# **Material and methods**

## Data Preparation

Our study aimed to evaluate the performance of our model across different scenarios, including various external scanners and radiotracers. The research utilized a primary dataset of Gallium PET/CT images for training and initial model validation, encompassing scans from multiple hospital imaging centers. To test the model's adaptability, a secondary dataset was incorporated, distinct in both the imaging centers and the type of radiotracer used. Additionally, a specialized set of images presenting with artifacts 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 68Ga-prostate-specific membrane antigen (PSMA)/DOTA-TATE (TOC) PET/CT imaging across five centers 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 attenuation and scatter correction (CT-ASC) was applied to amend PET images.

Table 1: Data information in 5 different imaging centers.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Center** | **No** | **Train** | **Validation** | **Test** | **Scanner** | **Reconstruction** | **Matrix size × Z\*** |
| Center 1 | 56 | 43 | 11 | 2 | Siemens Biograph 6 | 3D-OSEM | 168 × 168 |
| Center 2 | 31 | 25 | 4 | 2 | GE Discovery IQ | 3D-OSEM | 192 × 192 |
| Center 3 | 45 | 35 | 8 | 2 | Siemens mCT | 3D-OSEM | 200 × 200 |
| Center 4 | 40 | 28 | 10 | 2 | Siemens Biograph 6 | 3D-OSEM | 168 × 168 |
| External Center | 12 | - | - | 12 | Siemens Horizon | PSF+TOF+3D-OSEM | 180 × 180 |
| **Total** | **184** | **131** | **33** | **20** | - | - | - |
| \* 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 standardized uptake value (SUV) is a crucial quantitative measure that normalizes 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:

|  |  |
| --- | --- |
|  | ( 1) |

This conversion was applied uniformly across all MAC and NAC images in our dataset to standardize the voxel values into standardized uptake value (SUV) metrics.

To achieve uniformity across all images, the voxel intensities were normalized by dividing by a constant factor. Specifically, NAC images were scaled down by a factor of 2, and MAC images by a factor of 5. This approach of normalization by a constant factor simplifies the process by applying a uniform scale adjustment across the dataset, maintaining the relative differences in radiotracer uptake between various regions 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-normalization 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 standardize the dimensions to a uniform bounding box size of 168×168×Z (with 'Z' representing the count of slices), as illustrated in the 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. For uniformity and to enhance the training process's efficiency, all PET images were re-scaled to a voxel size of 4.07 × 4.07 × 3.0 mm³, the most common resolution across the collected data. This standardization was crucial for achieving consistent image quality throughout the dataset. Details regarding the initial voxel spacing are provided in Figure 2b.

|  |
| --- |
| A graph of a graph  Description automatically generated |
| Figure 1: Distribution of maximum intensity values for NAC and MAC images, displaying variations pre- and post-normalization to highlight data scaling effects. non-ASC images were scaled down by a factor of 2, and CT-ASC images by a factor of 5. |

|  |
| --- |
| A graph with green bars  Description automatically generated |
|  |
| Figure 2: **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 (1). We decomposed the complex end-to-end generation from NAC to MAC (model-based attenuation correction) into two components, **anatomy-independent textures** (relating to tracers and diseases) and **anatomy-dependent correction.**

In other word, 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 Anatomy-Dependent Correction Map (ADCM) at each voxel is defined by the ratio of the MAC intensity to the NAC intensity, encapsulated by the following conditional equation:

|  |  |
| --- | --- |
|  | ( 2) |

The threshold ε 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):

|  |  |
| --- | --- |
|  | ( 3) |

Sample cases are visualized in figure 4.

#### Normalization

In the calibration of ADCM, as we mentioned before, conventional min-max normalization techniques were consciously avoided to preserve the quantitative fidelity of SUV metrics—essential for valid clinical interpretations. We established an empirical normalization factor specifically for ADCM values, which was meticulously determined to bring the expansive range of the dataset within a more confined scope appropriate for deep learning applications. This factor ensures the broad spectrum of data, ranging from minimal to several thousand units, is normalized in a way that permits later recalibration into their original SUV metrics. Notably, extreme values which 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 normalized using a factor of 50 to maintain the relative and comparable and manageable values for training. The resultant histograms, illustrating the distribution of maximum values both pre- and post-normalization, are depicted in Figure 3.

A group of graphs showing different colors

Description automatically generated

Figure 3: Distribution of maximum intensity values for NAC and MAC images and ADCM metrics, displaying variations pre- and post-normalization to highlight data scaling effects. non-ASC images were scaled down by a factor of 2, CT-ASC images by a factor of 5 and ADCM by a factor of 50.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | | NAC-PET MAC-PET ADCM | | | NAC-PET MAC-PET ADCM | |
| **a)** | A comparison of a body scan  Description automatically generated with medium confidence | | | A comparison of a body scan  Description automatically generated with medium confidence | |
| **b)** | A several images of a person's body  Description automatically generated | | | A comparison of a body scan  Description automatically generated with medium confidence | |
| **c)** | A close-up of several images of a person's body  Description automatically generated | | | A comparison of x-ray images of a human body  Description automatically generated | |
|  | | | Figure 4: the middle slice of the coronal view for NAC, MAC, and ADCM images. Color bar unit: SUV | | | |

### FDG Datasets

To assess the model's performance with various radiotracers, our study incorporated a dataset of 20 whole-body 18F-FDG PET scans originating from two distinct hospitals, representing our external radiotracer dataset. In the preprocessing phase, voxel intensities from both MAC (CT-based attenuation corrected) and NAC (non-attenuation scatter corrected) images were standardized to SUVs, normalizing the dynamic range of intensities to optimize the efficiency of network training. Empirical scaling factors, 9 for MAC and 3 for NAC images, were applied to further constrain the dynamic range.

To achieve homogeneity across the dataset, we standardized 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 standardized spacings ensure uniformity across all scans in preparation for model training. Further details on the imaging parameters are presented in Table 2.

Table 2: "Overview of External Radiotracer Dataset Specifications.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Center** | **Test** | **Scanner** | **Reconstruction** | **Matrix size × Z** |
| Center 6 | 10 | ? | ? | 272 × 200 |
| Center 7 | 10 | ? | ? | 272 × 200 |
| Total | 20 | - | - | - |

### Artifact dataset

In this study, a secondary test set was utilized 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 artifacts. Artifacts included in this dataset were specifically chosen to test the model's robustness and ability to accurately interpret images compromised by common distortions found in clinical Ga imaging, such as motion and Halo artifact.

## Deep neural network

For final implementation, we leverage the Dyn-UNet architecture, renowned for its adaptability and efficiency in processing biomedical images (2). This model is particularly chosen for its dynamic configuration and deep supervision, enabling precise results tailored to the specific requirements of our dataset (2).

The initialization of the Dyn-UNet model is uniquely designed to compute the optimal kernel sizes and strides based on the input patch size and spacing of 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 the supervision heads. Deep supervision ensures that intermediate layers are also optimized 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 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 optimizing 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 ReLU activation function in the last layer, we can get the non-zero value with the concept of 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 optimization 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 optimization of the network was conducted using the Adam algorithm, with the aim to minimize 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 ±15 degrees and increasing the number of samples per patient from 4 to 20.

To maintain the integrity of the model, only artifact-free datasets were used during the network's training and validation stages. We trained the network over 400 epochs to ensure adequate convergence. 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, there was no overlap of patients across the training, testing, and validation datasets.

Details on alternative models tested, which did not meet our criteria for inclusion in the final report, are documented in Supplementary Material 2.

## Training approaches for deep learning models:

### Integrated multi-Center model:

A Dyn-Unet deep learning model was developed using a combined dataset from four different centers, all utilizing gallium-based tracers. This model was initially trained on a collective dataset and subsequently tested on an external center's data to evaluate its generalization capabilities. It was also tested within the originating dataset from each center. This approach aims to overcome the limitations of models trained on data from single centers, which may struggle with generalizability to new, unseen cases.

### Tune TL model:

To address the challenges encountered with different radiotracers, the model underwent tuning through Transfer Learning (TL). This method involves modifying the deep learning model by integrating learnings from decentralized 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.

### Decomposition Model:

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, is employed to focus exclusively on estimating the anatomy-dependent correction maps (ADCM).

This model's effectiveness is evaluated through its ability to generalize across different centers and tracers, testing its robustness in a variety of clinical settings.

## 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):** Reflects the average deviation across all voxels.

( 4)

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

( 5)

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

( 6)

* **Absolute Relative Error (ARE%):** Captures the absolute percentage difference, ensuring that all deviations are treated equally regardless of their direction.

( 7)

Where tot refers to the total number of voxels, and and 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.

|  |  |
| --- | --- |
|  | ( 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.

|  |  |
| --- | --- |
|  | ( 9) |

where: and are the averages of the pixel intensities in the predicted PET images () and the CT-attenuation corrected PET images (), respectively. and are the variances of the pixel intensities in the predicted and CT-attenuation corrected PET images, respectively. is the covariance of the predicted and CT-attenuation corrected PET images.

and are constants to stabilize the division with a weak denominator; L is the dynamic range of the pixel values (typically ). = 0.01 and =0.03 are default values for the stabilization constants.

1. 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).

2. Isensee F, Petersen J, Klein A, Zimmerer D, Jaeger PF, Kohl S, et al. nnU-Net: Self-adapting Framework for U-Net-Based Medical Image Segmentation. In: Informatik aktuell. 2019.