Given the ADCM method's focus on decomposing the PET image correction process into anatomy-independent textures and anatomy-dependent corrections, we expect that

This method's core innovation of estimating anatomy-dependent corrections while implicitly referencing its deep learning foundation and the specialized focus on anatomical nuances.

This is especially helpful if the data are private or sensitive and cannot be shared with a central server or other hospitals

The normalization of ADCM in this study represents a departure from conventional practices, driven by the imperative to uphold the interpretative power of SUVs in a clinical context. By selecting a unique normalization constant, we navigated the challenges presented by the heterogeneous scale of ADCM values—foregoing the typical min-max approach that would compromise quantitative depth for the sake of uniformity. The careful exclusion of extreme outliers underscored our commitment to model precision and robustness. This approach highlights the importance of customizing data preprocessing methods to the specific demands of the dataset and the analysis objectives, which, in this case, are governed by the exigencies of clinical applicability and diagnostic accuracy.

In our study, we applied the ADCM method to decompose PET image correction into anatomy-independent textures and anatomy-dependent corrections. Unlike the referenced method which utilized a resolution of 6.6 × 6.6 × 8 mm/voxel, our images were processed at a higher resolution of 4 × 4 × 3 mm/voxel (see Figure 5). This increase in resolution, although advantageous for capturing finer anatomical details, presented its own set of challenges, particularly in generating reliable anatomy-dependent textures in some instances.

Theoretically, higher resolution should enhance the ADCM's ability to delineate anatomy-dependent corrections. However, we observed that the ADCM occasionally failed to exhibit these textures distinctly, particularly in cases with lower radioisotope uptake or in regions where anatomical variations are subtle but critical. This limitation suggests that while our ADCM model benefits from higher resolution data, it does not consistently translate into improved delineation of anatomy-dependent features, which are crucial for accurate PET image correction.

This inconsistency might be attributed to several factors, including the inherent variability in radioisotope distribution and the complex interaction between image resolution and the deep learning model's ability to learn from such data. Given these observations, we cannot expect uniformly ideal results from the ADCM approach across all scenarios. The challenge remains to enhance the model's sensitivity to subtle anatomical details without compromising the overall accuracy, suggesting a potential area for future research to refine the approach and address these specific issues.

Our findings underscore the importance of tailoring the ADCM methodology to suit specific clinical scenarios and highlight the need for ongoing adjustments and validations to ensure its effectiveness across varying conditions. This underscores the need for further research incorporating a broader array of data conditions to better understand the limitations and optimize the technique for clinical application.

The approach to ADCM normalization adopted in this study was informed by the necessity to preserve the clinical significance of SUVs. This led to the selection of an empirical normalization constant, circumventing the use of standard min-max normalization, which could diminish the quantitative richness essential for clinical interpretation. The omission of extreme outliers further emphasizes the precision and robustness of our analytical model. The histograms in Figure X depict the effective normalization, capturing the essential data distribution while excluding the extremes, thereby underscoring the significance of tailoring preprocessing methods to the dataset's unique characteristics and the analysis goals—paramount among these being the exigencies of clinical utility and diagnostic precision.

The results from the joint histogram analysis raise pertinent questions about the calibration and reliability of the ADCM method in clinical settings. The observed overestimations by ADCM, particularly in the external center (C5), could potentially lead to misinterpretations in clinical diagnostics, where precision in SUV estimation is crucial. The systematic bias towards higher SUV values, although providing a superficial appearance of accuracy due to closer R-values to unity, suggests an underlying issue in the algorithm or its application across different PET systems.

Furthermore, the IMCM method, with its more consistent adherence to lower regression slopes and higher correlation coefficients, especially in internal centers, supports its suitability for clinical use by providing reliable SUV estimations. The variance in predictive performance between IMCM and ADCM highlights the necessity for rigorous validation of imaging algorithms to ensure they perform uniformly across different settings.

This discussion emphasizes the need for comprehensive assessments of PET imaging methods to ensure that they not only align statistically with expected outcomes but also adhere closely to clinical realities. Future work should aim to refine these methods, particularly focusing on eliminating biases and enhancing the predictive accuracy of techniques like ADCM to leverage their full potential in patient care and treatment planning.

Our decision to use the Wilcoxon signed-rank test was validated by the significant results obtained, which are critical for understanding the performance disparities between ADCM and IMCM. To ensure the robustness of our findings given multiple comparisons, we applied the False Discovery Rate (FDR) correction using the Benjamini-Hochberg procedure. This method adjusts for potential increases in Type I errors, providing more reliable p-values that enhance the validity of our conclusions about the differences between ADCM and IMCM. Detailed statistical methodologies and results are documented in supplementary material 2.

(Benjamini & Hochberg, 1995)

Future work:  
Furthermore, we can measured clinical imaging parameters such as SUVmean, SUVmax, total lesion metabolism, Table 1 | Information on patients as well as the most relevant radiomics features within the sphere

In future investigations, it will be crucial to assess the performance of the IMCM model specifically on artifact images, focusing on organ-specific evaluations. This approach would provide a more detailed understanding of the model's effectiveness in varied clinical scenarios. Additionally, conducting comprehensive statistical tests such as the Marginal Homogeneity Test or the McNemar test on categorized outcomes could offer deeper insights into the consistency and reliability of the model across different diagnostic categories. Such analyses will help refine the model's application and enhance its diagnostic accuracy in real-world settings.

The Multi-Center method exhibited enhanced performance compared to the ADCM method across various quantitative metrics, highlighting its effectiveness in providing more accurate and reliable PET imaging. This superior performance, especially noted in the consistency of Relative Errors and Absolute Relative Errors, underlines the potential of the Multi-Center method for clinical applications where precise imaging is critical.

Significant differences were observed between the methods when evaluated using the Wilcoxon test, with the Multi-Center method consistently showing reduced error magnitudes and higher structural similarity across all centers. These results highlight the Multi-Center method's superiority in precision and reliability for clinical PET imaging applications.

This structured and formal analysis underscores the efficacy of the Multi-Center imaging approach over ADCM, providing a clearer understanding of each method's performance across different centers, which is crucial for optimizing PET imaging protocols in clinical settings.

The analysis across both centers highlights a general trend: while the TL-MC model offers a more reliable and consistent estimation close to actual SUV values, it underestimates at C7, suggesting a potential calibration need. In contrast, the ADCM model, with its overestimation and lower correlation, may require adjustments to enhance its predictive accuracy and clinical utility. These findings emphasize the importance of model calibration to improve the reliability of SUV estimations for clinical applications, with each model showing distinct areas for enhancement.

We included disease-free and pathological patients with various indications, such as age, weight and disease type (for training and independent validation sets) resulting in a heterogeneous database. This study naturally bears a number of limitations. First, the training and independent validation processes were performed on only 18F-FDG as a tracer. For different radiotracer distributions, the network will need to be retrained on PET images using other radiotracers. However, the network trained on 18F-FDG PET images could be used to initialize networks trained on PET images acquired with other radiotracers via transfer learning [50]. This will also help to address the issue of having a limited training dataset. Second, all images were acquired on a PET/CT scanner. Hence, the network has not yet been validated for PET/MRI applications since this hybrid imaging uses rigid and surface MR coils (invisible in PET images) in the field-of-view. This will add challenges to the Deep-JASC approach, in particular for nonrigid surface coils as they may not impact PET-nonASC images similarly/repeatability across the subjects. The network should be validated for PET/MRI. Yet, PET/CT images included the CT couch, yet the network performed reasonably well.

Shibe balaye ADCM to FDG:  
This comparative analysis brings to light the critical aspect that a higher slope does not necessarily equate to better correlation or prediction accuracy. Instead, the consistency with which predictions align with actual values, as measured by correlation coefficients, provides a more substantial indication of a model's effectiveness. The TL-MC model, with its tighter adherence to the regression line despite lower slopes, ultimately demonstrates a more reliable and consistent performance in capturing the true behavior of SUVs across the studied centers.