**Supplementary Materials**

# Artificial Intelligence enables whole body Positron Emission Tomography Scans with minimal radiation exposure

Yan-Ran (Joyce) Wang[1], Lucia Baratto[1], K Elizabeth Hawk[1], Ashok J Theruvath[1], Allison Pribnow[3], Avnesh S Thakor[1], Sergios Gatidis[2], Rong Lu[4], Santosh E Gummidipundi[4], Jordi Garcia-Diaz[1], Daniel Rubin[1,3,*], and Heike E. Daldrup-Link[1,3,*]

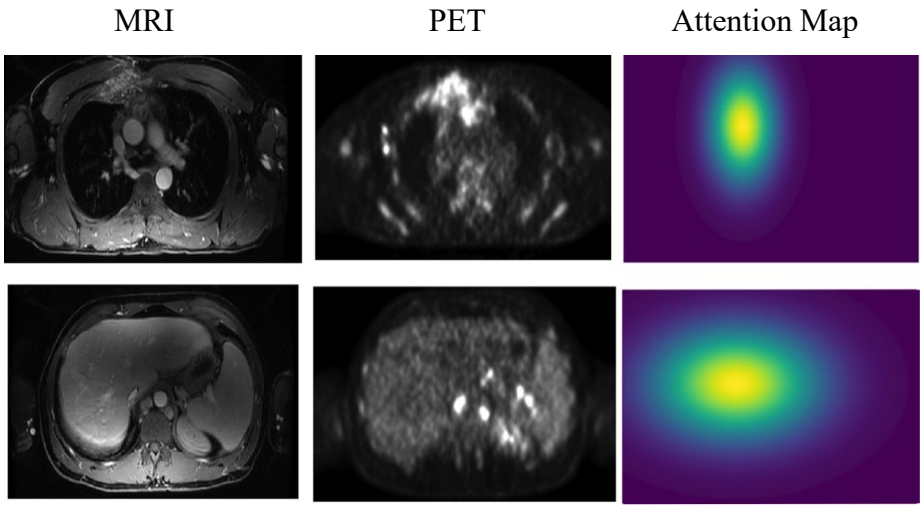
[1]Department of Radiology, Molecular Imaging Program at Stanford, Stanford University, Stanford, California 94304, USA

[2]Department of Diagnostic and Interventional Radiology, University Hospital Tuebingen, Tuebingen, Germany

[3]Department of Pediatrics, Pediatric Oncology, Lucile Packard Children's Hospital, Stanford University, Stanford, California 94304, USA

[4]Quantitative Sciences Unit, School of Medicine, Stanford University, Stanford, California 94304, USA

[*]**Co-Corresponding Authors**: Daniel Rubin, MD and Heike E. Daldrup-Link, Department of Radiology, Stanford University, 725 Welch Road, Stanford, CA 94304, USA; Email: drubin@stanford.edu and heiked@stanford.edu ; phone: (650) 497-8601




**Supplementary Fig. 1** Representative examples of the attention mask to attention-weighted loss. The attention map emphasizes the high-clinical-value regions. The top heatmap indicates tumor area and the bottom heatmap highlights the liver area. The idea is that the heatmap pixels near the mass center of the target region have high values, which smoothly but rapidly decrease farther away from the mass center

| Anatomical Regions (8) | Refined Regions (20) |
|---|---|
| CNS | CNS |
| paraspinal | paraspinal |
| neck | lymph nodes |
| | extralymphatic |
| | bone marrow |
| arms | lymph nodes |
| | extralymphatic |
| | bone marrow |
| chest | lymph nodes |
| | extralymphatic |
| | bone marrow |
| abdomen | lymph nodes |
| | extralymphatic |
| | bone marrow |
| pelvis | lymph nodes |
| | extralymphatic |
| | bone marrow |
| legs | lymph nodes |
| | extralymphatic |
| | bone marrow |

**Supplementary Table 1** The regions defined for the region-based clinical evaluation. The clinical imaging experts determined the presence or absence of tumor lesions in 8 anatomical regions including CNS (central nervous system), paraspinal, neck, arms, chest, abdomen, pelvis and legs. For the 8 anatomical regions besides CNS and paraspinal, we further divide them into lymph nodes, extralymphatic and bone marrow, resulting in 20 refined regions in total

| Procedure | TP | FP | TN | FN | Sensivitity (%) | Specificity (%) | PPV (%) | NPV (%) | Accuracy (%) |
|---|---|---|---|---|---|---|---|---|---|
| **Lymp Nodes** | | | | | | | | | |
| 100% standard-dose | 175 | 9 | 219 | 11 | 94.09 | 96.05 | 95.11 | 95.22 | 95.17 |
| 6.25% ultra-low-dose | 174 | 9 | 219 | 12 | 93.55 | 96.05 | 95.08 | 94.81 | 94.93 |
| AI-reconstructed PET | 174 | 9 | 219 | 12 | 93.55 | 96.05 | 95.08 | 94.81 | 94.93 |
| **Extralymphatic** | | | | | | | | | |
| 100% standard-dose | 36 | 12 | 360 | 6 | 85.71 | 96.77 | 75.00 | 98.36 | 95.65 |
| 6.25% ultra-low-dose | 33 | 11 | 361 | 9 | 78.57 | 97.04 | 75.00 | 97.57 | 95.17 |
| AI-reconstructed PET | **35** | 11 | 361 | 7 | 83.33 | 97.04 | 76.09 | 98.10 | **95.65** |
| **Bone Marrow** | | | | | | | | | |
| 100% standard-dose | 11 | 11 | 388 | 4 | 73.33 | 97.24 | 50.00 | 98.98 | 96.38 |
| 6.25% ultra-low-dose | 11 | 12 | 387 | 4 | 73.33 | 96.99 | 47.83 | 98.98 | 96.14 |
| AI-reconstructed PET | 11 | 12 | 387 | 4 | 73.33 | 96.99 | 47.83 | 98.98 | 96.14 |
| **Whole Body** | | | | | | | | | |
| 100% standard-dose | 225 | 32 | 1099 | 24 | 90.36 | 97.17 | 87.55 | 97.86 | 95.94 |
| 6.25% ultra-low-dose | 221 | 32 | 1099 | 28 | 88.76 | 97.17 | 87.35 | 97.52 | 95.65 |
| AI-reconstructed PET | **223** | 32 | 1099 | 26 | 89.56 | 97.17 | 87.45 | 97.69 | **95.80** |

**Supplementary Table 2** Performance of the simulated 6.25% ultra-low-dose PET, 100% standard-dose PET, and the AI-reconstructed PET for tumor detection for the Stanford cohort of 23 patients. Sensitivities, specificities, PPV (positive predictive value), NPV (negative predictive values) and balanced accuracies for the detection of pediatric malignancies are calculated across the three image modalities. There was no statistically significant difference in all metrics between different image modalities.

| Parameter | Procedure | TP | FP | TN | FN | Sensitivity | Specificity | PPV | NPV | Balanced.Accuracy |
|---|---|---|---|---|---|---|---|---|---|---|
| **Lymph Nodes** | 100% standard-dose | 28 | 0 | 38 | 0 | 100 | 100 | 100 | 100 | 100 |
| | 6.25% ultra-low-dose | 25 | 0 | 38 | 3 | 89 | 100 | 100 | 93 | 95 |
| | AI-reconstructed PET | 26 | 0 | 38 | 2 | 93 | 100 | 100 | 95 | 96 |
| **Extralymphatic** | 100% standard-dose | 4 | 0 | 62 | 0 | 100 | 100 | 100 | 100 | 100 |
| | 6.25% ultra-low-dose | 3 | 0 | 62 | 1 | 75 | 100 | 100 | 98 | 88 |
| | AI-reconstructed PET | 3 | 0 | 62 | 1 | 75 | 100 | 100 | 98 | 88 |
| **Bone Marrow** | 100% standard-dose | 5 | 0 | 61 | 0 | 100 | 100 | 100 | 100 | 100 |
| | 6.25% ultra-low-dose | 2 | 0 | 61 | 3 | 40 | 100 | 100 | 95 | 70 |
| | AI-reconstructed PET | 4 | 0 | 61 | 1 | 80 | 100 | 100 | 98 | 90 |
| **Whole Body** | 100% standard-dose | 37 | 0 | 183 | 0 | 100 | 100 | 100 | 100 | 100 |
| | 6.25% ultra-low-dose | 30 | 0 | 183 | 7 | 81 | 100 | 100 | 96 | 91 |
| | AI-reconstructed PET | 33 | 0 | 183 | 4 | 89 | 100 | 100 | 98 | 95 |

**Supplementary Table 3** Performance of the simulated 6.25% ultra-low-dose PET, 100% standard-dose PET, and the AI-reconstructed PET for tumor detection for the Tuebingen cohort of 11 patients. Sensitivities, specificities, PPV (positive predictive value), NPV (negative predictive values) and balanced accuracies for the detection of pediatric malignancies are calculated across the three image modalities. There was no statistically significant difference in all metrics between different image modalities.

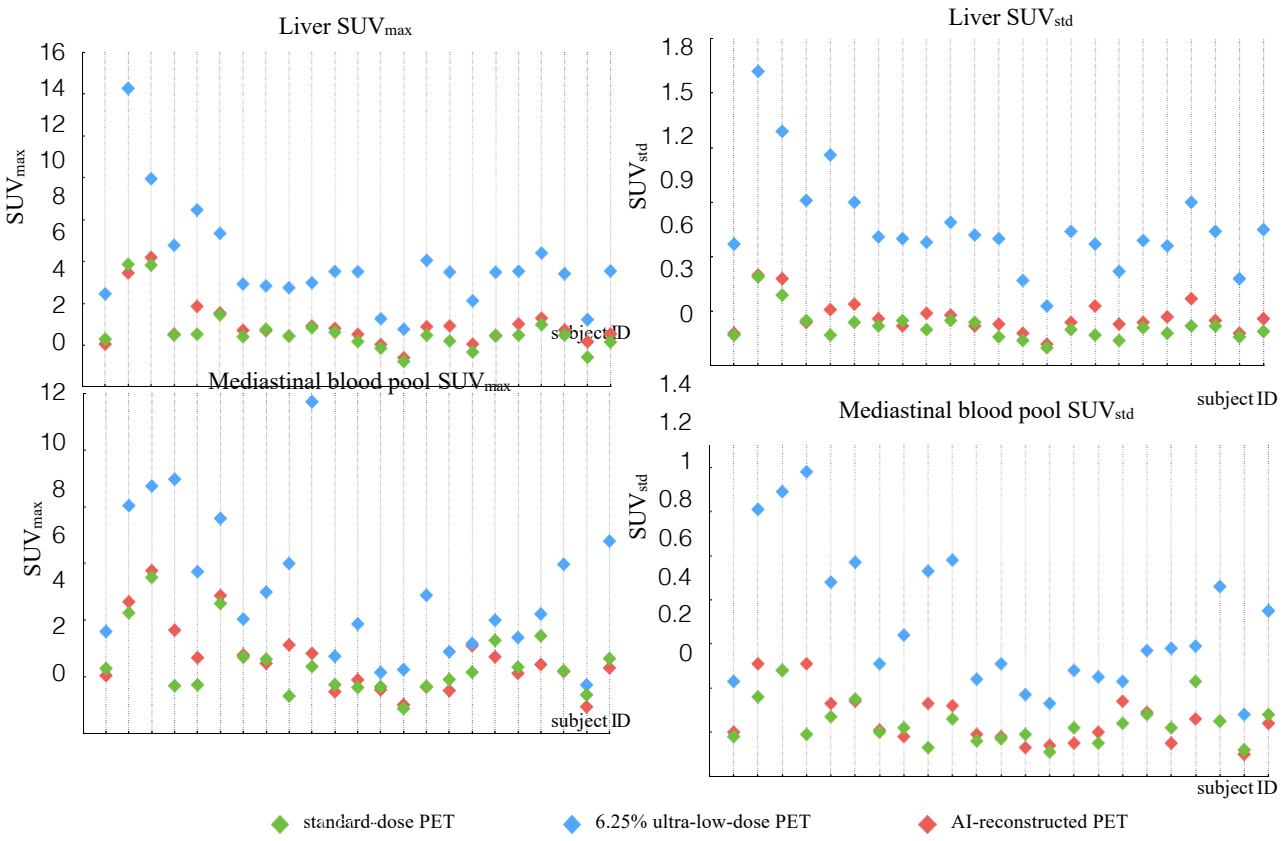
| Parameter | Procedure | Weighted Kappa (Linear) | Weighted Kappa (Quadratic) |
|---|---|---|---|
| Lymph Nodes | 100% standard-dose | 1.000 | 1.000 |
| | 6.25% ultra-low-dose | 0.887 | 0.906 |
| | AI-reconstructed PET | 0.926 | 0.959 |
| Extralymphatic | 100% standard-dose | 1.000 | 1.000 |
| | 6.25% ultra-low-dose | 0.758 | 0.790 |
| | AI-reconstructed PET | 0.849 | 0.849 |
| Bone Marrow | 100% standard-dose | 1.000 | 1.000 |
| | 6.25% ultra-low-dose | 0.316 | 0.346 |
| | AI-reconstructed PET | 0.735 | 0.787 |
| Whole Body | 100% standard-dose | 1.000 | 1.000 |
| | 6.25% ultra-low-dose | 0.834 | 0.852 |
| | AI-reconstructed PET | 0.912 | 0.937 |

**Supplementary Table 4 Agreement between expert reviewer tumor diagnoses on 100% standard-dose $^{18}$F-FDG PET and simulated 6.25% ultra-low-dose $^{18}$F-FDG PET and AI-reconstructed $^{18}$F-FDG PET scans (Tuebingen data set).** One expert reviewers determined the presence of absence of tumor lesions in 20 anatomical regions per patient according to a Likert scale (1 - tumor definitely not present, 2 - tumor probably not present, 3 - undecided, 4 - tumor probably present, 5 - tumor definitely present). The agreement between 100% standard-dose PET images and 6.25% ultra-low-dose 18F-FDG PET and AI-reconstructed 18F-FDG PET scans was calculated with weighted Kappa estimates.

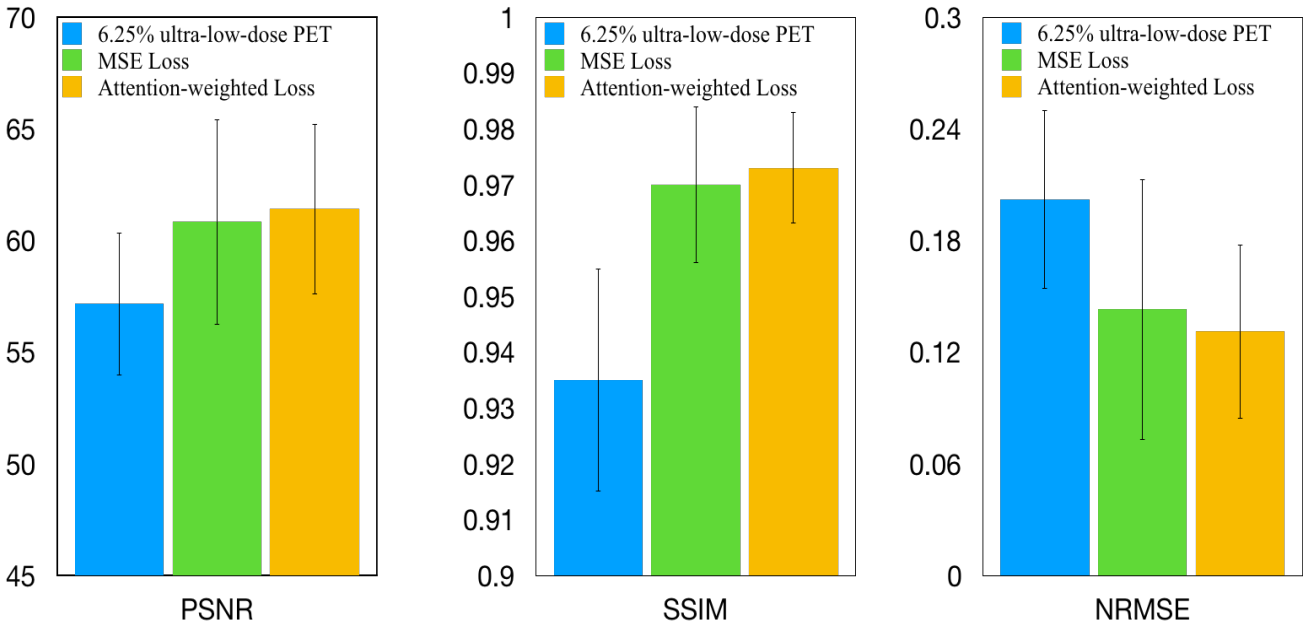
| | 100% standard-dose PET (N=37) | 6.25% ultra-low-dose PET (N=30) | AI-reconstructed PET (N=33) |
|---|---|---|---|
| **Lesion.SUVmax** | | | |
| Mean (SD) | 7.28 (4.97) | 9.66 (5.33) | 6.76 (4.88) |
| Median [Q1, Q3] | 5.72 [3.58, 8.48] | 8.46 [6.20, 10.6] | 5.30 [3.95, 8.13] |
| **Lesion.SUV.mean** | | | |
| Mean (SD) | 4.45 (3.61) | 4.70 (3.36) | 4.33 (3.37) |
| Median [Q1, Q3] | 3.52 [2.24, 5.49] | 3.85 [2.85, 5.34] | 3.60 [2.32, 4.94] |

| | 100% standard-dose PET (N=11) | 6.25% ultra-low-dose PET (N=11) | AI-reconstructed PET (N=11) |
|---|---|---|---|
| **Liver.SUV_max** | | | |
| Mean (SD) | 2.13 (0.492) | 4.07 (2.36) | 1.99 (0.601) |
| Median [Q1, Q3] | 2.06 [1.87, 2.40] | 3.81 [3.01, 4.46] | 2.04 [1.78, 2.30] |
| **Liver.SUV_mean** | | | |
| Mean (SD) | 1.53 (0.273) | 1.40 (0.448) | 1.40 (0.398) |
| Median [Q1, Q3] | 1.55 [1.46, 1.64] | 1.42 [1.24, 1.64] | 1.40 [1.27, 1.60] |
| **Liver.SUV_std** | | | |
| Mean (SD) | 0.194 (0.0644) | 0.642 (0.281) | 0.198 (0.0846) |
| Median [Q1, Q3] | 0.190 [0.150, 0.225] | 0.600 [0.505, 0.740] | 0.210 [0.145, 0.235] |

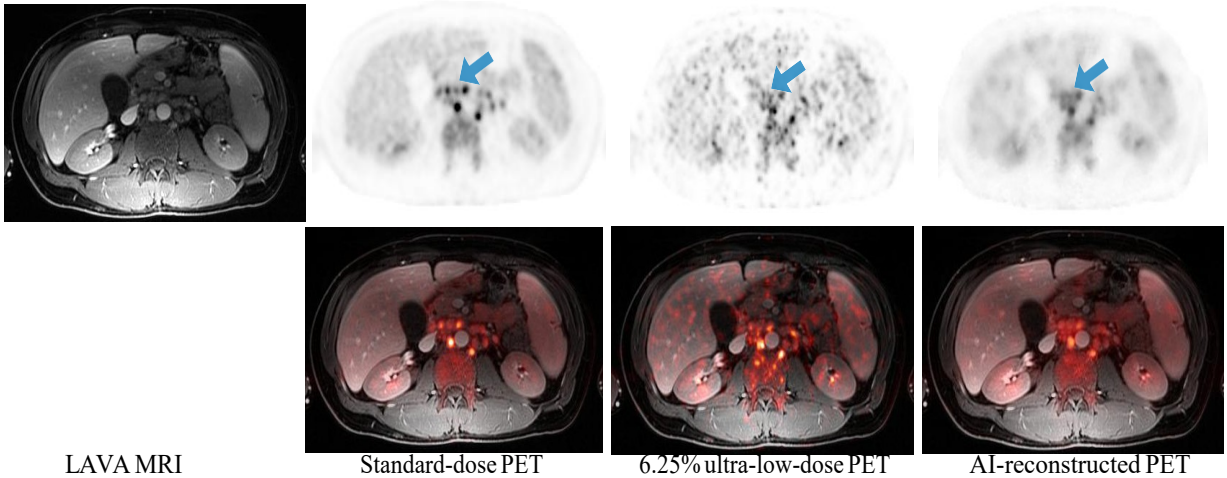
**Supplementary Table 5. Standardized uptake values (SUV), as measured on 100% standard-dose 18F-FDG PET, simulated 6.25% ultra-low-dose 18F-FDG PET and AI-reconstructed 18F-FDG PET scans.** Data represent mean and median SUVmax, SUVmean and SUVstd values of representative tumors (numbers of evaluated tumors are listed under each modality) and the liver (n=11 for each modality).




**Supplementary Fig. 2** *SUVmax* and *SUVstd* distribution at liver and mediastinal blood pool per different baseline scans (23 in total) and different PET modalities (3 in total) - standard-dose PET, 6.25% ultra-low-dose PET and AI-reconstructed PET. a). *SUVmax* at liver; b). *SUVstd* at liver; c). *SUVmax* at mediastinal blood pool; d). *SUVstd* at mediastinal blood pool; The simulated 6.25% ultra-low-dose PET has significantly higher *SUVmax* and *SUVstd* due to increased image noise and poor image quality. The AI-reconstructed PET recovers the image quality and retains *SUVmax* and *SUVstd* to the level of the ground truth standard-dose PET




**Supplementary Fig. 3** The quantitative performance comparison with and without attention-weighted loss. Image quality metrics - peak signal-to-noise ratio (PSNR), structural similarity (SSIM) and normalized root mean square error (NRMSE) - compare images from 6.25% ultra-low-dose PET, the AI-reconstructed PET by traditional mean square error (MSE) loss model, and the AI-reconstructed PET by the attention-weighted MSE loss model. For the three metrics, the comparison is to the ground truth standard-dose PET. The reconstructed images generated from the attention-weighted MSE loss model is superior for all three metrics




**Supplementary Fig. 4** Limitations of the study. PET/MRI scan of a 30-year old male patient with Hodgkin lymphoma (HL). Some small lymph nodes are less well delineated on the AI-reconstructed PET compared to the original standard-dose PET. The blue arrow points to the sub-centimeter hypermetabolic lymph nodes which were better delineated on the AI reconstructed scan than on the 6.25% dose scan. The AI reconstructed scan does not discriminate each small individual lesion as well as the original 100% standard-dose scan