Results

Quantitative analysis

Cross-Centre Results:

The two proposed DL algorithms were evaluated in this study on 68Ga-PET dataset (IMCM and ADCM). We tested the trained DL model with two internal and external test-set to evaluate the robustness, which internal test-set include 8 subjects from 4 different centres as external test-set and 12 subjects from an external non-seen centre.

Figure 3 provided quantitative accuracy of the DL attenuation scatter corrected PET ((DL ASC-PET) images to the ground-truth CT-based attenuation scatter corrected PET (CT ASC-PET) images on internal and external centres. As shown, both DL methods were capable of some degree of attenuation and scattering correction for internal and external centres. Center-wise analysi are available in supplemtary material.

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |

Figure ? in appendix 2, shows the relative and Absolute relative percentage error of entire image on all centers, calculated in reference to the CT ASC-PET, which demonstrates that IMCM performed the other on all scanners regarding local metrics as well. Results of peak signal-to noise ratio (PSNR) and structural similarity index measurement (SSIM) showed the same tendency as the ? results.

For the external center, ADCM yielded a ME of -0.631±0.965 (CI 95%: -1.23 to -0.03), MAE of 3.072±1.012 (CI 95%: 2.815 to 3.329), and a RE of -8.139±27.364% (CI 95%: -21.76 to 5.48). In contrast, the IMCM demonstrated improved consistency with a ME of -1.835±1.387 (CI 95%: -2.80 to -0.87) and a MAE of 2.588±0.931 (CI 95%: 2.386 to 2.790).

Internal centers analyzed collectively showed ADCM produced a ME of 0.373±1.455 (CI 95%: -0.55 to 1.30) and a MAE of 2.343±0.768 (CI 95%: 2.191 to 2.495). While IMCM showed lower ME of -0.364±0.841 (CI 95%: -0.76 to 0.03) and MAE of 1.415±0.327 (CI 95%: 1.360 to 1.470).

PSNR also favoured IMCM method, registering at 35.526±2.117 (CI 95%: 34.9 to 36.2) compared to 38.251±1.923 (CI 95%: 37.6 to 38.9) for the ADCM method. Notably, SSIM for IMCM at external centre was superior, recorded at 0.879±0.020 (CI 95%: 0.871 to 0.887). Detailes are available at supplementary material

In addition to voxel-wise assessments, model performance was further validated through various statistical tests, which compared image-derived metrics between different training models. The Wilcoxon test used here due to the non-normal distribution of the data as evidenced by Shapiro-Wilk tests.

The results of Wilcoxon test with the False Discovery Rate (FDR) method correction revealed significant differences for all metrics except for Relative Error across the ADCM and IMCM datasets. In other word, corrected p-values using the Benjamini-Hochberg procedure, indicated notable discrepancies in error measurements and image quality between the methods. except for the "Relative Error (SUV%)" where the corrected p-value does not indicate a statistically significant difference threshold of 0.05. IMCM showing consistently lower errors, higher PSNR, and higher SSIM values, indicating superior image quality and more reliable estimations.These findings are detailed further in the table1 in supplementary 1:

The joint histogram analysis, including Pearson correlation, visually represented the voxel-wise SUV correlations across different centers for both methods (Figures provided) showcased a discernible variance in the predictive accuracy and linearity in SUV estimation. The external center distinctly demonstrated this variance. IMCM depicted a regression slope of 0.65±0.02, R=0.949, indicating a consistent underestimation across the predicted SUV values when compared to the reference. Conversely, ADCM at the same center exhibited a slope of 1.18±0.10, R=0.901, suggesting a tendency towards overestimation, likely influenced by very high SUV values that might not be clinically advantageous.

In internal centers, both methods continued to show divergent behavior. IMCM maintained a closer approximation to ideal predictive accuracy, particularly notable at center C3 with a regression slope of 0.87±0.01, R=0.990, pointing towards an excellent match with reference SUVs. In contrast, the ADCM method presented slopes greater than one (C2: 1.13±0.03, C4: 1.19±0.03), indicating a potential over-calibration issue leading to systematic overestimations of SUV.

Voxel-wise analysis further substantiated these findings, revealing higher discrepancies in centers where ADCM predicted significantly higher values. These observations suggest that while ADCM may seem to offer a better fit in certain centers due to closer R values, the inflated SUV predictions cast doubts on its reliability and clinical utility. The IMCM provided image quality comparable with MAC and preserved more detailed information and less noise was observed compared to   
ADCM.

|  |  |
| --- | --- |
| Predicted (SUV) |  |
| Predicted (SUV) |  |
| Predicted (SUV) |  |
| Predicted (SUV) |  |
| Predicted (SUV) |  |
|  | Reference (SUV) Reference (SUV) |

Figure 1:

Cross-Tracer Results:

As part of our assessment of generalization capabilities across different tracer types, IMCM, was initially tested without specific tuning for cross-tracer variations. The results revealed that the IMCM, without prior tuning, struggled to maintain its efficacy when applied to different tracer types.

This outcome contrasts sharply with the claims from the ADCM approach, which posits that their model architecture inherently accommodates variations across tracers and anatomical structures without the need for additional adjustments.

Figure X showcases a sample coronal slices of IMCM, TL-MC and ADCM on cross-tracer subjects. The significant drop in accuracy and increased error rates highlight the challenges in achieving robust cross-tracer generalization with a single, unified model approach. These findings underscore the necessity for targeted model tuning to adapt to the unique characteristics of each tracer, thereby enhancing the model's applicability and effectiveness in diverse clinical settings.

A comparison of a person's body

Description automatically generatedA comparison of images of a human body

Description automatically generated

A comparison of images of a person's body

Description automatically generatedA comparison of images of a dog

Description automatically generated

The two approaches,TL-MC (Tuned version of IMCM) and ADCM evaluated in this part for FDG imaging for two centres and indicates a significant difference between the ADCM and IMCM techniques.

Both ME and MAE indicated much smaller error margins for the TL-MC, with the overall mean values reflecting better accuracy than the ADCM. The TL-MC ME deviated narrowly by, -0.10±0.76, while the ADCM deviated by 0.82±0.70, signifying a much wider spread of the SUV estimates.

These are shown as RE%. This also agree that TL-MC had a better performance. RE spread was relatively lower for TL-MC, averaging at 30±50%, in contrast with ADCM, where the spread was a lot broader at 50±100%.

TL-MC gave a lower RMSE of 2.0 ± 0.6, which pointed out consistency and reliability in comparison to ADCM's 3.2 ± 1.1. It was also better than ADCM in the metrics of image quality by higher values of PSNR and higher SSIM, which showed tighter control over noise and structural fidelity.

Taken together, these findings point toward superiority in the use of TL-MC over ADCM in terms of accuracy and consistency in all major key PET imaging metrics, and the use of this approach is recommended in clinical practice where precision is critical. Data make a compelling case that TL-MC should be preferred with respect to strong performance in consistently keeping lower errors in the images. For a comprehensive view and deeper analysis, refer to the box plots in Figure X. Detailed statistical comparisons of these metrics are illustrated in the supplementary material 2, table X provided.

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |

Figure 2: Comparative Analysis of Imaging Metrics Between ADCM and IMCM Methods. The box plots depict the distribution of Mean Error (SUV), Mean Absolute Error (SUV), Relative Error (SUV%), Absolute Relative Error (SUV%), Root Mean Squared Error, Peak Signal-to-Noise Ratio, and Structural Similarity Index across centers C6 and C7.

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |

|  |  |
| --- | --- |
| A close-up of a person's body | A close-up of a person's body  Description automatically generated |
| A close-up of a hand scan  Description automatically generated | A close-up of a person's body |
| A close-up of a person's body  Description automatically generated | A close-up of a scan of a person  Description automatically generated |
| A close-up of a person's body  Description automatically generated | A close-up of a person's body  Description automatically generated |

Qualitative assessment of artefacted images

|  |  |
| --- | --- |
| A close-up of a scan of a dog  Description automatically generated | A close-up of a scan of a person  Description automatically generated |
| A close-up of a person's body  Description automatically generated | A comparison of a baby body  Description automatically generated with medium confidence |
| A close-up of x-ray images  Description automatically generated | A close-up of a person's body  Description automatically generated |
| A close-up of a person's body  Description automatically generated | A close-up of x-ray images  Description automatically generated |
|  |  |
|  |  |
|  |  |
|  |  |

|  |
| --- |
| A comparison of a person's body  Description automatically generated |
| A close-up of a pair of images  Description automatically generated |
| A close-up of a scan of a person  Description automatically generated |
| A close-up of a pair of eyes  Description automatically generated |
| A close-up of x-ray images  Description automatically generated |
| A close-up of a pair of eyes  Description automatically generated |
| A close-up of a person's body  Description automatically generated |
| A close-up of a person's body  Description automatically generated |
| A close-up of a pair of eyes  Description automatically generated |
| A comparison of a person's body  Description automatically generated |
| A close-up of a pair of eyes  Description automatically generated |