

## Deep Learning-Based PET Image Correction Toward Quantitative Imaging



**Zohreh Shahpouri**

Supervisor: Isaac Shiri Lord, Ph.D.

Institute: Artificial Intelligence in Cardiac Imaging Laboratory, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland

# Introduction

Positron Emission Tomography (PET) is a key molecular imaging modalities used during in vivo studies for the assessment of various diseases in a non-invasive manner (1–3). The use of PET is important in the clinical oncology field, including diagnostic, staging, restaging, assessment of therapeutic response, and radiation therapy planning (4–6). Artefact minimization and high-quality imaging are, therefore, imperatives for the role they play in the qualitative interpretation and quantitative analysis of PET scans (7,8).

Artefacts in PET imaging are anomalies in the final images that do not correspond to the true distribution of the radiotracer within the body. These can be caused by a variety of factors, including patient motion, improper scanner calibration, and physiological processes that interfere with signal acquisition. Artefacts can lead to misinterpretations in clinical diagnosis, making it essential to identify and correct them to enhance the accuracy of PET scans (9–12).

Still, artefacts in medical imaging are a recurring challenge that can lower the quality of images and make them less reliable. This can lead to erroneous interpretations that could adversely influence clinical decisions (13,14).

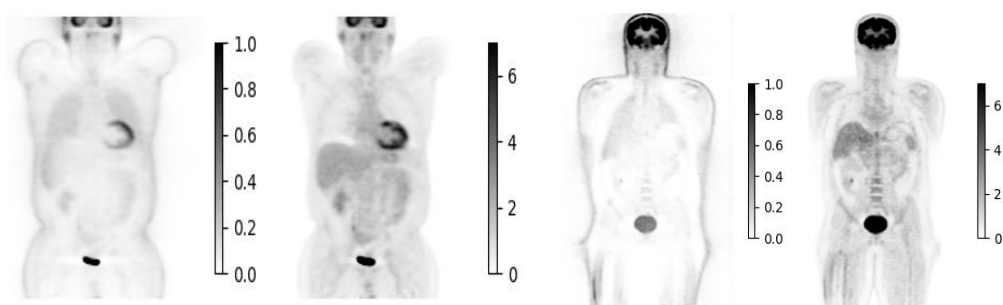
Scatter correction is a technique used to enhance the quality of PET images by removing scatter radiation that can blur images and obscure details. During a PET scan, photons emitted by the radiotracer can scatter as they collide with other particles before reaching the detectors (15–17). This scatter distorts the image by introducing signals from incorrect locations. Scatter correction algorithms estimate the number of scattered photons and subtract them from the detected signals, thereby improving image clarity and contrast (15).

Attenuation correction (AC) is another critical process in PET imaging, which compensates for the loss of signal intensity due to the absorption of photons within the body. Different tissues absorb photons at varying rates, which can lead to underestimation of tracer concentration in areas like bones or organs with higher densities. AC uses information from a transmission scan (using either a radioactive source or a CT scan) to accurately map the absorption properties of various tissues and adjust the PET signal accordingly (18,19)). This correction is crucial for providing quantitatively accurate images that reflect the true distribution of the radiotracer (13,14,20–23).

Attenuation and scatter correction (ASC) are critical during PET image reconstruction, primarily aimed at enhancing image clarity and accuracy. Despite the implementation of these corrections, artefacts can still occur, particularly under complex scenarios such as high radiotracer activity or patient movement (7,9,11,12). For an illustrative example, see Figure 1. Common artefacts encountered in PET imaging can be categorised as follows: (a) those associated with the distribution of the tracer, such as halo artefacts; (b) those that arise from the alignment of PET with Computed Tomography (CT) or MR images, including mismatch, misregistration, or motion artefacts; and (c) those transmitted from CT or MRI to PET images, such as errors caused by metals, contrast agents, and image truncation (9,10,24,25).

Halo artefacts are very common in PET imaging, especially with compounds that are labelled with gallium-68 ( $^{68}\text{Ga}$ ). They make it hard to correctly interpret high-activity regions adjacent to organs. In fact, these are a type of radiopharmaceutical artefact that happens when too much radiopharmaceutical builds up and makes it harder to see what's going on in nearby tissues (26,27).

These artefacts are primarily induced by incorrect scatter correction during image reconstruction, where negative values near regions of intense radiopharmaceutical accumulation—such as the bladder or kidneys due to urinary excretion of the tracer—lead to the assignment of zero values to these voxels due to the non-negativity constraint in statistical reconstruction algorithms. This phenomenon results in the formation of a "halo" or photogenic area around these high-activity zones, potentially obscuring faint abnormalities and impacting the diagnosis, staging, and treatment planning for cancer patients. When halo artefacts are present on PET images, especially near primary tumours or areas where pelvic cancers tend to come back locally, they can lead to a wrong diagnosis because they hide or change how the images are seen and interpreted quantitatively. Trying to get rid of these artefacts, like giving diuretics, often makes the patient more uncomfortable and increases the chance of motion artefacts, which makes the image quality and readability even worse (26,28)).



*Figure 1: showcases examples of PET images before and after attenuation and scatter correction.*

Most PET acquisition settings are performed with arms up (to decrease photon scatter). As arm raising is uncomfortable for patients, this will cause arm motion during sequential PET and CT/MRI scans. This is one example of the mismatch effect (29–31).

Mismatch artefacts constitute a significant challenge in PET imaging, particularly when discrepancies arise between PET and anatomical scans such as CT or MRI. These discrepancies can stem from both voluntary and involuntary movements of organs, potentially leading to the misidentification, mis-localization, and inaccurate quantification of lesions. This issue is critical as it can result in misdiagnoses and, subsequently, inappropriate patient management (32–34). Techniques such as deformable image registration have been developed to mitigate these effects, though they cannot always eliminate the problem. Mismatched artefacts mean that PET imaging needs to be done in a more complex way, using different CT acquisition protocols and other techniques to reduce the chance of misalignments (35,36).

Truncation artefacts in PET imaging emerge primarily due to the disparities in the trans axial fields of view (FOVs) between PET and CT/MRI modalities. These artefacts are particularly prevalent in scenarios involving obese patients or when patients have their arms down during the scanning process, as well as in cases where PET/CT or PET/MR scans are utilised for treatment planning (37–39). Truncation artefacts happen when there aren't any matching parts of the attenuation map for structures that go beyond the CT/MR images. This makes standardised uptake value (SUV) estimates wrong, usually giving too high of an estimate around the edges and too low of an estimate in the middle of the image. The issue is compounded when anatomical images truncate parts of the patient's body, leading to artefacts and distorted activity quantification in PET images. Optimally positioning the patient in the centre of the FOV with arms-up can decrease such artefacts, yet specific conditions, like scanning for melanoma or head-neck cancer, necessitate arms-down positioning. To fix or lessen truncation artefacts, different methods have been tried, such as extended FOV CT scans, extrapolation of CT projections, specialised MR sequences, and manual or semi-automatic in-painting algorithms. Despite these efforts, it is still hard to deal with truncation artefacts, especially in obese patients where more photons are attenuated and scattered, making the image quality and quantitative accuracy even worse (38–40).

Halo and mismatch artefacts are notably frequent in PET imaging using gallium-68 ( $^{68}\text{Ga}$ )-labelled radiopharmaceuticals. These artefacts might be overlooked if they are subtle, yet when pronounced, they can significantly degrade the image quality, necessitating additional scans. However, even repeated scanning often fails to correct these artefacts, as they are sometimes inherent and unavoidable in specific situations (11,12,41,42).

ASC techniques are necessary for making reconstructed and quantitative PET images, but they also cause some artefacts, which means that PET imaging needs to be done in a more nuanced way. Understanding the limitations and potential pitfalls of these techniques is crucial for radiologists and clinicians to interpret PET images accurately. Ensuring meticulous calibration, considering patient-specific factors, and using

advanced correction algorithms are essential steps in minimising the impact of these artefacts on clinical outcomes.

Integration into CT or MRI is necessary for quantitatively accurate and visually readable PET images. ASC are required to make PET images that are both quantitatively accurate and easy to read visually (13,43). Typically, an unenhanced, low-dose CT scan is conducted alongside PET/CT scans for ASC, and occasionally, a diagnostic CT scan with a contrast agent may serve the same function (44,45). Elimination of the CT component could be particularly beneficial for patients requiring repeated PET/CT scans, notably paediatric patients, as even marginal reductions in cumulative radiation exposure are of significance (46,47).

The integration of CT in PET/CT imaging, while invaluable for AC and precise anatomical localization, significantly contributes to the total ionising radiation dose received by patients. Innovations such as long axial field of view (LAFOV) total-body PET scanners have markedly improved image resolution and quantification while reducing the need for high radiopharmaceutical doses (16–18,21,22). Nonetheless, the aspiration for entirely CT-free PET imaging methodologies is driven by the imperative to diminish radiation exposure in vulnerable populations and during repeated examinations or longitudinal studies (48).

Historically, the field has seen progress through magnetic resonance (MR)-based strategies and algorithmic advances such as the maximum likelihood estimation of activity and attenuation (MLAA), further refined with time of flight (TOF) enhancements (49–51). Despite these developments, the interplay between activity distribution and attenuation remains a challenging frontier, compounded by scanner-specific noise and resolution discrepancies (23).

Deep learning (DL) has emerged as a groundbreaking approach in PET imaging, revolutionising AC, and artefact reduction (52–57). DL-based methods have been developed for a variety of applications, including the synthesis of pseudo-CT images from MRI or uncorrected PET data, prediction of scatter maps from emission data (36,39,58–61), and direct generation of ASC PET images from uncorrected inputs (53,57,62). These advancements not only demonstrate the vast potential of DL in enhancing the safety and efficacy of PET imaging but also highlight its capability to improve image quality significantly and reduce errors caused by metal artefacts and truncation effects in both PET/CT and PET/MRI modalities (37,63).

Deep learning presents a promising paradigm capable of transcending traditional challenges in PET imaging, such as the activity-attenuation crosstalk and the noise inherent in the imaging process (36,55). However, the success of DL critically hinges on its adaptability to the dynamic nature of PET tracers, particularly  $^{68}\text{Ga}$ , and the variability across imaging platforms. This includes dealing with differences in scanners,

creating new tracers that have specific biodistributions, and the fact that PET imaging domains are naturally diverse (35,64).

Furthermore, the use of federated learning (FL) addresses critical challenges such as data privacy and limited dataset sizes in medical imaging. FL lets DL models be trained in different places without compromising data privacy. This makes the models more stable and able to adapt to differences in data and imaging protocols (11,12,42,65). This approach is invaluable in environments where data sharing is restricted by ethical and regulatory considerations.

Yet, the quest for CT-free PET imaging avenues, particularly beneficial in paediatric scans, repetitive examinations, and pharmaceutical research, underscores the need for novel correction techniques devoid of additional radiation risks. However, there are still some problems with how well DL approaches work in PET imaging. This is why we need to create a DL model that doesn't depend on the centre or tracer used, so it can be used for all  $^{68}\text{Ga}$ -PET imaging problems. As the technology progresses, further research is necessary to address the emerging challenges, particularly in adapting DL models to handle the rapid advancements in imaging technologies and tracer development. The ongoing refinement of these innovative methodologies will be crucial for achieving widespread clinical acceptance and enhancing the diagnostic capabilities of PET imaging (52).

One significant challenge for deep learning methods in PET imaging is the struggle to adapt to the inherent heterogeneity across various domains. Variations in spatial resolution and sensitivity among scanners, coupled with the ongoing changes in biodistributions of new tracers, pose challenges in developing a comprehensive training dataset. This diversity can compromise the consistency and reliability of deep learning approaches for AC, making it challenging to capture the full range of variables in real-world situations.

Studies in the past have shown that direct ASC frameworks can help fix artefacts in  $^{18}\text{F}$ -FDG PET/CT images. However, gallium-based PET images often exhibit lower quality and resolution, potentially due to their unique characteristics and interactions within the body. As a result, these images require more nuanced approaches to ensure accurate interpretation and analysis (36,65).

This thesis will try to take a step into the problematic field of correction in PET imaging artefacts, with especially high-prevalence ones: mismatch and halo artefacts in  $^{68}\text{Ga}$  PET imaging. The aim of this paper is to look at several deep learning models and methodologies to design a multi-centre model that allows semi- and un-direct data sharing at each centre due to some demerits of conventional deep learning techniques. This is made possible using a novel deep learning architecture called Dyn-Unet with a sophisticated 3D convolutional capability, allowing for precise disentangling and correction of artefacts.



We will use our approach to estimate and compare the performance of models under both strategies within different levels of tracer dynamics and multi-centre data environments. In particular, we will integrate domain expertise into our deep learning framework in order to detect and correct artefacts more efficiently in multi-centre studies.

This research will aim to be a demonstration of the possibility, as well as the superior performance, of the deep learning models for real clinical settings, which will potentially set a new standard of CT-free PET imaging that enhances diagnostic accuracy while minimising radiation exposure and procedural complexity.

# Material and methods

## Data Preparation

Our study aimed to evaluate the performance of our model across different scenarios, including various external scanners and radiotracers. Multiple hospital imaging centres in five different places were used as a primary dataset for training and initial model validation. The dataset contained Gallium PET/CT scans. To test the model's adaptability, a secondary dataset was incorporated, distinct in both the imaging centres (external to the primary dataset) and the type of radiotracer used ( $^{18}\text{F}$ -FDG PET scans from two different hospitals). Additionally, a specialised set of images presenting artefacts was included to assess the model's capability to identify and correct for image quality issues. The Ethics Committee of the Geneva University Hospital approved this retrospective study, which spans across several institutions.

### Gallium PET/CT dataset

A cohort of more than 1000 patients underwent  $^{68}\text{Ga}$ -prostate-specific membrane antigen (PSMA)/DOTA-TATE (TOC). PET/CT imaging across five centres located in different countries. To ensure the integrity of the data for model training, an expert in nuclear medicine evaluated all the scans, identifying 184 images of optimal quality without artefacts from the total pool. Detailed information on the datasets collected from the various locations is outlined in Table 1. The method of CT-based ASC was applied to amend PET images for accurate correction of attenuation and scatter effects in the images. For this study, non-attenuation corrected images will be referred to as NAC, and CT-based attenuation scatter corrected images will be denoted as MAC.

Table 1: Data information in 5 different imaging centers.

Center	No	Train	Validation	Test	Scanner	Reconstruction	Matrix size $\times Z^*$
Center 1	56	43	11	2	Siemens Biograph 6	3D-OSEM	168 $\times$ 168
Center 2	31	25	4	2	GE Discovery IQ	3D-OSEM	192 $\times$ 192
Center 3	45	35	8	2	Siemens mCT	3D-OSEM	200 $\times$ 200
Center 4	40	28	10	2	Siemens Biograph 6	3D-OSEM	168 $\times$ 168
External Center	12	-	-	12	Siemens Horizon	PSF+TOF+3D-OSEM	180 $\times$ 180
<b>Total</b>	<b>184</b>	<b>131</b>	<b>33</b>	<b>20</b>	-	-	-

\*  $Z'$  representing the number of slices in the axial view, depends on body length, scanner resolution, scan protocol, and patient positioning. So, it is different patiently.



## Normalization of PET Imaging Data

In PET imaging, the SUV is a crucial quantitative measure that normalises the detected radiotracer concentration in a way that allows comparison between patients and scans. It corrects for the injected dose of the radiotracer and the patient's body weight. This conversion is essential as it factors in variations due to patient size and the amount of radiotracer administered. The SUV is typically calculated using the formula:

$$\text{SUV} = \frac{\text{Voxel Activity Concentration}_{(\text{Bq/ml})}}{\text{Injected Dose}_{(\text{Bq})} / \text{Body Weight}_{(\text{kg})}} \quad (1)$$

To turn the voxel values into SUV metrics, this conversion was done the same way on all MAC and NAC images.

To achieve uniformity across all images, the voxel intensities were normalised by dividing by a constant factor. MAC images underwent a factor of 5 scaling, while NAC images underwent a factor of 2. By applying a uniform scale adjustment across the dataset, this method of normalisation by a constant factor makes the process easier while keeping the relative differences in radiotracer uptake between different areas within and across the images.

This method ensures that the data remains quantitatively comparable while being computationally straightforward. By scaling the intensity values in this manner, we were able to preserve the quantitative nature of PET imaging, which is vital for accurate diagnosis and assessment of metabolic activity. The histogram of the images post-normalisation illustrates the effect of this scaling on the distribution of voxel intensities, confirming the consistency of intensity levels across the processed images.

## Data Transformation and Augmentations:

For training data preparation, each PET image was initially trimmed to fit the body's outline, followed by the addition of zero-padding to standardise the dimensions to a uniform bounding box size of 168×168×Z (with 'Z' representing the count of slices), as illustrated in Figure 2a, ensuring the retention of original image resolution and anatomical structure.

This ensured the preservation of the original resolution and the fidelity of the anatomical representation. To ensure uniformity and enhance the training process's efficiency, all PET images were re-scaled to a voxel size of  $4.07 \times 4.07 \times 3.0 \text{ mm}^3$ , the most common resolution across the collected data and crucial for consistent image analysis. This standardisation was crucial for achieving consistent image quality throughout the dataset. Details regarding the initial voxel spacing are provided in Figure 2b.

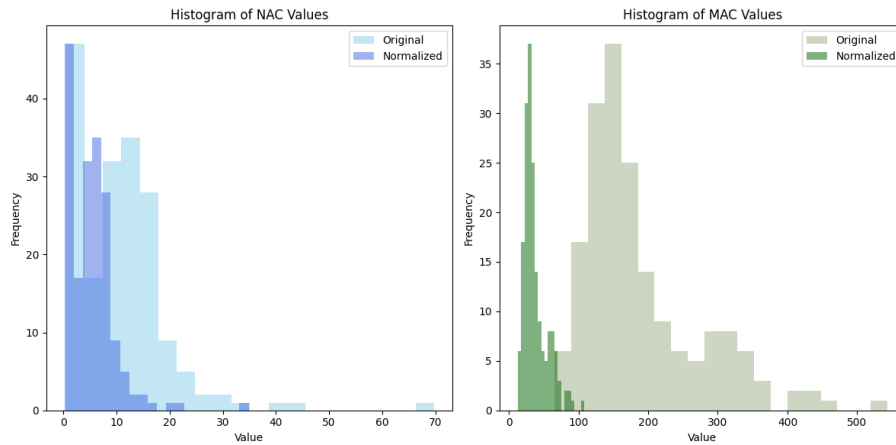


Figure 2: Distribution of maximum intensity values for NAC and MAC images, displaying variations pre- and post-normalization to highlight data scaling effects. NAC images were scaled down by a factor of 2, and MAC images by a factor of 5.

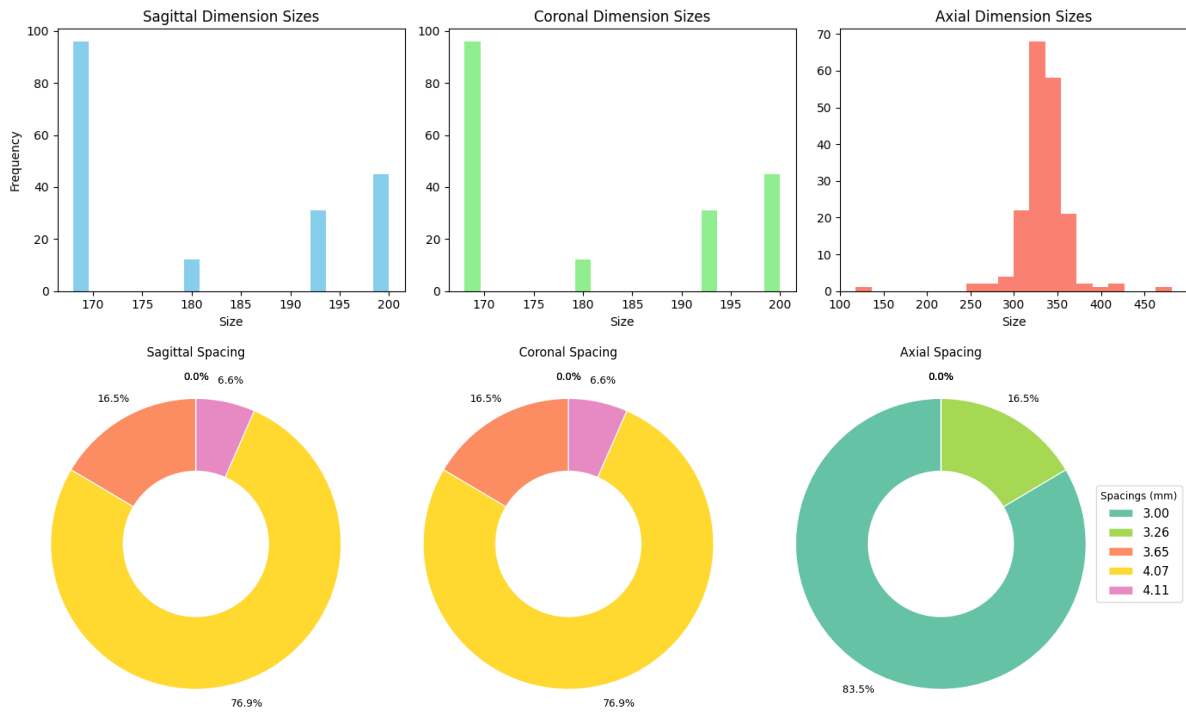


Figure 3: **A)** Distribution of initial PET image dimensions across sagittal, coronal, and axial planes. Each bar represents the frequency of occurrence for specific dimension sizes within the dataset. **B)** Proportion of different voxel spacings utilized in PET image preprocessing. The donut charts depict the percentage of images corresponding to each voxel spacing dimension in millimeters across sagittal, coronal, and axial views.

## Generation of Anatomy-Dependent Correction Maps (ADCM)

In exploring advanced techniques for PET image correction, we examine a decomposition-based deep learning approach previously outlined in the literature (52). From NAC to MAC, the complex end-to-end generation was broken down into two parts: anatomy-independent textures (related to tracers and diseases) and anatomy-dependent correction.

In other words, this method involves dividing the MAC image into these two key component maps. Anatomy-independent information, which correlates with tracer type and disease pathology, and another component, anatomy-dependent factors necessary for image correction.

The conditional equation that captures the ratio of the MAC intensity to the NAC intensity defines the anatomy-dependent correction map (ADCM) at each voxel:

$$\begin{aligned} & \text{If } PET_{NAC}[x, y, z] \geq \varepsilon \text{ then} \\ & PET_{ADCM}[x, y, z] = PET_{MAC}[x, y, z] / PET_{NAC}[x, y, z] \\ & \text{else } PET_{ADCM}[x, y, z] = PET_{MAC}[x, y, z] \end{aligned} \quad (2)$$

The threshold  $\varepsilon$  ensures that division by zero is avoided, defaulting to the MAC intensity where necessary.

In the evaluation phase, our trained model predicts the DL-ADCM for a given NAC. We then employ the following transformation (equation 2) to achieve the DL model-based attenuation correction (DL):

$$\begin{aligned} & \text{If } PET_{NAC}[x, y, z] > \varepsilon \text{ then} \\ & PET_{DL}[x, y, z] = PET_{NAC}[x, y, z] * PET_{DL-ADCM}[x, y, z] \\ & \text{else } PET_{DL}[x, y, z] = PET_{NAC}[x, y, z] \end{aligned} \quad (3)$$

Sample cases are visualised in Figure 3.

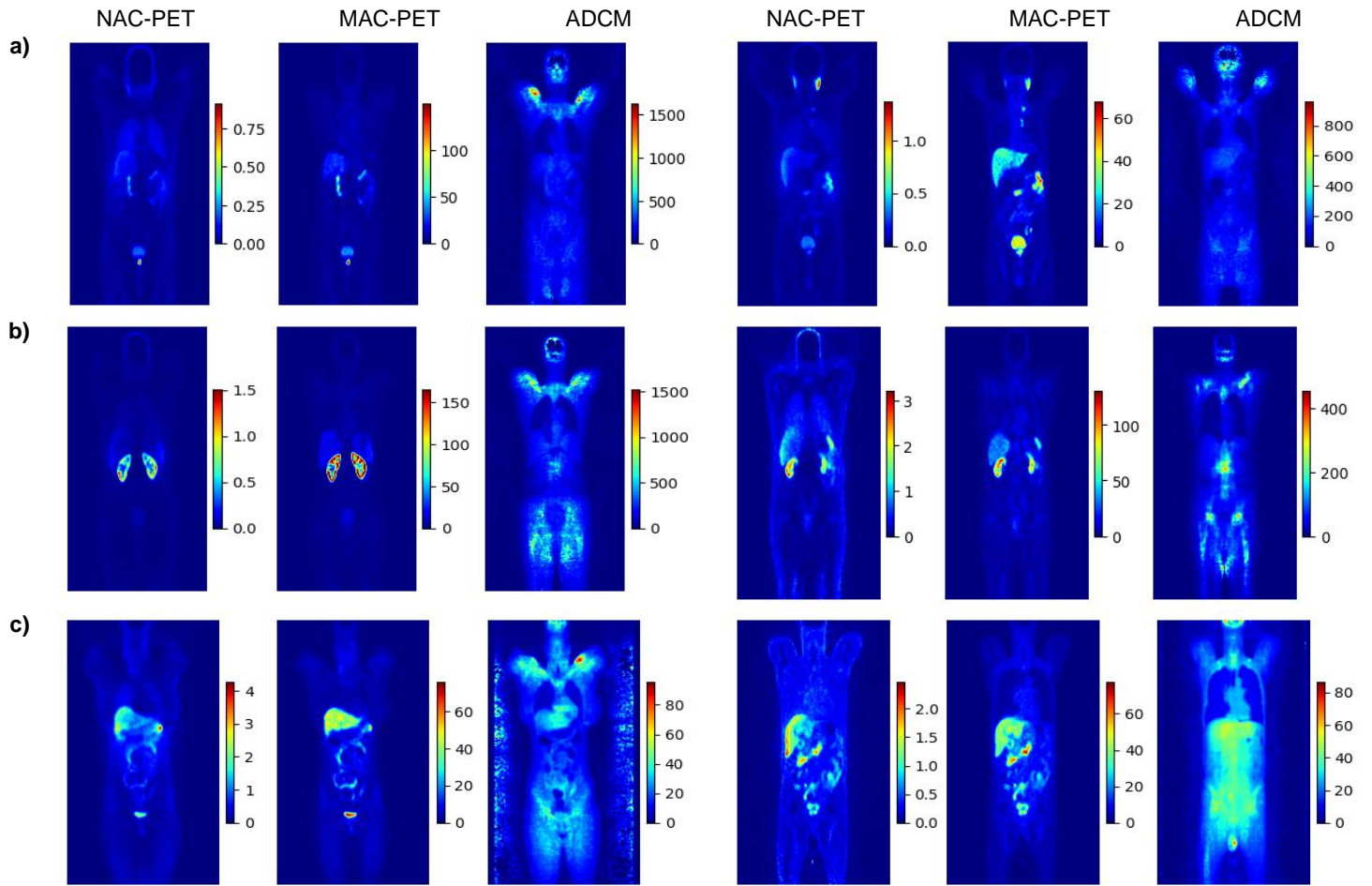


Figure 4: The middle slice of the coronal view for NAC, MAC, and ADCM images. Color bar unit: SUV

## Normalization

As we already mentioned, famous normalisation methods were not used to calibrate ADCM to protect the quantitative accuracy of SUV metrics, which is necessary for accurate clinical interpretations. We came up with an empirical normalisation factor just for ADCM values. This factor was carefully chosen to bring the dataset's wide range of values into a more manageable range that is good for deep learning applications. This factor ensures the broad spectrum of data, ranging from minimal to several thousand units, is normalised in a way that permits later recalibration into their original SUV metrics. Notably, extreme values that could bias the model (such as outliers with values of 28180 and 7300) were carefully excluded to align the focus with the representative range critical for analysis. Then, voxel intensities were normalised using a factor of 50 to maintain relative, comparable, and manageable values for training. The resultant histograms, illustrating the distribution of maximum values both pre- and post-normalisation, are depicted in Figure 5.

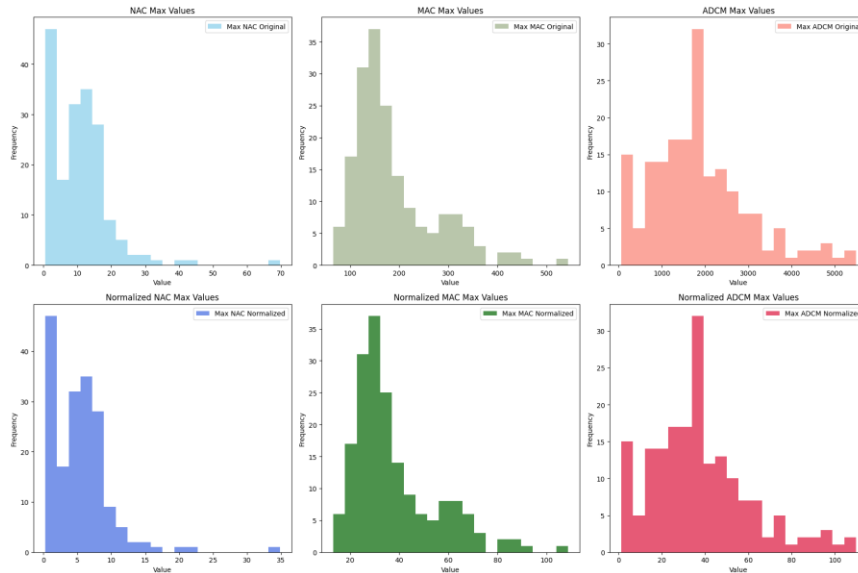


Figure 5: shows the range of highest intensity values for NAC and MAC images as well as ADCM metrics, showing changes before and after normalization to show how data scaling affects the images. Scaling down NAC images by a factor of 2, MAC images by a factor of 5, and ADCM by a factor of 50.

## FDG Datasets

To assess the model's performance with various radiotracers, our study incorporated a dataset of 98 whole-body  $^{18}\text{F}$ -FDG PET scans originating from two distinct hospitals, representing our external radiotracer dataset (Figure 6). During the preprocessing phase, the intensities of voxels in both MAC and NAC images were standardised to SUVs. This made the dynamic range of intensities more uniform so that network training would work better. Empirical scaling factors, 9 for MAC and 3 for NAC images, were applied to further constrain the dynamic range.

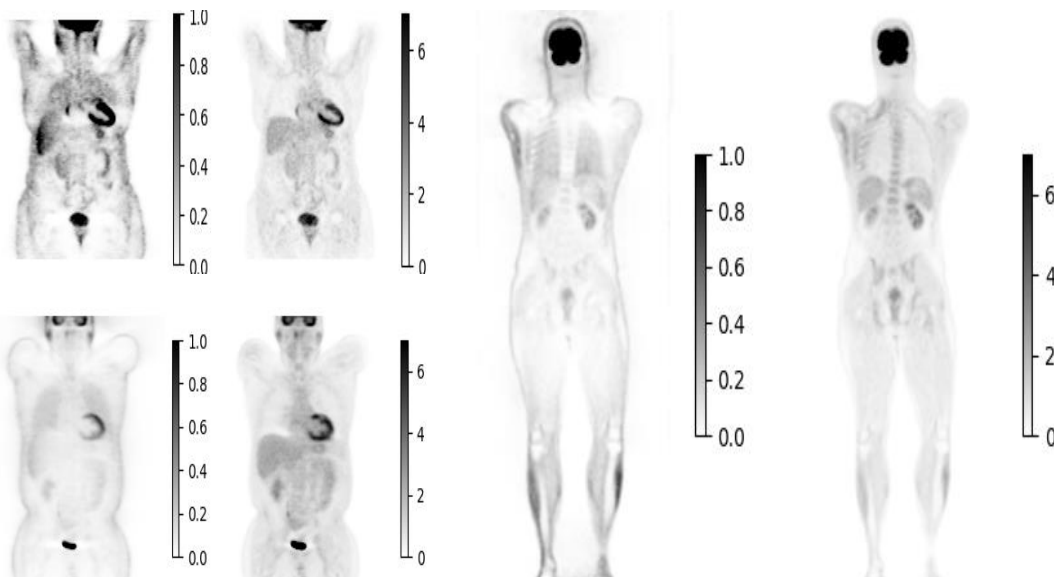


Figure 6: Sample of coronal slices from an FDG dataset, illustrating the range in axial slice counts, which vary from 180 to 600 based on the organ of interest.

To achieve homogeneity across the dataset, we standardised the voxel spacing to 1.92 mm for both coronal and sagittal planes, with an axial dimension set to 3.0 mm, which aligns with 52% of the existing data. These standardised spacings ensure uniformity across all scans in preparation for model training.

Table 2: "Overview of External Radiotracer Dataset Specifications.

Center	No	Train	Validation	Test	Matrix size × Z
Center 6	55	39	6	11	272 × 200
Center 7	43	23	9	10	272 × 200
Total	98	62	15	21	-
* Z' representing the number of slices in the axial view, depends on body length, scanner resolution, scan protocol, and patient positioning. So, it is different patiently.					

## Artifact dataset

In this study, a third test set was utilised to rigorously evaluate the performance of the developed model under more challenging conditions. This set consisted of imaging data from 198 patients, each displaying various types of artefacts. The artefacts in this dataset were chosen to test how well the model can handle and correctly interpret images that are distorted by common problems seen in clinical Ga imaging, like motion and Halo artefacts.

## Deep neural network

For final implementation, we leverage the Dyn-UNet architecture, renowned for its adaptability and efficiency in processing biomedical images (Isensee et al., 2019). This model is particularly chosen for its dynamic configuration and deep supervision, enabling precise results tailored to the specific requirements of our dataset.

The Dyn-UNet model's initialization is specially made to find the best kernel sizes and strides based on the size and spacing of the input patches in our dataset. By evaluating the spatial dimensions and resolution of the input data, these parameters were determined, ensuring the network architecture is directly aligned with the inherent characteristics of our medical images.

The Dyn-UNet model is specified with supervision heads, which ensure that intermediate layers are optimised for accurate prediction, enhancing learning efficiency and model robustness. Deep supervision ensures that intermediate layers are also optimised for accurate prediction, not just the final output layer. This strategy boosts the learning efficiency and enhances the robustness of the model, making it adept at segmenting complex anatomical structures with high fidelity.

For the Gallium dataset, the computed kernel sizes and strides are set to four layers of [3, 3, 3] kernels, with strides transitioning from [1, 1, 1] in the initial layer to [2, 2, 1] in the deeper layers. Thereby ensuring a balanced focus on capturing both high-resolution details and broader anatomical structures. Additionally, the implementation of deep supervision, with two supervision heads, was a critical decision aimed at enhancing the learning process by optimising both the final and intermediate

layers of the network. This comprehensive approach to selecting hyperparameters underscores our commitment to leveraging the Dyn-UNet's full potential for achieving precise and robust segmentation results, tailored specifically to the complexity and variety of our imaging data. By adjusting the ReLU activation function in the last layer, we can get the non-zero value from the concept of the PET image.

Our deep learning network was designed to process NAC images as inputs with the objective of generating MAC images. For certain scenarios, which will be elaborated upon later, the output included the anatomy-dependent correction maps (ADCM), derived from the MAC images.

Network training involved using 3D patches sized at 168x168x16 and 20 sample patches per patient. The key training parameters were as follows: a learning rate of 0.001, and the optimisation of the mean squared error (MSE) loss function—also referred to as the squared L2 norm. The MSE loss function was employed to measure the deviation of the network's output from the MAC ground truth.

The optimisation of the network was conducted using the Adam algorithm, with the aim of minimising the loss function effectively. The beta coefficients, set at 0.5 and 0.999, governed the moment estimates' exponential decay rates. The architecture of our network is detailed in Supplemental Material 1.

To enhance the robustness of our model, we implemented specific data augmentations. These included adding rotations of  $\pm 15$  degrees and increasing the number of samples per patient from 4 to 20.

To maintain the integrity of the model, only artefact-free datasets were used during the network's training and validation stages. We trained the network over 400 epochs to ensure adequate convergence and comprehensive learning from the dataset. An epoch represents a complete iteration over the entire training dataset. The validation set, separate from the training set, was used solely to assess the network's performance, and was not involved in fine-tuning any hyperparameters. To prevent data leakage and ensure data integrity, there was no overlap of patients across the training, testing, and validation datasets, maintaining the independence of each dataset.

Details on alternative models tested, including those that did not meet our criteria for inclusion in the final report, are documented in Supplementary Material 1 for transparency and completeness.

## Training approaches for deep learning models:

### Integrated multi-Centre model (IMCM):

A Dyn-Unet deep learning model was developed using a combined dataset from four different centres, all utilising gallium-based tracers. This model was initially trained on a collective dataset and subsequently tested on an external centre's data to evaluate its generalisation capabilities. It was also tested within the originating dataset from each centre. This approach aims to overcome the limitations of models trained on data from single centres, which may struggle with generalizability to new, unseen cases. The training and validation losses for the IMCM are illustrated in Figure 7.



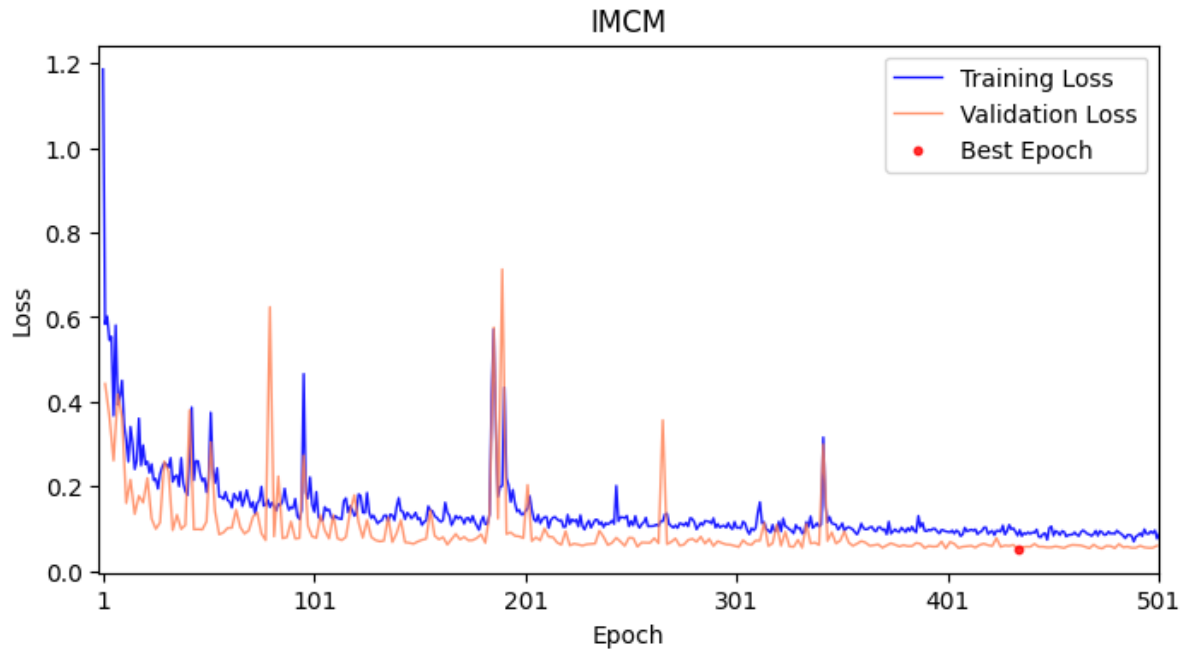


Figure 7: Training and validation loss for the Integrated Multi-Center Model showing a best metric of 0.0527 at epoch 434.

### Anatomy-Dependent Correction Model (ADCM):

This methodology adopts a new approach by decomposing the transformation from non-attenuation corrected PET (NAC-PET) to model-based attenuation corrected PET (MAC-PET) into two distinct components. Specifically, the model targets anatomy-independent features associated with tracers and diseases and anatomy-dependent corrections that are crucial for accurate image interpretation. This decomposition enables a more targeted and efficient handling of the data during the deep learning process.

The previous network was employed to focus exclusively on estimating the anatomy-dependent correction maps (ADCM). This model's effectiveness is evaluated through its ability to generalise across different centres and tracers, testing its robustness in a variety of clinical settings. The training progress and validation stability for the ADCM are detailed in Figure 8.

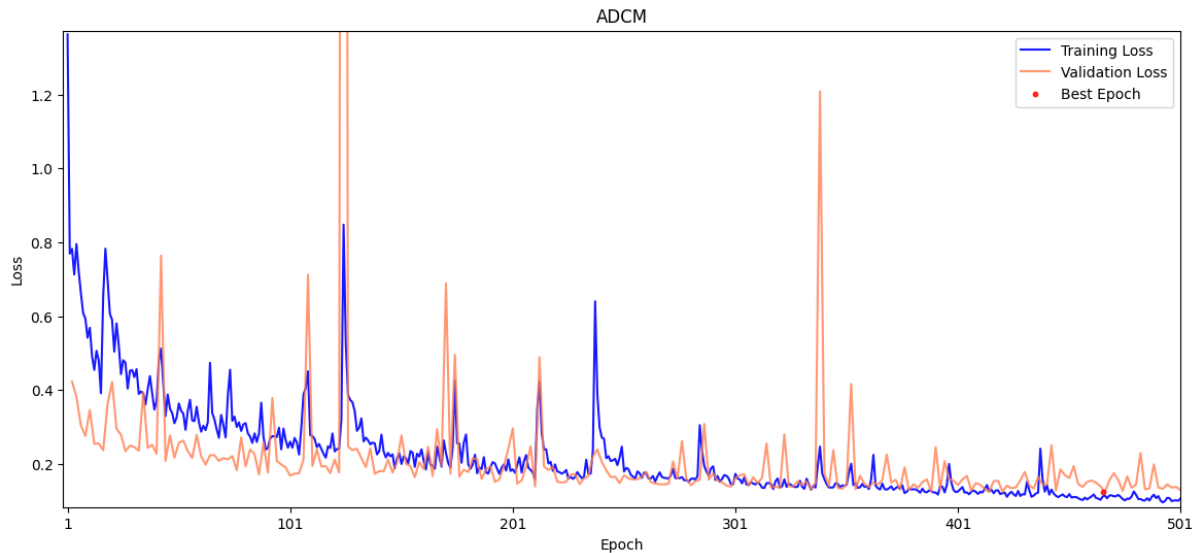


Figure 8: Training and validation loss for the ADCM model, where the best metric of 0.1237 was reached at epoch 466.

### Tuned Transfer Learning for IMCM model (TL-MC):

To address the challenges encountered with different radiotracers, the IMCM model underwent tuning through transfer learning (TL). This method involves modifying the deep learning model by integrating learning from decentralised data sources without requiring direct data sharing. This refinement was aimed at enhancing the model's performance and adaptability across different tracer types, providing a more robust solution that could potentially handle variability more effectively. The effectiveness of the TL approach is depicted in Figure 9, demonstrating rapid convergence and effective transfer learning.

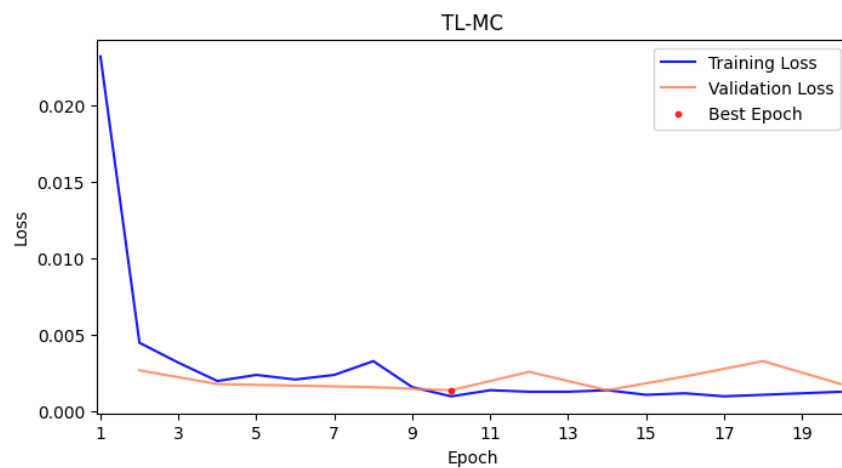


Figure 9: Training and validation loss for the Tune TL Model with a best metric of 0.0014 achieved at epoch 10, demonstrating rapid convergence and effective transfer learning.

## Quantitative evaluation:

The model's efficacy was rigorously quantified using a range of statistical metrics, calculated by comparing the DL-predicted PET images against the ground truth CT-based attenuation/scatter corrected images. These voxel-wise metrics computed as follows:

- **Mean Error (ME):** This reflects the average deviation across all voxels.

$$ME = \frac{1}{tot} \sum_{v=1}^{tot} PET_{pred}(v) - PET_{ref}(v) \quad (4)$$

- **Mean Absolute Error (MAE):** Measures the average magnitude of errors without considering their direction.

$$MAE = \frac{1}{tot} \sum_{v=1}^{tot} |PET_{pred}(v) - PET_{ref}(v)| \quad (5)$$

- **Relative Error (RE%):** Provides a percentage error relative to the true values, indicating the proportion of the deviation.

$$RE (\%) = \frac{1}{tot} \sum_{v=1}^{tot} \frac{(PET_{pred})_v - (PET_{ref})_v}{(PET_{ref})_v} \times 100\% \quad (6)$$

- **Root Mean Squared Error (RMSE):** Measures the average of the squared differences between the predicted and reference values. It is useful for quantifying the deviation in predictions from the observed values across the dataset.

$$RMSE = \sqrt{\frac{1}{tot} \sum_{v=1}^{tot} ((PET_{pred})_v - (PET_{ref})_v)^2} \quad (7)$$

Where tot refers to the total number of voxels, and  $PET_{pred}$  and  $PET_{ref}$  indicate the predicted image via DL model and the ground truth image, respectively.

- **Peak Signal-to-Noise Ratio (PSNR):** Evaluates the ratio of the maximum possible signal to the corrupting noise.

$$\text{PSNR(dB)} = 10\log_{10}\left(\frac{\text{Peak}^2}{\text{MSE}}\right) \quad (8)$$

In Eq. 8, Peak represents the maximum intensity value in the image.

- **Structural Similarity Index (SSIM):** Assesses the perceptual quality of the predicted images relative to the reference images.

$$\text{SSIM}(PET_{pred}, PET_{ref}) = \frac{(2\mu_{pred}\mu_{ref} + c_1)(2\sigma_{pred,ref} + c_2)}{(\mu_{pred}^2 + \mu_{ref}^2 + c_1)(\sigma_{pred}^2 + \sigma_{ref}^2 + c_2)} \quad (9)$$

where:  $\mu_{pred}$  and  $\mu_{ref}$  are the averages of the pixel intensities in the predicted PET images ( $PET_{pred}$ ) and the CT-attenuation corrected PET images ( $PET_{ref}$ ), respectively.  $\sigma_{pred}^2$  and  $\sigma_{ref}^2$  are the variances of the pixel intensities in the predicted and CT-attenuation corrected PET images, respectively.  $\sigma_{pred,ref}$  is the covariance of the predicted and CT-attenuation corrected PET images.

$c_1 = (k_1L)^2$  and  $c_2 = (k_2L)^2$  are constants to stabilize the division with a weak denominator; L is the dynamic range of the pixel values (typically  $2^{\text{bit per pixel}} - 1$ ).  $k_1=0.01$  and  $k_2=0.03$  are default values for the stabilization constants.

# Results

## Quantitative assessment

### Cross-Centre Results:

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

Figure 6 displays the quantitative accuracy of the deep learning-based images compared to the ground-truth, MAC images for both internal and external centres. The results demonstrate that both DL methods effectively performed some degree of attenuation and scattering correction across these centres. For a detailed centre-wise analysis, refer to the Supplementary Material in Figure 1.

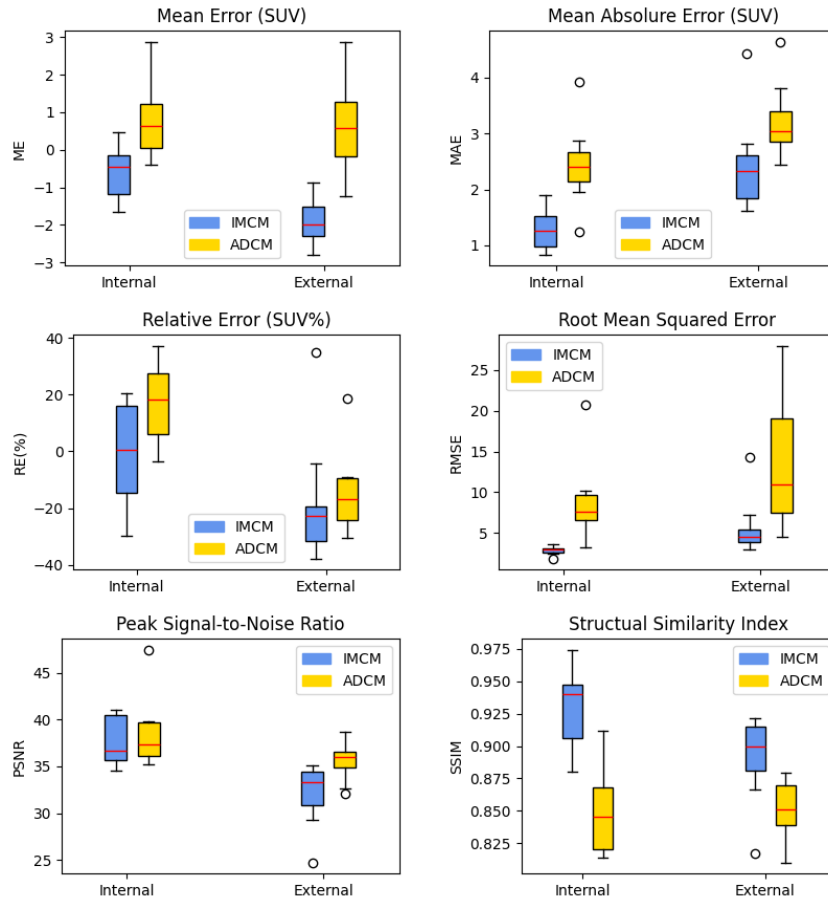


Figure 10: Quantitative metrics for the IMCM and ADCM methods across internal and external centres, including mean error (SUV), mean absolute error (SUV), relative error (SUV%), root mean squared error, peak signal-to-noise ratio, and structural similarity index.

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

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

PSNR also favoured the IMCM method, registering at  $35.526 \pm 2.117$  (CI 95%: 34.9 to 36.2) compared to  $38.251 \pm 1.923$  (CI 95%: 37.6 to 38.9) for the ADCM method. Notably, SSIM for IMCM at the external centre was superior, recorded at  $0.879 \pm 0.020$  (CI 95%: 0.871 to 0.887). Details are available in the Supplementary Material, table 1.

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 was used here due to the non-normal distribution of the data, as evidenced by the Shapiro-Wilk tests.

The Wilcoxon test with the False Discovery Rate (FDR) method correction showed that the ADCM and IMCM datasets were significantly different for all metrics except for relative error. In other words, corrected p-values using the Benjamini-Hochberg procedure indicated notable discrepancies in error measurements and image quality between the methods (66). except for the Relative Error (SUV%), where the corrected p-value does not indicate a statistically significant difference threshold of 0.05. IMCM shows consistently lower errors, a higher PSNR, and higher SSIM values, indicating superior image quality and more reliable estimations. These findings are further detailed in Supplementary Material, Statistical test.

In the analysis of the joint histograms, such as Pearson correlation, the voxel-wise correlation across the different centres for both methods were visualised in Figure 11. A clear difference in predictive accuracy and linearity in SUV estimation was demonstrated. In the external centre, the IMCM regression slope of  $0.65 \pm 0.02$  with an R-value of 0.949 clearly showed a systematic underestimation over the range of predicted SUV values, compared to ADCM, which showed a slope of  $1.18 \pm 0.10$  and an R-value of 0.850, suggesting a trend towards overestimation potentially linked to very high SUV values that might not be clinically advantageous.

In internal centres, the behaviour of the methods differed, with the IMCM method closer to ideal prediction, especially evident at centre C3 with a regression slope of  $0.87 \pm 0.01$  and an R-value of 0.988. On the other hand, the ADCM method had slopes greater than one, specifically  $1.13 \pm 0.03$  at C2 and  $1.19 \pm 0.03$  at C4. This could mean that the method wasn't calibrated correctly, leading to an overall overestimation of the SUV.

Voxel-wise analysis further confirmed these findings, showing larger discrepancies in centres where ADCM predicted significantly higher values. Overall, these results demonstrate that ADCM appears to be closer to the truth in some centres because the R-values are higher. However, the reliability and clinical usefulness of ADCM can be called into question. IMCM demonstrated image quality comparable to MAC and preserved more detailed information with lower noise compared to ADCM.

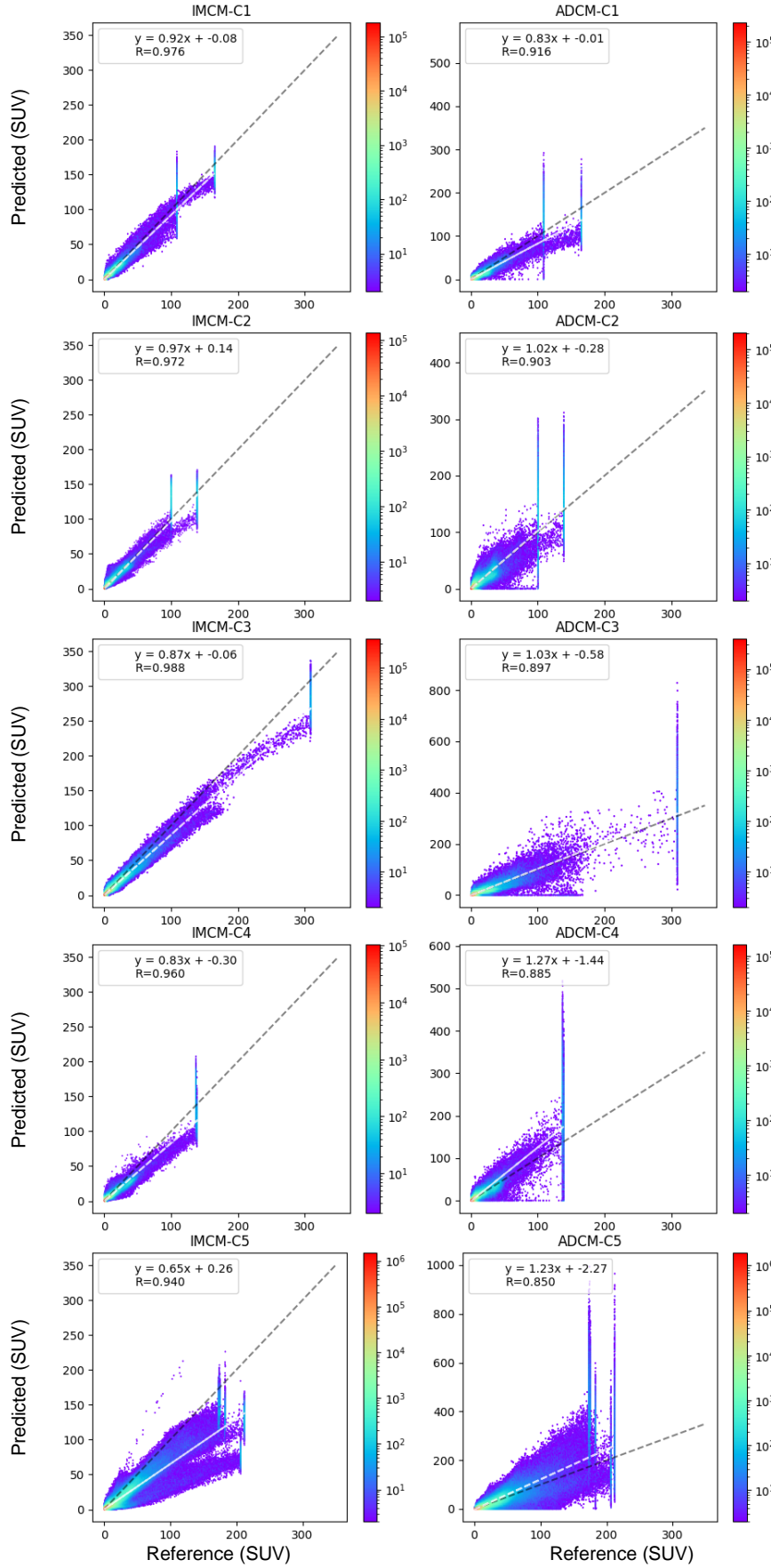


Figure 11: Joint histogram analysis displaying the correlation between activity concentration in DL-IMCM and DL-ADCM images versus reference MAC images serving as the ground truth. Note that a logarithmic scale was used to display the SUV levels. C1-4 are internal centres, while C5 is an external centre.



## Cross-Tracer Results:

As part of our assessment of generalisation capabilities across different tracer types, IMCM was initially tested without specific tuning for cross-tracer variations. As proved before, the results revealed that the IMCM, without prior tuning, struggled to maintain its efficacy when applied to different radiopharmaceutical tracers (36). But this outcome contrasts sharply with the claims from the ADCM approach, which posits that the ADCM model architecture inherently accommodates variations across tracers and anatomical structures without the need for additional adjustments.

So, the  $^{18}\text{F}$ -FDG-PET dataset was used as a cross-tracer in this study to test the two proposed DL algorithms: TL-MC (the tuned version of IMCM) and ADCM. We tested the trained DL model to evaluate its robustness, which included 20 subjects from 2 different centres as external non-seen centres.

Figure 12 showcases a sample coronal slice 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 generalisation with a single, unified model approach. These results show how important it is to tune the model specifically to each tracer's specific properties. This will make the model more useful and accurate in a wider range of clinical settings.

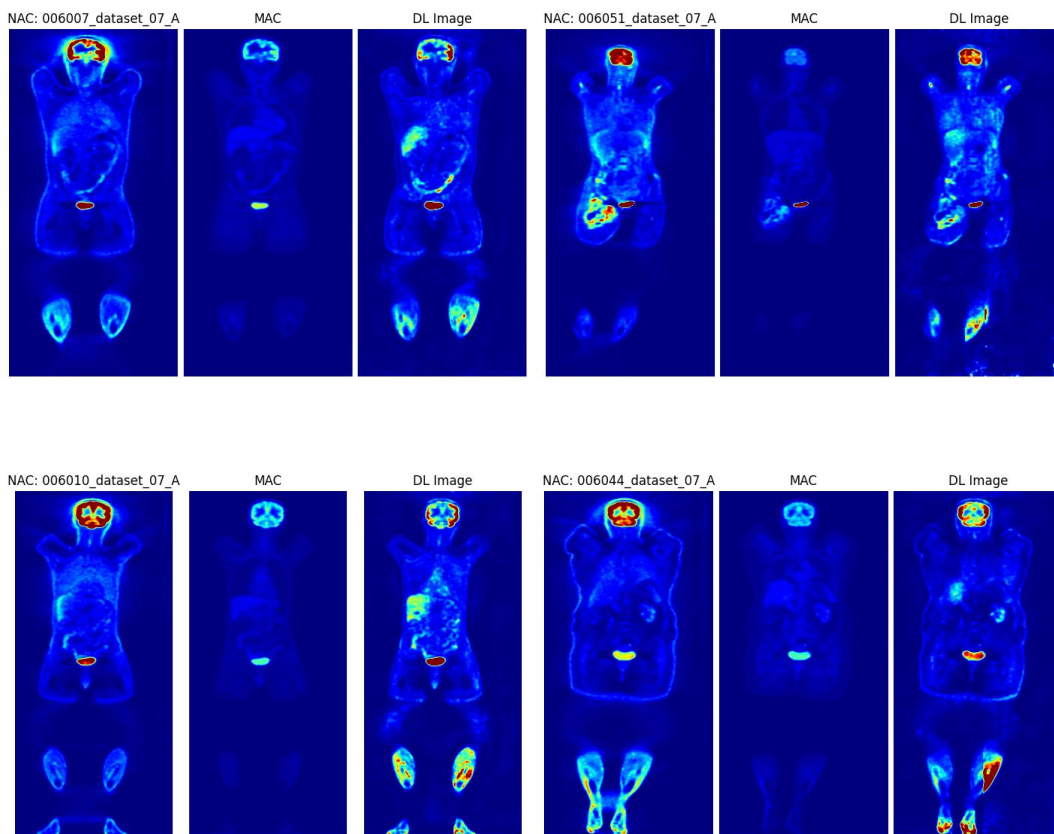


Figure 12: From left to right, a coronal slice of NAC, MAC, IMCM, TL-MC, and ADCM on cross-tracer subjects, respectively.

The two approaches, TL-MC and ADCM, indicate significant differences in error metrics. 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  $-0.10 \pm 0.76$ , while the ADCM deviated by  $0.82 \pm 0.70$ , signifying a much wider spread of the SUV estimates (Figure 13).

These are shown as RE%. This also confirms that TL-MC had a better performance. The RE spread was relatively lower for TL-MC, averaging at  $30 \pm 50\%$ , in contrast with ADCM, where the spread was much broader at  $50 \pm 100\%$ .

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

All together, these findings point towards 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. The data make a compelling case that TL-MC should be preferred with respect to its strong performance in consistently keeping lower errors in the images. For a comprehensive view and deeper analysis, refer to the box plots in Figure 9. Detailed statistical comparisons of these metrics are illustrated in Supplementary Material 2, table 3 & 4, provided.

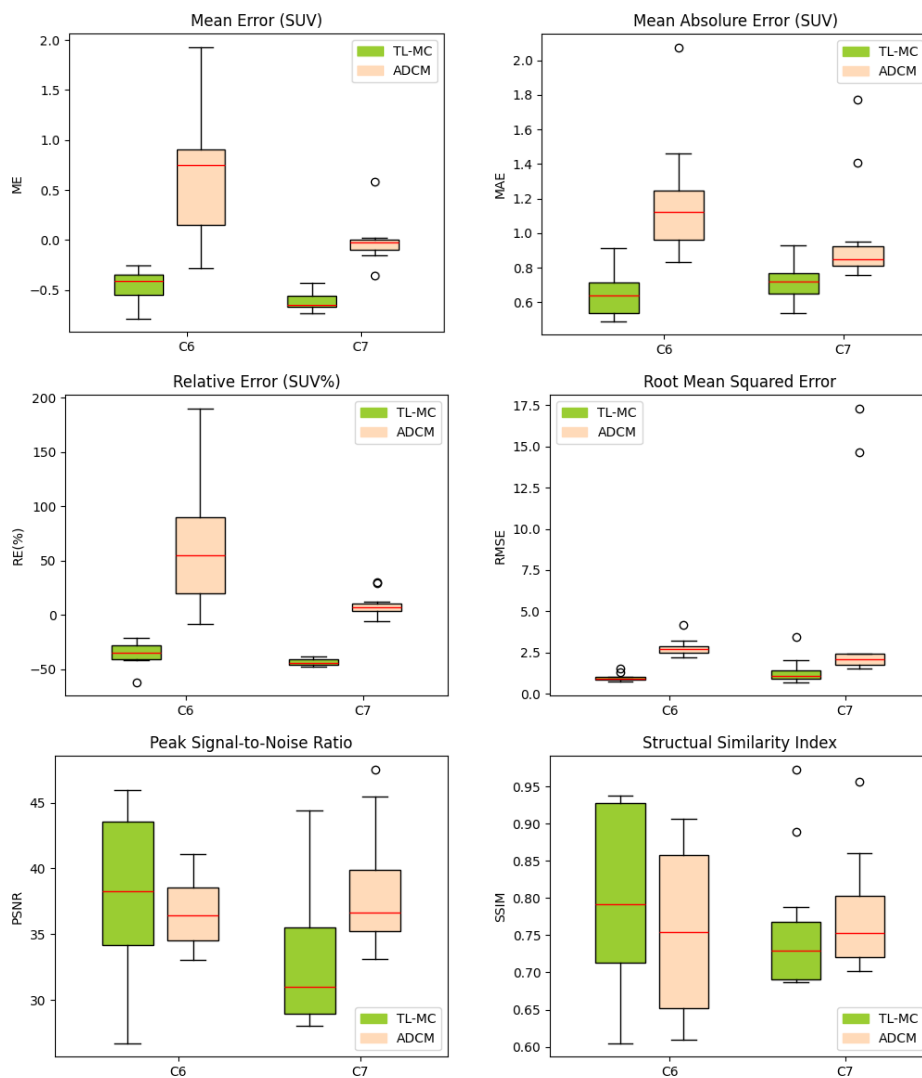


Figure 13: Comparative Analysis of Imaging Metrics Between ADCM and IMCM Methods. The box plots depict the distribution of mean error (SUV), mean absolute error (SUV), relative error (SUV%), root mean squared error, peak signal-to-noise ratio, and structural similarity index across centres C6 and C7.

Upon further investigation through joint histogram analysis of the TL-MC and ADCM models across different centres, a nuanced understanding of each model's predictive capabilities for standardised uptake values (SUVs) emerges. The TL-MC model aligns closely with reference values, as evidenced by regression slopes of  $0.98 \pm 0.38$  and  $0.69 \pm 0.08$  at two respective centres. Notably, this model also shows high correlation coefficients of 0.915 and 0.918, underscoring its precision in SUV prediction despite a tendency to slightly underestimate values, particularly at Centre C7 as depicted in the analysis.

On the other hand, the ADCM model has lower correlation coefficients of 0.660 and 0.678, even though its regression slopes are higher at  $1.10 \pm 0.46$  and  $1.35 \pm 0.66$ , which means it overestimates the data. This discrepancy highlights the lesser consistency and reliability of its predictions when compared to TL-MC. Contrary to intuitive expectations of better correlation, the higher slopes observed in ADCM indicate a greater deviation from the reference line, pointing to a systematic error in overestimating SUVs.

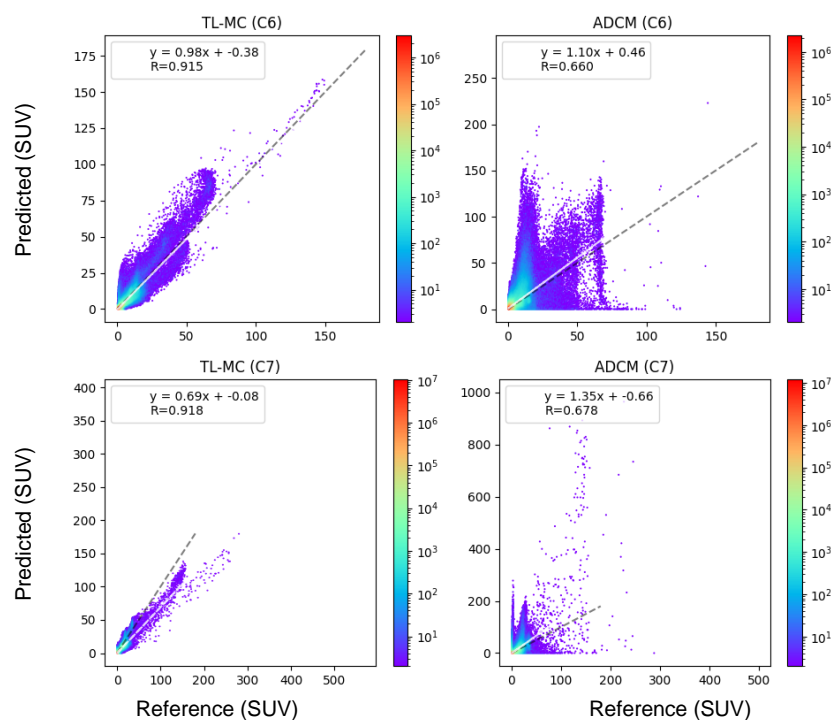


Figure 14: Joint histogram analysis displaying the correlation between activity concentration in TL-MC and ADCM images versus reference MAC images serving as the ground truth for cross-tracer. Note that a logarithmic scale was used to display the SUV levels.

## Case Study on artefact images

In this section, a series of case studies involving repeated scans were examined. These repeated scans have been requested by nuclear medicine physicians shortly after initial assessments.

Figure 15 and 16 and 17 displays the imaging results for patients with halo artefacts in the pelvic, kidney, diaphragm, lung, liver and spleen regions. These artefacts were removed in the repeated scan.

The ICMC method produced artefact-free images of high quality, diagnostic confidence, and nearly identical to the initial scan.

Figure 18 features patients with a halo artefact in the kidneys. A repeated scan was conducted in this region due to the initial scan's low image quality and diagnostic confidence. Unfortunately, for some cases the repeated scan could not remove these artifacts. Nonetheless, the ICMC model successfully eliminated the artefact in both the original and subsequent scans.

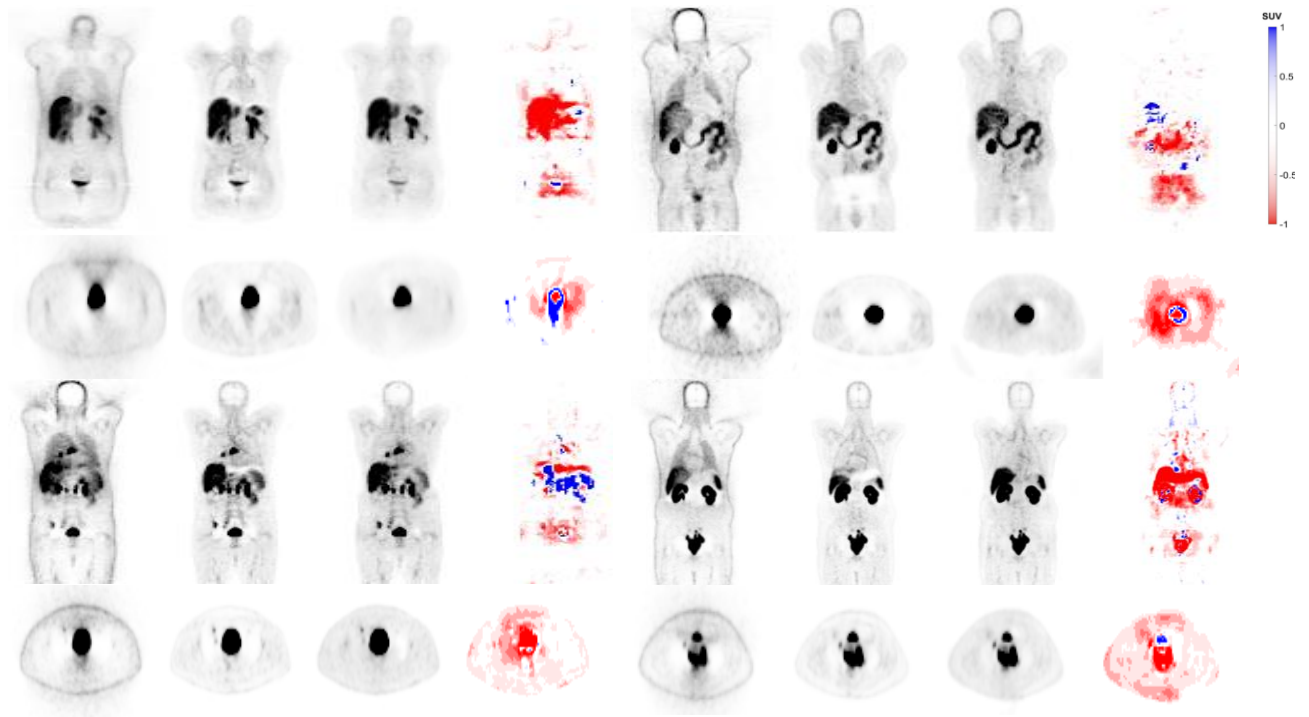


Figure 15: Coronal and axial views of 12 clinical studies showing from left to right NAC, MAC, IMCM-DL and the difference images of MAC and DL image. The images generated using the IMCM approach successfully corrected the halo artefact in pelvic area.

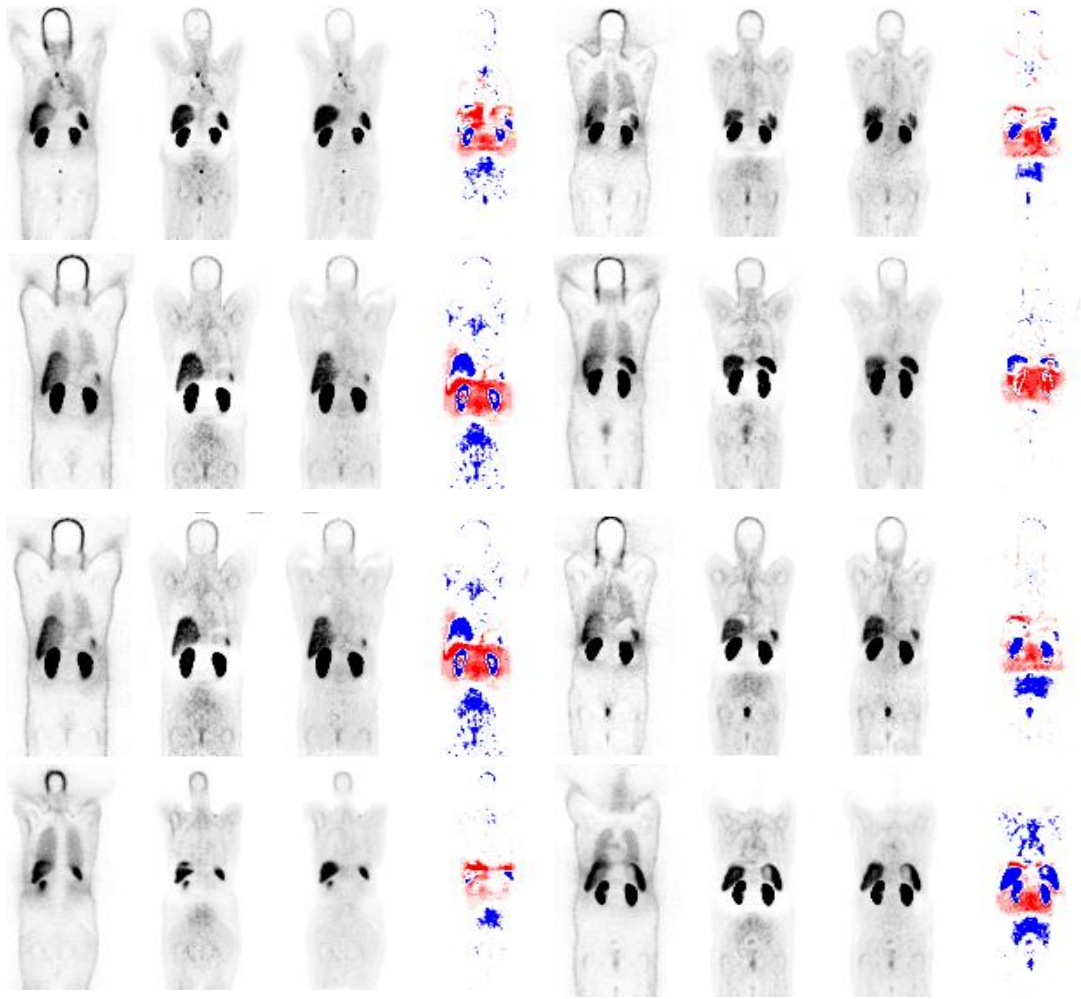


Figure 16: Coronal views of 8 clinical studies, representing from left to right: NAC, MAC, IMCM-DL and the difference images of MAC and DL image. Our method effectively disentangles halo artefacts in the kidney area.

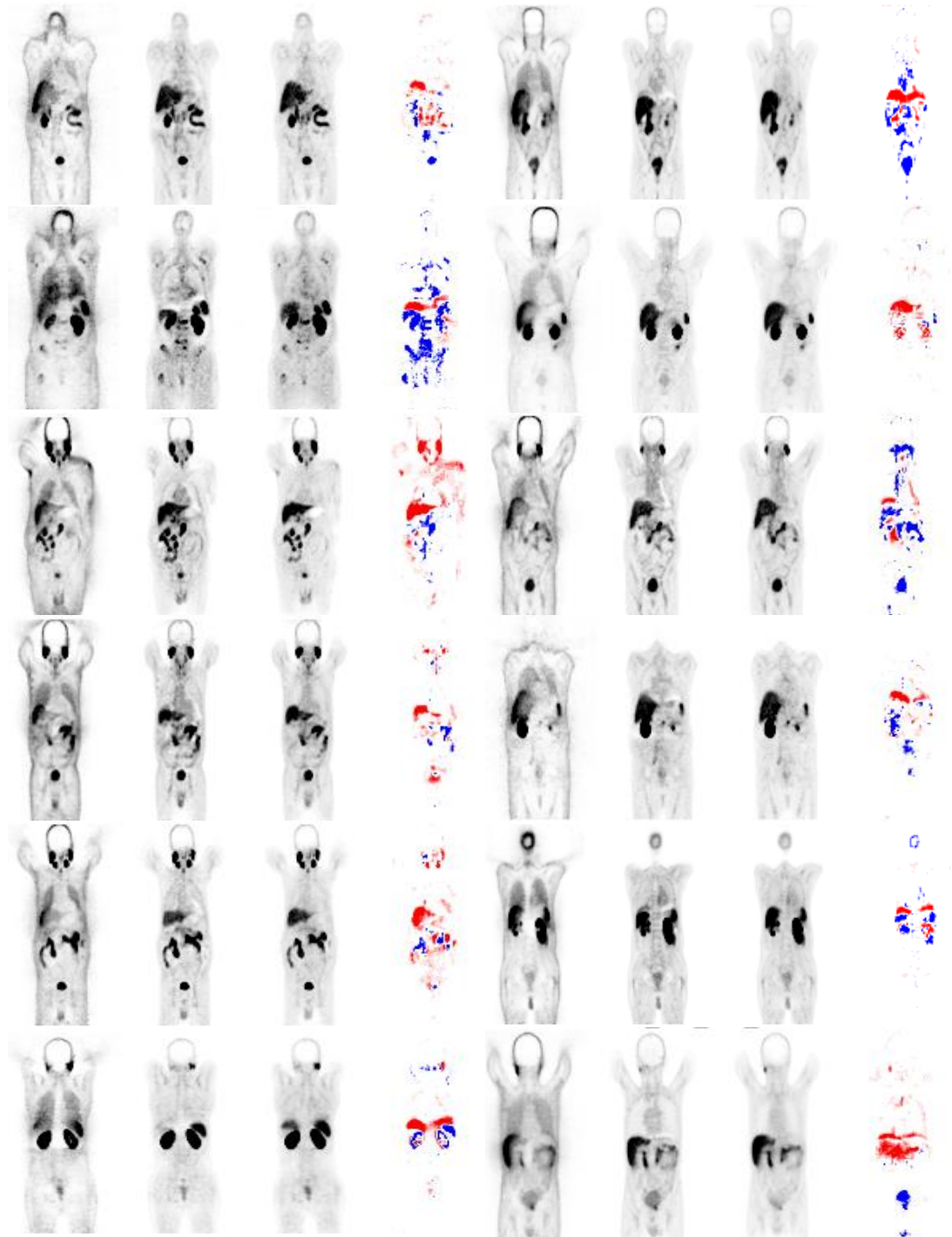


Figure 17: Coronal views of 12 clinical studies showing from left to right NAC, MAC, IMCM-DL and the difference images of MAC and DL image. The images generated using the IMCM approach successfully corrected the mismatch artefact in the diaphragm, lung, liver and spleen regions.

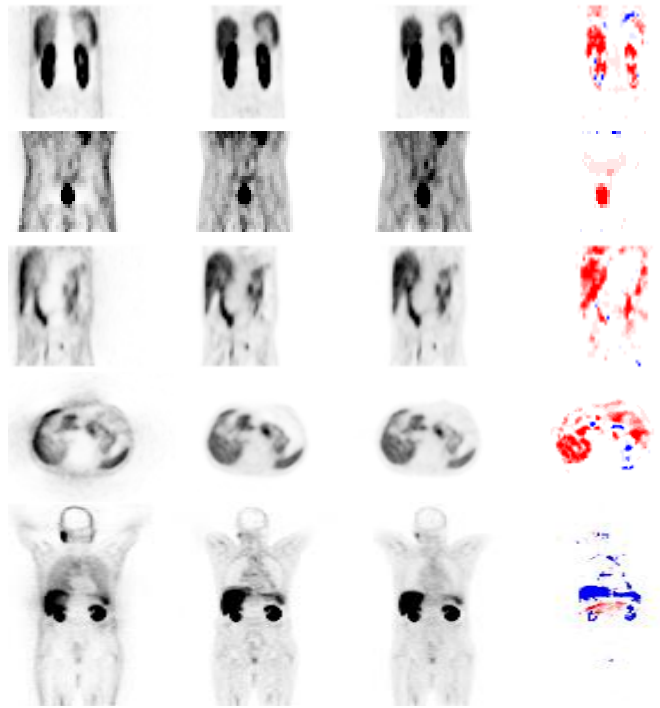


Figure 18: Coronal and axial views showing from left to right NAC, MAC, IMCM-DL and the difference images of MAC and DL image. The repeated scan which was requested right after the initial scan. The IMCM image recovered high quality and high diagnostic confidence for both scans.



# Discussion

Various deep learning-based attenuation scatter correction (DL-ASC) methods have been developed for PET imaging. These include indirect approaches that generate attenuation maps from MRI and NAC images, or maximum likelihood estimation of activity and attenuation (MLAA). For instance, studies have employed generative adversarial networks (GANs) to derive pseudo-CT images from PET non-attenuation-corrected (non-AC) scans in both brain and whole-body PET imaging. Furthermore, MLAA methodologies have been enhanced by incorporating deep learning to mitigate common issues such as crosstalk artefacts, slow convergence, and the generation of noisy attenuation maps. Direct DL-ASC methods get around traditional methods by making ASC PET images directly from NAC images. This was first used in brain PET imaging, then tested in 18F-FDG PET studies of the whole body.

A significant challenge arises with the low tracer activity and the extensive positron range of  $^{68}\text{Ga}$ -labelled pharmaceuticals, which generally produce lower-quality images compared to  $^{18}\text{F}$ -labelled compounds. Initially, employing DL for direct ASC in PET might seem overly reliant on advanced technology. However, our findings indicate that not only does it enhance both quantitative and qualitative aspects of PET images, but it also effectively identifies and corrects mismatches and halo artefacts without needing anatomical images. While indirect techniques require reconstructions to produce ASC PET images, they often fail to address halo artefacts that arise during the reconstruction phase and are predominantly influenced by the PET images themselves.

This study has demonstrated that a single universal model may not be effective due to variations in tracer-injected activity across different hospitals. There is a need to tune radiotracer-wise models using heterogeneous datasets to address these discrepancies. But the good news is that using large and heterogeneous datasets from different hospitals in the same tracer can compensate for the differences in equipment, image acquisition, and reconstruction strategies. In our research, we utilised differential data from various hospitals, which enhanced the accuracy of attenuation scatter correction (ASC) in PET images when implementing a shared model across different hospitals for identical radiotracer imaging.

Furthermore, we employed the integrated multi-centre model (IMCM) for additional qualitative analysis. Through quantitative assessments, we observed the substantial impact that radiotracers and scanners have on model performance. Notably, IMCM greatly increased the quantitative accuracy across various scanners, indicating the need for model tuning using transfer learning that is tailored to specific tracer situations and thus performing better than ADCM. IMCM showed enhanced efficiency when different scanners utilised the same radiotracer, compared to when various radiotracers were employed on the same scanner. We also found that the source of the data, including the type of scanner and radiotracer used, significantly affected ADCM's effectiveness, contrary to initial assumptions.

While the ADCM method focuses on decomposing the PET image correction process into anatomy-independent textures and anatomy-dependent corrections, there are noteworthy limitations to this approach. Even though the method uses deep learning to estimate anatomy-dependent corrections and focuses on anatomical details, our results suggest that the ADCM method may not be able to handle the differences that come from using different scanners and radiotracers well. This limitation could undermine its utility in scenarios where data privacy concerns or logistical constraints prevent

the sharing of sensitive information across centres or with a central server. Consequently, while ADCM's premise is theoretically sound, its practical application across diverse clinical environments and varied technological setups appears limited. This underscores the need for more robust, adaptable models like IMCM, which not only accommodate but thrive on the heterogeneity inherent in multi-centre clinical data.

Regarding the ADCM method implementation, the dataset was processed at a higher resolution of  $4 \times 4 \times 3$  mm/voxel compared to the previously referenced method, which used a resolution of  $6.6 \times 6.6 \times 8$  mm/voxel. While this increased resolution aided in capturing finer anatomical details, it did not help in our scenario. Furthermore, we observed that the ADCM occasionally failed to exhibit these textures clearly, particularly in cases with lower radioisotope uptake or in regions where anatomical variations are subtle but critical. One of the most significant differences between our work and the original paper was the normalisation method. Our focus on ADCM normalisation adopted in this study was informed by the necessity to preserve the clinical significance of SUVs. This led to the selection of an empirical normalisation constant, circumventing the use of standard min-max normalisation, which could diminish the quantitative richness essential for clinical interpretation. Other issue during this normalisation process was that extreme values and outliers were created because of the nature of dividing by small values. Another contributing factor to these differences lies in the choice of deep learning algorithms utilized. This study employs a U-Net architecture, in contrast to the use of GANs in the original research.

The joint histogram analysis raised pertinent questions regarding the calibration and reliability of the ADCM method in clinical settings. Notably, the overestimations seen by ADCM, especially in cross-centre, could lead to incorrect diagnoses in situations where accuracy in SUV estimation is critical. The systematic bias towards higher SUV values, although providing a superficial appearance of accuracy due to closer R-values to unity, suggests underlying issues in the algorithm or its application across different PET systems.

In contrast, the IMCM method, with its adherence to lower regression slopes and higher correlation coefficients, particularly in internal centres, underscores its suitability for clinical applications 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 uniform performance across different settings.

The analysis across cross-tracer highlights 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 behaviour of SUVs across the studied centres.

CT-based attenuation scatter corrections (CT-ASC) are a primary adjustment for quantitative  $^{68}\text{Ga}$  PET imaging. However, this process can introduce mismatches and halo artefacts in  $^{68}\text{Ga}$  PET images, potentially altering patient diagnosis and prognosis. These artefacts are challenging to detect and correct in real clinical settings.

Our developed model does not require iterative image reconstruction with ASC. Qualitative analysis underscored the effectiveness of our proposed model in detecting and correcting mismatches and halo artefacts in the chest, abdomen, and pelvic regions without needing ground truth in  $^{68}\text{Ga}$  PET images. We also observed scenarios in which repeated scans, typically conducted to eliminate

artefacts, failed, and even exacerbated them. Here, our DL algorithms were able to distinguish and correct these issues independently of the ground truth.

The predominant limitation of previous studies lies in their single-centre datasets, which restrict the generalizability of DL models. Our current study employs a multi-centre approach to address this issue.

One of the limitations of this study was the restricted access to advanced computational resources, including high-performance GPUs. This constraint affected the efficiency of data processing and extended the runtimes considerably, which posed challenges in model training and validation phases.

Moving forward, future research should explore clinical imaging parameters such as  $SUV_{mean}$ ,  $SUV_{max}$ , and total lesion metabolism, providing a more comprehensive analysis of the IMCM model's performance. These metrics, along with an assessment of the most relevant radiomic features within the sphere of influence, will provide crucial insights into the effectiveness of the model under various clinical conditions.

Additionally, future investigations should focus on the performance of the IMCM model, specifically concerning artefact images, with a particular emphasis on organ-specific evaluations. This targeted approach would afford a nuanced understanding of how well the model performs in diverse clinical scenarios, potentially leading to significant improvements in model precision and utility. Furthermore, rigorous statistical tests on categorised outcomes, such as the marginal homogeneity test or the McNemar test, will be paramount. These tests will offer deeper insights into the consistency and reliability of the model across different diagnostic categories, helping to refine the model's application and enhance its diagnostic accuracy in practical healthcare settings.

# Conclusion

In this thesis, we have demonstrated the efficacy of an Integrated multi-Centre Dynamic Unet deep learning framework for artefact detection and correction in PET imaging of  $^{68}\text{Ga}$ -labelled compounds. The approach leverages large datasets from multiple centres. Through the incorporation of transfer learning concepts, we have developed site-specific models that significantly outperform centralized models and those based on single-centre data, thereby addressing a major limitation in the field of medical imaging.

Our model effectively detected and corrected artefacts. This enhancement is vital for making therapeutic decisions in the field of oncology, where PET imaging plays a central role in diagnosing, planning treatments, and evaluating responses. By using Dyn-Unet architecture and other advanced deep learning techniques, our method has not only improved image quality but also greatly decreased the appearance of common artefacts like halo and mismatch artefacts, especially in  $^{68}\text{Ga}$ -PET imaging. The effective implementation of our models in different centres highlights their resilience and flexibility, which are essential for general acceptance in clinical settings.

# References

1. Cerqueira MD. Cardiac SPECT or PET?: Is there still a debate? Vol. 29, *Journal of Nuclear Cardiology*. 2022.
2. Sarikaya I. Cardiac applications of PET. *Nucl Med Commun* [Internet]. 2015 Oct;36(10):971–85. Available from: <https://journals.lww.com/00006231-201510000-00002>
3. Catana C, Procissi D, Wu Y, Judenhofer MS, Qi J, Pichler BJ, et al. Simultaneous in vivo positron emission tomography and magnetic resonance imaging. *Proc Natl Acad Sci U S A*. 2008;105(10).
4. Boellaard R, Delgado-Bolton R, Oyen WJG, Giammarile F, Tatsch K, Eschner W, et al. FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. Vol. 42, *European Journal of Nuclear Medicine and Molecular Imaging*. 2015.
5. Karakatsanis NA, Fokou E, Tsoumpas C. Dosage optimization in positron emission tomography: state-of-the-art methods and future prospects. *Am J Nucl Med Mol Imaging*. 2015;5(5).
6. Fahey FH, Treves ST, Adelstein SJ. Minimizing and communicating radiation risk in pediatric nuclear medicine. *J Nucl Med Technol*. 2012;40(1).
7. Pettinato C, Nanni C, Farsad M, Castellucci P, Sarnelli A, Civollani S, et al. Artefacts of PET/CT images. *Biomed Imaging Interv J*. 2006;2(4).
8. Lammertsma AA. Forward to the past: The case for quantitative PET imaging. Vol. 58, *Journal of Nuclear Medicine*. 2017.
9. Sureshbabu W, Mawlawi O. PET/CT Imaging Artifacts\* [Internet]. Vol. 33, *J Nucl Med Technol*. 2005. Available from: [http://www.snm.org/ce\\_online](http://www.snm.org/ce_online)
10. Mawlawi O, Pan T, Macapinlac HA. PET/CT Imaging Techniques, Considerations, and Artifacts. *J Thorac Imaging* [Internet]. 2006;21(2). Available from: [https://journals.lww.com/thoracicimaging/fulltext/2006/05000/pet\\_ct\\_imaging\\_techniques,\\_considerations,\\_and.2.aspx](https://journals.lww.com/thoracicimaging/fulltext/2006/05000/pet_ct_imaging_techniques,_considerations,_and.2.aspx)
11. Shiri I, Salimi Y, Maghsudi M, Jenabi E, Harsini S, Razeghi B, et al. Differential privacy preserved federated transfer learning for multi-institutional 68Ga-PET image artefact detection and disentanglement. *Eur J Nucl Med Mol Imaging*. 2023;
12. Shiri I, Salimi Y, Hervier E, Pezzoni A, Sanaat A, Mostafaei S, et al. Artificial Intelligence-Driven Single-Shot PET Image Artifact Detection and Disentanglement: Toward Routine Clinical Image Quality Assurance. *Clin Nucl Med*. 2023 Dec 1;48(12):1035–46.
13. Lindemann ME, Nensa F, Quick HH. Impact of improved attenuation correction on 18F-FDG PET/MR hybrid imaging of the heart. *PLoS One*. 2019;14(3).
14. McQuaid SJ, Hutton BF. Sources of attenuation-correction artefacts in cardiac PET/CT and SPECT/CT. *Eur J Nucl Med Mol Imaging*. 2008;35(6).
15. Zaidi H, MML. Scatter Compensation Techniques in PET. *PET clinics*. *PET Clin* [Internet]. 2007 [cited 2023 Nov 20];2(2):219–34. Available from: <https://doi.org/10.1016/j.cpet.2007.10.003>

16. Baer M, Kachelrie M. Hybrid scatter correction for CT imaging. *Phys Med Biol.* 2012;57(21).
17. Watson CC, Casey ME, Michel C, Bendriem B. Advances in scatter correction for 3D PET/CT. In: *IEEE Nuclear Science Symposium Conference Record.* 2004.
18. Carney JPJ, Townsend DW, Rappoport V, Bendriem B. Method for transforming CT images for attenuation correction in PET/CT imaging. *Med Phys.* 2006;33(4).
19. Buchbender C, Hartung-Knemeyer V, Forsting M, Antoch G, Heusner TA. Positron emission tomography (PET) attenuation correction artefacts in PET/CT and PET/MRI. *British Journal of Radiology.* 2013;86(1025).
20. Watson CC. New, faster, image-based scatter correction for 3D PET. *IEEE Trans Nucl Sci.* 2000;47(4 PART 2).
21. Kinahan PE, Townsend DW, Beyer T, Sashin D. Attenuation correction for a combined 3D PET/CT scanner. *Med Phys.* 1998;25(10).
22. Alessio AM, Kohlmyer S, Branch K, Chen G, Caldwell J, Kinahan P. Cine CT for attenuation correction in cardiac PET/CT. *Journal of Nuclear Medicine.* 2007;48(5).
23. Akbarzadeh A, Ay MR, Ahmadian A, Riahi Alam N, Zaidi H. MRI-guided attenuation correction in whole-body PET/MR: Assessment of the effect of bone attenuation. *Ann Nucl Med.* 2013;27(2).
24. Abdoli M, Dierckx RAJO, Zaidi H. Metal artifact reduction strategies for improved attenuation correction in hybrid PET/CT imaging. Vol. 39, *Medical Physics.* 2012.
25. Ghafarian P, Aghamiri SMR, Ay MR, Rahmim A, Schindler TH, Ratib O, et al. Is metal artefact reduction mandatory in cardiac PET/CT imaging in the presence of pacemaker and implantable cardioverter defibrillator leads? *Eur J Nucl Med Mol Imaging.* 2011;38(2).
26. Heußner T, Mann P, Rank CM, Schäfer M, Dimitrakopoulou-Strauss A, Schlemmer HP, et al. Investigation of the halo-artifact in 68Ga-PSMA-11-PET/MRI. *PLoS One.* 2017;12(8).
27. Hong I, Nekolla SG, Michel C. Improving Scatter Correction for Ga-68 PSMA PET Studies. In: *2017 IEEE Nuclear Science Symposium and Medical Imaging Conference, NSS/MIC 2017 - Conference Proceedings.* 2018.
28. Magota K, Numata N, Shinyama D, Katahata J, Munakata Y, Maniawski PJ, et al. Halo artifacts of indwelling urinary catheter by inaccurate scatter correction in 18F-FDG PET/CT imaging: incidence, mechanism, and solutions. *EJNMMI Phys.* 2020;7(1).
29. Afshar-Oromieh A, Wolf M, Haberkorn U, Kachelrieß M, Gnirs R, Kopka K, et al. Effects of arm truncation on the appearance of the halo artifact in 68Ga-PSMA-11 (HBED-CC) PET/MRI. *Eur J Nucl Med Mol Imaging.* 2017;44(10).
30. Sarikaya I, Sarikaya A. PET/CT Image Artifacts Caused by the Arms. *J Nucl Med Technol.* 2021;49(1).
31. Lodge MA, Mhlanga JC, Cho SY, Wahl RL. Effect of patient arm motion in whole-body PET/CT. *Journal of Nuclear Medicine.* 2011;52(12).
32. Dinges J, Nekolla SG, Bundschuh RA. Motion artifacts in oncological and cardiac PET imaging. Vol. 8, *PET Clinics.* 2013.

33. Presotto L. The long fight against motion artifacts in cardiac PET. Vol. 29, *Journal of Nuclear Cardiology*. 2022.
34. Piccinelli M, Votaw JR, Garcia E V. Motion Correction and Its Impact on Absolute Myocardial Blood Flow Measures with PET. Vol. 20, *Current Cardiology Reports*. 2018.
35. Shiri I, Sanaat A, Salimi Y, Akhavanallaf A, Arabi H, Rahmim A, et al. PET-QA-Net: Towards Routine PET Image Artifact Detection and Correction using Deep Convolutional Neural Networks. In: 2021 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC). 2021. p. 1–3.
36. Shiri I, Arabi H, Geramifar P, Hajianfar G, Ghafarian P, Rahmim A, et al. Deep-JASC: joint attenuation and scatter correction in whole-body 18F-FDG PET using a deep residual network. *Eur J Nucl Med Mol Imaging*. 2020 Oct 1;47(11):2533–48.
37. Arabi H, Zaidi H. Truncation compensation and metallic dental implant artefact reduction in PET/MRI attenuation correction using deep learning-based object completion. *Phys Med Biol*. 2020;65(19).
38. Mawlawi O, Erasmus JJ, Pan T, Cody DD, Campbell R, Lonn AH, et al. Truncation artifact on PET/CT: Impact on measurements of activity concentration and assessment of a correction algorithm. *American Journal of Roentgenology*. 2006;186(5).
39. Lindemann ME, Gratz M, Blumhagen JO, Jakoby B, Quick HH. MR-based truncation correction using an advanced HUGE method to improve attenuation correction in PET/MR imaging of obese patients. *Med Phys*. 2022;49(2).
40. Yoon SH, Jang JS, Park C. The feasibility of maximum likelihood estimation of activity and attenuation (MLAA) algorithm for reduction of truncation artifact in the breast PET/MRI. *Journal of the Korean Physical Society*. 2022;81(2).
41. Panagiotidis E, Alshammari A, Michopoulou S, Skoura E, Naik K, Maragkoudakis E, et al. Comparison of the impact of 68Ga-DOTATATE and 18F-FDG PET/CT on clinical management in patients with Neuroendocrine tumors. *Journal of Nuclear Medicine*. 2017;58(1).
42. Shiri I, Vafaei Sadr A, Akhavan A, Salimi Y, Sanaat A, Amini M, et al. Decentralized collaborative multi-institutional PET attenuation and scatter correction using federated deep learning. *Eur J Nucl Med Mol Imaging*. 2023 Mar 1;50(4):1034–50.
43. Lassen ML, Rasul S, Beitzke D, Stelzmüller ME, Cal-Gonzalez J, Hacker M, et al. Assessment of attenuation correction for myocardial PET imaging using combined PET/MRI. *Journal of Nuclear Cardiology*. 2019;26(4).
44. Presotto L, Busnardo E, Perani D, Gianolli L, Gilardi MC, Bettinardi V. Simultaneous reconstruction of attenuation and activity in cardiac PET can remove CT misalignment artifacts. *Journal of Nuclear Cardiology*. 2016;23(5).
45. Mostafapour S, Greuter M, van Snick JH, Brouwers AH, Dierckx RAJO, van Sluis J, et al. Ultra-low dose CT scanning for PET/CT. *Med Phys*. 2024;51(1).
46. Xia T, Alessio AM, Kinahan PE. Limits of ultra-low dose CT attenuation correction for PET/CT. In: *IEEE Nuclear Science Symposium Conference Record*. 2009.



47. Prieto E, García-Velloso MJ, Aquerreta JD, Rosales JJ, Bastidas JF, Soriano I, et al. Ultra-low dose whole-body CT for attenuation correction in a dual tracer PET/CT protocol for multiple myeloma. *Physica Medica*. 2021;84.
48. Zaidi H, Koral KF. Scatter modelling and compensation in emission tomography. Vol. 31, *European Journal of Nuclear Medicine and Molecular Imaging*. 2004.
49. Li S, Wang G. Modified kernel MLAA using autoencoder for PET-enabled dual-energy CT. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*. 2021;379(2204).
50. Chun SY, Kim KY, Lee JS, Fessler JA. Joint estimation of activity distribution and attenuation map for TOF-PET using alternating direction method of multiplier. In: *Proceedings - International Symposium on Biomedical Imaging*. 2016.
51. Mehranian A, Arabi H, Zaidi H. Vision 20/20: Magnetic resonance imaging-guided attenuation correction in PET/MRI: Challenges, solutions, and opportunities. *Med Phys*. 2016;43(3).
52. Guo R, Xue S, Hu J, Sari H, Mingels C, Zeimpekis K, et al. Using domain knowledge for robust and generalizable deep learning-based CT-free PET attenuation and scatter correction. *Nat Commun*. 2022 Dec 1;13(1).
53. Yang J, Sohn JH, Behr SC, Gullberg GT, Seo Y. Ct-less direct correction of attenuation and scatter in the image space using deep learning for whole-body fdg pet: Potential benefits and pitfalls. *Radiol Artif Intell*. 2021 Mar 1;3(2).
54. Shiri I, Ghafarian P, Geramifar P, Leung KHY, Ghelichoghli M, Oveisi M, et al. Direct attenuation correction of brain PET images using only emission data via a deep convolutional encoder-decoder (Deep-DAC). *Eur Radiol*. 2019 Dec 1;29(12):6867–79.
55. Lee JS. A Review of Deep-Learning-Based Approaches for Attenuation Correction in Positron Emission Tomography. Vol. 5, *IEEE Transactions on Radiation and Plasma Medical Sciences*. 2021.
56. Qian H, Rui X, Ahn S. Deep Learning Models for PET Scatter Estimations. In: *2017 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*. 2017. p. 1–5.
57. Liu F, Jang H, Kijowski R, Zhao G, Bradshaw T, McMillan AB. A deep learning approach for 18 f-fdg pet attenuation correction. *EJNMMI Phys*. 2018;5(1).
58. Wafa B, Moussaoui A. A review on methods to estimate a CT from MRI data in the context of MRI-alone RT. *Medical Technologies Journal*. 2018;2(1).
59. Sun H, Xi Q, Fan R, Sun J, Xie K, Ni X, et al. Synthesis of pseudo-CT images from pelvic MRI images based on an MD-CycleGAN model for radiotherapy. *Phys Med Biol*. 2022;67(3).
60. Wang T, Manohar N, Lei Y, Dhabaan A, Shu HK, Liu T, et al. MRI-based treatment planning for brain stereotactic radiosurgery: Dosimetric validation of a learning-based pseudo-CT generation method. *Medical Dosimetry*. 2019;44(3).
61. Jabbarpour A, Mahdavi SR, Vafaei Sadr A, Esmaili G, Shiri I, Zaidi H. Unsupervised pseudo CT generation using heterogeneous multicentric CT/MR images and CycleGAN: Dosimetric assessment for 3D conformal radiotherapy. *Comput Biol Med*. 2022;143.

62. Liu F, Jang H, Kijowski R, Bradshaw T, McMillan AB. Deep learning MR imaging-based attenuation correction for PET/MR imaging. *Radiology*. 2018;286(2).
63. Arabi H, Zaidi H. Deep learning-based metal artefact reduction in PET/CT imaging. *Eur Radiol*. 2021;31(8).
64. Shiri I, Salimi Y, Sanaat A, Saberi A, Amini M, Akhavanallaf A, et al. Fully Automated PET Image Artifacts Detection and Correction Using Deep Neural Networks </strong> Journal of Nuclear Medicine [Internet]. 2022 Jun 1;63(supplement 2):3218. Available from: [http://jnm.snmjournals.org/content/63/supplement\\_2/3218.abstract](http://jnm.snmjournals.org/content/63/supplement_2/3218.abstract)
65. Shiri I, Sadr A V, Sanaat A, Ferdowsi S, Arabi H, Zaidi H. Federated Learning-based Deep Learning Model for PET Attenuation and Scatter Correction: A Multi-Center Study. In: 2021 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC). 2021. p. 1–3.
66. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*. 1995;57(1).