidney	-0.07 ± 0.53	-0.2 to -0.17	0.39 ± 0.37	0.3 to 0.32	0.883
	L4	-0.1 ± 0.26	-0.2 to -0.15	0.17 ± 0.23	0.1 to 0.12	0.361
	L5	-0.04 ± 0.27	-0.1 to -0.1	0.17 ± 0.21	0.1 to 0.13	0.737
	Right GluteusMaximus	0.01 ± 0.1	0 to -0.01	0.06 ± 0.08	0 to 0.04	0.656
	Left GluteusMaximus	0.02 ± 0.08	0 to 0	0.05 ± 0.06	0 to 0.04	0.251
	All Segments	0 ± 0.12	0 to -0.02	0.21 ± 0.06	0.2 to 0.2	0.874

Supplemental Table 5. Intra- and inter-reader correlation coefficient ICC (95% CI) for image quality, artifacts, confidence, and lesion detectability in the different regions.

Region	Image Quality	Artifact	Confidence	Lesion
Intra-reader correlation coefficient ICC				
Brain	0.72 (0.59-0.81)	0.79 (0.69-0.86)	0.69 (0.56-0.79)	0.87 (0.81-0.92)
Chest	0.65 (0.50-0.76)	0.74 (0.62-0.83)	0.72 (0.59-0.81)	0.89 (0.84-0.93)
Chest/Abd interface	0.54 (0.36-0.68)	0.74 (0.62-0.82)	0.57 (0.39-0.70)	0.88 (0.82-0.92)
Abdomen	0.62 (0.46-0.74)	0.72 (0.59-0.81)	0.56 (0.39-0.69)	0.91 (0.86-0.94)
Pelvis	0.68 (0.54-0.78)	0.74 (0.62-0.82)	0.64 (0.50-0.75)	0.96 (0.93-0.97)
Extremities	0.71 (0.58-0.80)	0.75 (0.64-0.83)	0.68 (0.54-0.80)	0.85 (0.77-0.90)
All regions	0.67 (0.61-0.72)	0.78 (0.74-0.81)	0.65 (0.60-0.70)	0.90 (0.89-0.92)
Inter-reader correlation coefficient using ICC				
Brain	0.74 (0.66-0.80)	0.80 (0.68-0.90)	0.71 (0.62-0.77)	0.80 (074-0.85)
Chest	0.58 (0.46-0.67)	0.71 (0.58-0.78)	0.53 (0.40-0.64)	0.56 (0.29-0.66)
Chest/Abd interface	0.58 (0.50-0.65)	0.54 (0.45-0.65)	0.52 (0.44-0.61)	0.58 (0.42-0.72)
Abdomen	0.61 (0.47-0.73)	0.56 (0.48-0.67)	0.51 (0.40-0.63)	0.67 (0.55-0.76)
Pelvis	0.53 (0.39-0.63)	0.60 (0.45-0.74)	0.44 (0.35-0.55)	0.66 (0.56-0.77)
Extremities	0.67 (0.58-0.74)	0.76 (0.59-0.89)	0.53 (0.39-0.63)	0.28 (0.10-0.45)
All regions	0.66 (0.60-0.69)	0.68 (0.62-0.74)	0.53 (0.48-0.59)	0.61 (0.51-0.72)

Supplemental Table 6. Comparison of image quality between CT-ASC and PET-QA-NET generated PET images.

Region	Image Quality	p-value*	p-value#	p-value%
Head and Neck	Excellent	0.552	0.340	0.313
	High	0.636		
	Average	0.633		
	Poor	**<0.05**		
	Very poor	0.406		
Chest	Excellent	0.130	0.660	0.623
	High	0.760		
	Average	0.527		
	Poor	0.992		
	Very poor	0.990		
Chest/Abd Int	Excellent	0.096	**<0.005**	**<0.002**
	High	**<0.001**		
	Average	**<0.05**		
	Poor	**<0.05**		
	Very poor	0.881		
Abdomen	Excellent	0.732	0.854	0.543
	High	0.342		
	Average	0.761		
	Poor	0.417		
	Very poor	0.999		
Pelvis	Excellent	0.311	0.591	0.487
	High	0.359		
	Average	0.993		
	Poor	0.231		
	Very poor	0.999		
Extremities	Excellent	0.549	0.479	0.325
	High	0.223		
	Average	0.438		
	Poor	0.147		
	Very poor	0.999		
All regions	Excellent	0.430	0.442	0.411
	High	0.154		
	Average	0.387		
	Poor	**<0.05**		
	Very poor	0.652		

P-value* is based on the McNemar test; P-value# is based on the marginal homogeneity test; P-value% is based on a generalized linear model after adjustment of ICC for all raters (CT-ASC and PET-QA-NET) and readers (1 and 2);

Supplemental Table 7. Comparison of confidence between CT-ASC and PET-QA-NET.

Region	Confidence	P-value*	P-value[#]	P-value[%]
Head and Neck	Excellent	0.834	0.799	0.704
	High	0.448		
	Average	0.485		
	Poor	0.999		
	Very poor	0.311		
Chest	Excellent	0.321	0.835	0.698
	High	0.877		
	Average	0.433		
	Poor	0.699		
	Very poor	0.990		
Chest/Abd Int	Excellent	0.406	0.076	**<0.05**
	High	**<0.01**		
	Average	**<0.01**		
	Poor	0.515		
	Very poor	0.990		
Abdomen	Excellent	0.773	0.625	0.487
	High	0.165		
	Average	0.133		
	Poor	0.699		
	Very poor	0.991		
Pelvis	Excellent	0.999	0.148	**<0.05**
	High	**<0.02**		
	Average	**<0.01**		
	Poor	0.992		
	Very poor	0.881		
Extremities	Excellent	0.502	0.085	**<0.05**
	High	**<0.05**		
	Average	**<0.05**		
	Poor	0.649		
	Very poor	0.560		
All regions	Excellent	0.861	**<0.05**	**<0.05**
	High	**<0.05**		
	Average	**<0.02**		
	Poor	0.777		
	Very poor	0.314		

P-value* is based on the McNemar test; P-value[#] is based on the marginal homogeneity test; P-value[%] is based on a generalized linear model after adjustment of ICC for all raters (CT-ASC and PET-QA-NET) and readers (reader 1 and 2);

Supplemental Table 8. Comparison of the occurrence of artifacts between CT-ASC and PET-QA-NET.

Region	Artifact	P-value*	P-value#	P-value%
Head and Neck	None	0.194	0.741	0.664
	Minor	0.395		
	Moderate	0.999		
	Major	0.311		
	Unacceptable	0.990		
Chest	None	**<0.02**	0.190	0.098
	Minor	0.066		
	Moderate	0.173		
	Major	0.560		
	Unacceptable	0.990		
Chest/Abd Int	None	**<0.001**	**<0.001**	**<0.001**
	Minor	0.093		
	Moderate	**<0.001**		
	Major	**<0.005**		
	Unacceptable	0.998		
Abdomen	None	0.876	0.938	0.689
	Minor	0.636		
	Moderate	0.406		
	Major	0.881		
	Unacceptable	0.998		
Pelvis	None	0.272	0.818	0.754
	Minor	0.172		
	Moderate	0.787		
	Major	0.891		
	Unacceptable	0.787		
Extremities	None	**<0.05**	0.228	0.106
	Minor	0.552		
	Moderate	**<0.05**		
	Major	0.649		
	Unacceptable	0.560		
All regions	None	0.084	0.071	**<0.05**
	Minor	0.692		
	Moderate	**<0.02**		
	Major	0.306		
	Unacceptable	0.562		

P-value* is based on the McNemar test; P-value# is based on the marginal homogeneity test; P-value% is based on a generalized linear model after adjustment of ICC for all raters (CT-ASC and PET-QA-NET) and readers (reader 1 and 2);

Supplemental Table 9. Comparison of the number of detected lesions between CT-ASC and PET-QA-NET.

Region	Number	P-value[#]	P-value[%]
Brain	0	0.060	**<0.05**
	1		
	2		
	3		
	4		
	5≤		
Chest	0	0.337	0.219
	1		
	2		
	3		
	4		
	5≤		
Chest/Abd Int	0	0.401	0.378
	1		
	2		
	3		
	4		
	5≤		
Abdomen	0	0.412	0.391
	1		
	2		
	3		
	4		
	5≤		
Pelvis	0	0.093	0.069
	1		
	2		
	3		
	4		
	5≤		
Extremities	0	0.523	0.398
	1		
	2		
	3		
	4		
	5≤		
All regions	0	0.972	0.891
	1		
	2		
	3		
	4		
	5≤		

P-value[#] is based on the linear-by-linear association using the Mantel-Haenszel test; P-value[%] is based on a generalized linear model after adjustment of ICC for all raters (CT-ASC and PET-QA-NET) and readers (reader 1 and 2).

Supplemental Table 10. Mean error and absolute mean error of SUV_{peak}, SUV_{max}, and SUV_{mean} across the different VOIs for different lesions in the chest/abdomen regions (motion artifact cases). CT-ASC were defined as reference for SUVs comparison.

SUV	Mean	95% CI	Mean	95% CI	P-value
SUV_{max}	-1.77±1.22	-2.33 to -1.22	1.78±1.22	1.24 to 2.34	0.0001
SUV_{mean}	-1.05±0.70	-1.37 to -0.73	1.05±0.70	0.73 to 1.37	0.0001
SUV_{peak}	-1.50±1.00	-2.00 to -1.05	1.50±1.00	1.10 to 1.11	0.0001

References

1. Shiri I, Vafaei Sadr A, Akhavan A, et al. Decentralized collaborative multi-institutional PET attenuation and scatter correction using federated deep learning. Eur J Nucl Med Mol Imaging. 2023;50:1034-1050.
2. Shiri I, Sanaat A, Salimi Y, et al. PET-QA-NET: Towards routine PET image artifact detection and correction using deep convolutional neural networks. 2021 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), IEEE, 2021, pp. 1-3.
3. Shiri I, Salimi Y, Sanaat A, et al. Fully Automated PET Image Artifacts Detection and Correction Using Deep Neural Networks. Journal of Nuclear Medicine 2022, pp. 3218.
4. Shiri I, Salimi Y, Maghsudi M, et al. Deep Adaptive Transfer Learning for Site-Specific PET Attenuation and Scatter Correction from Multi-National/Institutional Datasets. 2021 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), IEEE, 2022, pp. 1-3.
5. Shiri I, Sadr AV, Sanaat A, et al. Federated learning-based deep learning model for PET attenuation and scatter correction: a multi-center study. 2021 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), IEEE, 2021, pp. 1-3.
6. Shiri I, Arabi H, Geramifar P, et al. Deep-JASC: joint attenuation and scatter correction in whole-body (18)F-FDG PET using a deep residual network. Eur J Nucl Med Mol Imaging. 2020;47:2533-2548.
7. Qin X, Zhang Z, Huang C, et al. U2-Net: Going deeper with nested U-structure for salient object detection. Pattern Recognition. 2020;106:107404.
8. Bradshaw TJ, Boellaard R, Dutta J, et al. Nuclear Medicine and Artificial Intelligence: Best Practices for Algorithm Development. Journal of nuclear medicine : official publication, Society of Nuclear Medicine. 2022;63:500-510.
9. Saboury B, Bradshaw T, Boellaard R, et al. Artificial Intelligence in Nuclear Medicine: Opportunities, Challenges, and Responsibilities Toward a Trustworthy Ecosystem. Journal of nuclear medicine : official publication, Society of Nuclear Medicine. 2022.
10. Clark K, Vendt B, Smith K, et al. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. J Digit Imaging. 2013;26:1045-1057.
11. Machtay M, Duan F, Siegel BA, et al. Prediction of survival by [18F]fluorodeoxyglucose positron emission tomography in patients with locally advanced non-small-cell lung cancer undergoing definitive chemoradiation therapy: results of the ACRIN 6668/RTOG 0235 trial. J Clin Oncol. 2013;31:3823-3830.
12. Kinahan P, Muzi M, Bialecki B, et al. Data from the ACRIN 6668 Trial NSCLC-FDG-PET. Cancer Imaging Arch. 2019;10.
13. Bakr S, Gevaert O, Echegaray S, et al. Data for NSCLC radiogenomics collection. The Cancer Imaging Archive. 2017.
14. Bakr S, Gevaert O, Echegaray S, et al. A radiogenomic dataset of non-small cell lung cancer. Sci Data. 2018;5:180202.
15. Gevaert O, Xu J, Hoang CD, et al. Non-small cell lung cancer: identifying prognostic imaging biomarkers by leveraging public gene expression microarray data--methods and preliminary results. Radiology. 2012;264:387-396.
16. Grossberg A, Elhalawani H, Mohamed A, et al. MD Anderson Cancer Center Head and Neck Quantitative Imaging Working Group.(2020) HNSCC [Dataset]. The Cancer Imaging Archive DOI: https://doi org/107937/k9/tcia. 2020;a8sh-7363.
17. Grossberg AJ, Mohamed ASR, Elhalawani H, et al. Imaging and clinical data archive for head and neck squamous cell carcinoma patients treated with radiotherapy. Sci Data. 2018;5:180173.
18. Matched computed tomography segmentation and demographic data for oropharyngeal cancer radiomics challenges. Sci Data. 2017;4:170077.
19. Shiri I, Amini M, Salimi Y, et al. Multi-institutional PET/CT image segmentation using a decentralized federated deep transformer learning algorithm. Journal of Nuclear Medicine 2022, pp. 3348.
20. Shiri I, Razeghi B, Vafaei Sadr A, et al. Multi-institutional PET/CT image segmentation using federated deep transformer learning. Comput Methods Programs Biomed. 2023;240:107706.
21. Shiri I, Showkatian E, Mohammadi R, et al. Collaborative Multi-Institutional Prostate Lesion Segmentation from MR images Using Deep Federated Learning Framework. 2022 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), IEEE, 2022, pp. 1-3.
22. Shiri I, Vafaei Sadr A, Amini M, et al. Decentralized Distributed Multi-institutional PET Image Segmentation Using a Federated Deep Learning Framework. Clin Nucl Med. 2022;47:606-617.