



# Self-supervised Learning for Physiologically-Based Pharmacokinetic Modeling in Dynamic PET

Francesca De Benetti<sup>1(✉)</sup>, Walter Simson<sup>2</sup>, Magdalini Paschali<sup>3</sup>, Hasan Sari<sup>4</sup>,  
Axel Rominger<sup>5</sup>, Kuangyu Shi<sup>1,5</sup>, Nassir Navab<sup>1</sup>, and Thomas Wendler<sup>1</sup>

<sup>1</sup> Chair for Computer-Aided Medical Procedures and Augmented Reality,  
Technische Universität München, Garching, Germany  
[francesca.de-benetti@tum.de](mailto:francesca.de-benetti@tum.de)

<sup>2</sup> Department of Radiology, Stanford University School of Medicine, Stanford, USA

<sup>3</sup> Department of Psychiatry and Behavioral Sciences,  
Stanford University School of Medicine, Stanford, USA

<sup>4</sup> Advanced Clinical Imaging Technology, Siemens Healthcare AG,  
Lausanne, Switzerland

<sup>5</sup> Department of Nuclear Medicine, Bern University Hospital, Bern, Switzerland

**Abstract.** Dynamic Positron Emission Tomography imaging (dPET) provides temporally resolved images of a tracer. Voxel-wise physiologically-based pharmacokinetic modeling of the Time Activity Curves (TAC) extracted from dPET can provide relevant diagnostic information for clinical workflow. Conventional fitting strategies for TACs are slow and ignore the spatial relation between neighboring voxels. We train a spatio-temporal UNet to estimate the kinetic parameters given TAC from dPET. This work introduces a self-supervised loss formulation to enforce the similarity between the measured TAC and those generated with the learned kinetic parameters. Our method provides quantitatively comparable results at organ level to the significantly slower conventional approaches while generating pixel-wise kinetic parametric images which are consistent with expected physiology. To the best of our knowledge, this is the first self-supervised network that allows voxel-wise computation of kinetic parameters consistent with a non-linear kinetic model.

**Keywords:** Kinetic modelling · PBPK models · Dynamic PET

## 1 Introduction

Positron Emission Tomography (PET) is a 3D imaging modality using radiopharmaceuticals, such as F-18-fluorodeoxyglucose (FDG), as tracers. Newly introduced long axial field-of-view PET scanners have enabled dynamic PET (dPET)

**Supplementary Information** The online version contains supplementary material available at [https://doi.org/10.1007/978-3-031-43907-0\\_28](https://doi.org/10.1007/978-3-031-43907-0_28).

with frame duration  $< 1$  min [18], allowing the observation of dynamic metabolic processes throughout the body. For a given voxel in space, the radioactivity concentration over time can be described by a characteristic curve known as Time Activity Curve (TAC), measured in [Bq/ml]. TACs can be described by mathematical functions, called physiologically-based pharmacokinetic (PBPK) models or kinetic models (KM) [14]. The parameters of the KM represent physiologically relevant quantities and are often called *micro-parameters*, whereas their combinations are called *macro-parameters* [14,20]. While the former can be retrieved only by methods that directly use the KM function, the latter can be computed by simplified linearized methods (such as the Logan and the Patlak-Gjedde plots). The approaches to extract KM parameters can be split in two categories: Volume of Interest (VoI) methods, in which the average TAC in a VoI is used, or voxel-based methods. Despite the former displaying less noise and, therefore, lower variance in the kinetic parameters (KPs), VoI-based methods only provide organ-wise information. On the other hand, voxel-based methods allow the generation of parametric images (KPIs), in which the KPs are visualized at a voxel level, but suffer from motion and breathing artifacts, and require more computational power or simplified linearized methods.

Parametric images are reported to be superior in lesion detection and delineation when compared to standard-of-care activity- and weight-normalized static PET volumes, known as Standard Uptake Value (SUV) volumes [6,7]. Changes in the KPs during oncological therapy are associated with pathological response to treatment, whereas this is not true for changes in SUV [15]. Despite the advantages of KPIs in diagnosis, the generation of accurate micro-parametric images is not yet possible in clinical practice.

To address the problem of the generation of micro-parametric images, we propose a custom 3D UNet [3] to estimate kinetic *micro-parameters* in an unsupervised setting drawing inspiration from physics-informed neural networks (PINN). The main contributions of this work are:

- A self-supervised formulation of the problem of kinetic *micro-parameters* estimation
- A spatio-temporal deep neural network for parametric images estimation
- A quantitative and qualitative comparison with conventional methods for PBPK modeling

The code is available at:

[https://github.com/FrancescaDB/self\\_supervised\\_PBPK\\_modelling](https://github.com/FrancescaDB/self_supervised_PBPK_modelling).

## 1.1 Related Work

Finding the parameters of a KM is a classical optimization problem [2,19,21] solved by fitting the KM equation to a measured TAC in a least squares sense [1,14,17]. The non-linearity of the KM functions makes this approach prone to overfitting and local minima, and sensitive to noise [14]. Therefore, non-linear parametric imaging is still too noisy for clinical application [20].

To limit the drawbacks of the non-linear parameter fitting, the identification of KPs is commonly performed using simplified linearized versions of the KM [6,20], such as the Logan and the Patlak-Gjedde plots [5,16], which are often included in clinical software for KM such as PMOD<sup>1</sup>.

Preliminary works towards KM parameter estimation in dPET imaging have recently begun to be explored. Moradi et al. used an auto-encoder along with a Gaussian process regression block to select the best KM to describe simulated kinetic data [13]. A similar approach was presented for the quantification of myocardial blood flow from simulated PET sinograms [11]. Huang et al. used a supervised 3D U-Net to predict a macro-parametric image using an SUV image as input, and a Patlak image derived from dPET acquisition as ground truth [9]. Cui et al. proposed a conditional deep image prior framework to predict a macro-parametric image using a DNN in an unsupervised setting [4]. Finally, a supervised DNN was used to predict Patlak KPIs from Dynamic PET sinograms [12]. Until now, methods used simulated data [11,13] or static PET [9], were supervised [9,11–13] or predicted *macro-parameters* [4,9,12].

## 2 Methodology and Materials

We propose to compute the kinetic *micro-parameters* in a self-supervised setting by directly including the KM function in the loss function and comparing the predicted TAC to the measured TAC. For this reason, an understanding of the KM is fundamental to describing our pipeline.

### 2.1 Kinetic Modelling

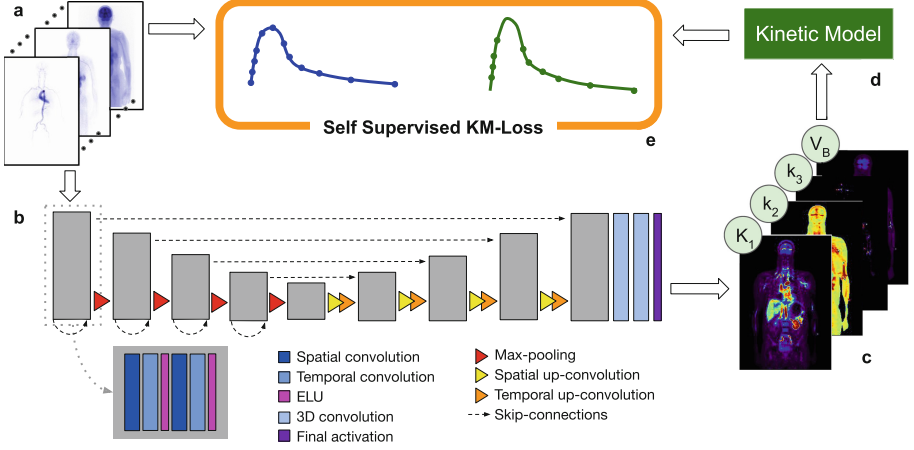
The concentration of the tracer  $\tilde{C}(t)$  [Bq/ml] in each tissue can be described as a set of ordinary differential equations [20]. It represents the interaction of two compartments,  $F(t)$  (free) and  $B(t)$  (bound), and takes as input the radioactivity concentration in blood or plasma  $A(t)$  [14]:

$$\begin{aligned}\frac{dF(t)}{dt} &= K_1 A(t) - (k_2 + k_3)F(t) \\ \frac{dB(t)}{dt} &= k_3 F(t) - k_4 B(t) \\ \tilde{C}(t) &= F(t) + B(t)\end{aligned}\tag{1}$$

where  $K_1$  [ml/cm<sup>3</sup>/min],  $k_2$  [1/min],  $k_3$  [1/min] and  $k_4$  [1/min] are the *micro-parameters* [6,20]. Equation 1 describes a general two-tissue compartment (2TC) kinetic model. However, an FDG-TAC is conventionally described by an irreversible 2TC, in which  $k_4$  is set to 0 [20]. Therefore, in the following, we will use  $k_4 = 0$ .

Moreover, including the blood fraction volume  $V_B$  [·] allows to correctly model the contribution to the radioactivity in a voxel coming from vessels that are too small to be resolved by the PET scanner [14].

<sup>1</sup> <https://www.pmod.com>.



**Fig. 1.** Proposed Pipeline: a sequence of dPET slices (a) is processed by the proposed DNN (b) to estimate the 4 KPIs (c). During training, the predicted TACs are computed by the KM (d) and compared to the measured TACs using the mean square error (e).

Together, the TAC of each voxel in an FDG dPET acquisition can be modeled as  $C(t) = (1 - V_B)\tilde{C}(t) + V_B A(t)$ , and solved using the Laplace transform [20]:

$$C(t) = (1 - V_B) \left[ \frac{K_1}{k_2 + k_3} [k_3 + k_2 e^{-(k_2 + k_3)t}] * A(t) \right] + V_B A(t). \quad (2)$$

## 2.2 Proposed Pipeline

Our network takes as input a sequence of 2D axial slices and returns a 4-channel output representing the spatial distribution of the KM parameters of a 2TC for FDG metabolism [14]. The network has a depth of four, with long [3] and short skip connections [10]. The kernel size of the max-pooling is [2, 2, 2]. After the last decoder block, two 3D convolutional layers (with kernel size [3, 3, 3] and [64, 1, 1]) estimate the KPs per voxel given the output feature of the network. Inside the network the activation function is ELU and critically batch normalization is omitted. The network was trained with an initial learning rate of  $10^{-4}$ , which was divided by half every 25 epochs, for a maximum of 500 epochs.

Following the approach taken by Küstner et al. for motion correction of 4D spatio-temporal CINE MRI [10], we replaced a conventional 3D convolutional layer with (2+1)D spatial and temporal convolutional layers. The spatial convolutional layer is a 3D convolutional layer with kernel size [1, 3, 3] in [t, x, y]. Similarly, the temporal convolutional layer has a kernel size of [3, 1, 1].

We imposed that the KPs predicted by the network satisfy Eq. 2 by including it in the computation of the loss. At a pixel level, we computed the mean squared error between the TAC estimated using the corresponding predicted parameters ( $\tilde{TAC}_i$ ) and the measured one ( $TAC_i$ ), as seen in Fig. 1.

We introduced a final activation function to limit the output of the network to the valid parameter domain of the KM function. Using the multi-clamp function, each channel of the logits is restricted to the following parameter spaces:  $K_1 \in [0.01, 2]$ ,  $k_2 \in [0.01, 3]$ ,  $k_3 \in [0.01, 1]$ , and  $V_B \in [0, 1]$ . The limits of the ranges were defined based on the meaning of the parameter (as in  $V_B$ ), mathematical requirements (as in the minimum values of  $k_2$  and  $k_3$ , whose sum can not be zero) [6] or previous knowledge on the dataset derived by the work of Sari et al. [16] (as in the maximum values of  $K_1$ ,  $k_2$  and  $k_3$ ).

We evaluated the performance of the network using the Mean Absolute Error (MAE) and the Cosine Similarity (CS) between  $TAC_i$  and  $\hat{TAC}_i$ .

### 2.3 Curve Fit

For comparison, parameter optimization via non-linear fitting was implemented in Python using the `scipy.optimize.curve_fit` function (version 1.10), with step equal to 0.001. The bounds were the same as in the DNN.

### 2.4 Dataset

The dataset is composed of 23 oncological patients with different tumor types. dPET data was acquired on a Biograph Vision Quadra for 65 min, over 62 frames. The exposure duration of the frames were  $2 \times 10$  s,  $30 \times 2$  s,  $4 \times 10$  s,  $8 \times 30$  s,  $4 \times 60$  s,  $5 \times 120$  s and  $9 \times 300$  s. The PET volumes were reconstructed with an isotropic voxel size of 1.6 mm. The dataset included the label maps of 7 organs (bones, lungs, heart, liver, kidneys, spleen, aorta) and one image-derived input function  $A(t)$  [Bq/ml] from the descending aorta per patient. Further details on the dataset are presented elsewhere [16].

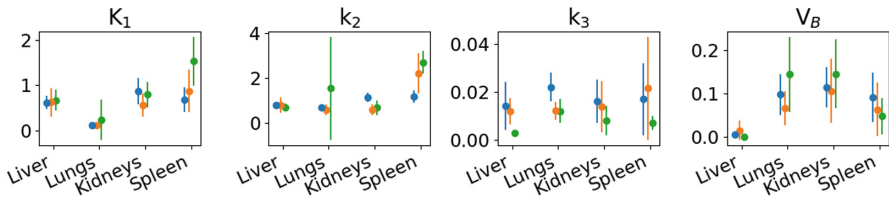
The PET frames and the label map were resampled to an isotropic voxel size of 2.5 mm. Then, the dataset was split patient-wise into training, validation, and test set, with 10, 4, and 9 patients respectively. Details on the dataset split are available in the Supplementary Material (Table 1). The training set consisted of 750 slices and the validation consisted of 300. In both cases, 75 axial slices per patient were extracted in a pre-defined patient-specific range from the lungs to the bladder (included) and were cropped to size  $112 \times 112$  pixels.

## 3 Results

Table 1 shows the results of the 8 ablation studies we performed to find the best model. We evaluated the impact of the design of the convolutional and max-pooling kernels, as well as the choice of the final activation function. The design of the max pooling kernel (i.e., kernel size equal to  $[2, 2, 2]$  or  $[1, 2, 2]$ ) had no measurable effects in terms of CS in most of the experiments, with the exception of Exp. 3.2, where max-pooling only in space resulted in a drop of 0.06. When evaluating the MAE, the use of 3D max-pooling was generally better.

**Table 1.** Configurations and metrics of ablation studies for architecture optimization.

Exp	Convolution	Pooling	Final activation	CS $\uparrow$	MAE $\downarrow$
1.1	3D	3D	Absolute	$0.74 \pm 0.05$	$3.55 \pm 2.12$
1.2	3D	space	Absolute	$0.74 \pm 0.05$	$3.64 \pm 2.21$
2.1	space + time	3D	Absolute	$0.75 \pm 0.05$	$3.59 \pm 2.33$
2.2	space + time	space	Absolute	$0.75 \pm 0.05$	$3.67 \pm 2.20$
3.1	space + time	3D	Clamp	$0.75 \pm 0.05$	$3.48 \pm 2.04$
3.2	space + time	space	Clamp	$0.69 \pm 0.05$	$3.55 \pm 2.07$
4.1	space + time	3D	Multi-clamp	<b><math>0.78 \pm 0.05</math></b>	$3.28 \pm 2.03$
4.2	space + time	space	Multi-clamp	$0.77 \pm 0.05$	<b><math>3.27 \pm 2.01</math></b>

**Fig. 2.** Comparison between the kinetic parameters obtained with different methods:  $KP_{DNN}$  in blue,  $KP_{CF}$  in orange and, as plausibility check,  $KP_{CF}^{ref}$  [16] in green. The exact values are reported in the Supplementary Material (Table 3 and 4) and in [16]. (Color figure online)

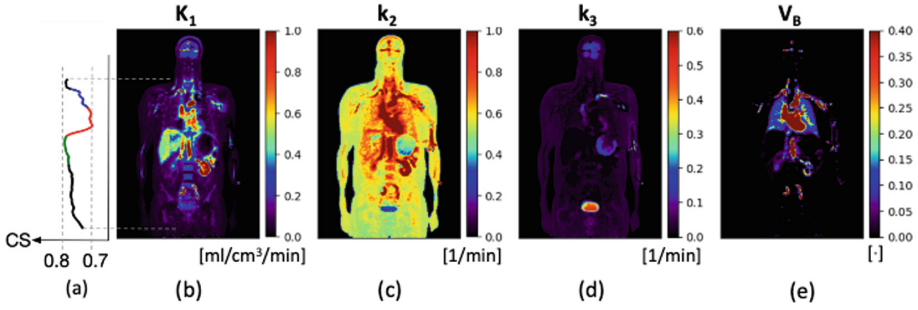
The most important design choice is the selection of the final activation function. Indeed, the multi-clamp final activation function was proven to be the best both in terms of CS (Exp 4.1:  $CS = 0.78 \pm 0.05$ ) and MAE (Exp 4.2:  $MAE = 3.27 \pm 2.01$ ). Compared to the other final activation functions, when the multi-clamp is used the impact of the max-pooling design is negligible also in terms of MAE. For the rest of the experiments, the selected configuration is the one from Exp. 4.1 (see Table 1).

Figure 2 shows the KPs for four selected organs as computed with the proposed DNN ( $KP_{DNN}$ ), as computed with curve fit using only the 9 patients of the test set ( $KP_{CF}$ ) and using all 23 patients ( $KP_{CF}^{ref}$ ) [16]. The voxel-wise KPs predicted by the DNN were averaged over the available organ masks.

In terms of run-time, the DNN needed  $\approx 1$  min to predict the KPs of the a whole-body scan ( $\approx 400$  slices), whereas curve fit took 8.7 min for a single slice: the time reduction of the DNN is expected to be  $\approx 3.500$  times.

## 4 Discussion

Even though the choice of the final activation function has a greater impact, the selection of the kernel design is important. Using spatial and temporal convo-



**Fig. 3.** (a) Cosine similarity (CS) per slice in patient 23 (blue: lungs; red: lungs and heart; green: liver). (b–e) Parametric images of a coronal slice for the same patient. (Color figure online)

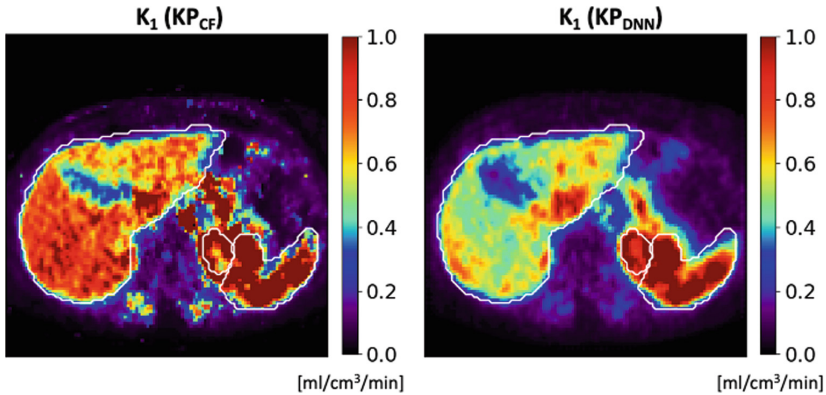
lution results in an increase in the performance (+0.01 in CS) and reduces the number of trainable parameters (from 2.1 M to 8.6 K), as pointed out by [10]. Therefore, the convergence is reached faster. Moreover, the use of two separate kernels in time and space is especially meaningful. Pixel counts for a given exposure are affected by the neighboring count measurements due to the limited resolution of the PET scanner [20]. The temporally previous or following counts are independent. In general, there is good agreement between  $KP_{DNN}$ ,  $KP_{CF}$  and  $KP_{CF}^{ref}$ . The DNN prediction of  $K_1$  and  $k_2$  in the spleen and  $k_3$  in the lungs is outside the confidence interval of the results published by Sari et al. [16].

An analysis per slice of the metrics shows that the CS between  $TAC_i$  and  $\tilde{TAC}_i$  changes substantially depending on the region:  $CS_{max} = 0.87$  within the liver boundaries and  $CS_{min} = 0.71$  in the region corresponding to the heart and lungs (see Fig. 3a). This can be explained by the fact that  $V_B$  is underestimated for the heart and aorta. The proposed network predicts  $V_B^{heart} = 0.376 \pm 0.133$  and  $V_B^{aorta} = 0.622 \pm 0.238$  while values of nearly 1 are to be expected. This is likely due to breathing and heartbeat motion artifacts, which cannot be modeled properly with a 2TC KM that assumes no motion between frames.

Figure 3b–e shows the central coronal slice of the four KPIs in an exemplary patient. As expected,  $K_1$  is high in the heart, liver, and kidney. Similarly, the blood fraction volume  $V_B$  is higher in the heart, blood vessels, and lungs.

The  $KP_{DNN}$  are more homogeneous than  $KP_{CF}$ , as can be seen in the exemplary  $K_1$  axial slice shown in Fig. 4. A quantitative evaluation of the smoothness of the images is reported in the Supplementary Material (Fig. 1). Moreover, the distribution in the liver is more realistic in  $KP_{DNN}$ , where the gallbladder can be seen as an ellipsoid between the right and left liver lobes. High  $K_1$  regions are mainly within the liver, spleen, and kidney for  $KP_{DNN}$ , while they also appear in unexpected areas in the  $KP_{CF}$  (e.g., next to the spine or in the region of the stomach).

The major limitation of this work is the lack of ground truth and a canonical method to evaluate quantitatively its performance. This limitation is inherent



**Fig. 4.** Comparison of  $K_1$  parametric images for an axial slice of patient 2, with contours of the liver (left), the spleen (center) and the left kidney (right).

to PBPK modeling and results in the need for qualitative analyses based on expected physiological processes. A possible way to leverage this would be to work on simulated data, yet the validity of such evaluations strongly depends on how realistic the underlying simulation models are. As seen in Fig. 3a, motion (gross, respiratory, or cardiac) has a major impact on the estimation quality. Registering different dPET frames has been shown to improve conventional PBPK models [8] and would possibly have a positive impact on our approach.

## 5 Conclusion

In this work, inspired by PINNs, we combine a self-supervised spatio-temporal DNN with a new loss formulation considering physiology to perform kinetic modeling of FDG dPET. We compare the best DNN model with the most commonly used conventional PBPK method, curve fit. While no ground truth is available, the proposed method provides similar results to curve fit but qualitatively more plausible images in physiology and with a radically shorter run-time.

Further, our approach can be applied to other KMs without significantly increasing the complexity and the need for computational power. In general, Eq. 2 should be modified to represent the desired KM [20], and the number of channels of the output of the network should be the same as the KP to be predicted.

Overall, this work offers scalability and a new research direction for analysing pharmacokinetics.

**Acknowledgements.** This work was partially funded by the German Research Foundation (DFG, grant NA 620/51-1).



## References

1. Avula, X.J.: Mathematical modeling. In: Meyers, R.A. (ed.) *Encyclopedia of Physical Science and Technology*, 3rd edn., pp. 219–230. Academic Press, New York (2003)
2. Besson, F.L., et al.: 18F-FDG PET and DCE kinetic modeling and their correlations in primary NSCLC: first voxel-wise correlative analysis of human simultaneous [18F] FDG PET-MRI data. *EJNMMI Res.* **10**(1), 1–13 (2020)
3. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) *MICCAI 2016*. LNCS, vol. 9901, pp. 424–432. Springer, Cham (2016). [https://doi.org/10.1007/978-3-319-46723-8\\_49](https://doi.org/10.1007/978-3-319-46723-8_49)
4. Cui, J., Gong, K., Guo, N., Kim, K., Liu, H., Li, Q.: Unsupervised PET logan parametric image estimation using conditional deep image prior. *Med. Image Anal.* **80**, 102519 (2022)
5. Dias, A.H., Hansen, A.K., Munk, O.L., Gormsen, L.C.: Normal values for 18F-FDG uptake in organs and tissues measured by dynamic whole body multiparametric FDG PET in 126 patients. *EJNMMI Res.* **12**(1), 1–14 (2022)
6. Dimitrakopoulou-Strauss, A., Pan, L., Sachpekidis, C.: Kinetic modeling and parametric imaging with dynamic PET for oncological applications: general considerations, current clinical applications, and future perspectives. *Eur. J. Nucl. Med. Mol. Imaging* **48**, 21–39 (2021). <https://doi.org/10.1007/s00259-020-04843-6>
7. Fahrni, G., Karakatsanis, N.A., Di Domenicantonio, G., Garibotto, V., Zaidi, H.: Does whole-body Patlak <sup>18</sup>F-FDG PET imaging improve lesion detectability in clinical oncology? *Eur. Radiol.* **29**, 4812–4821 (2019). <https://doi.org/10.1007/s00330-018-5966-1>
8. Guo, X., Zhou, B., Chen, X., Liu, C., Dvornek, N.C.: MCP-Net: inter-frame motion correction with Patlak regularization for whole-body dynamic PET. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) *MICCAI 2022, Part IV*. LNCS, vol. 13434, pp. 163–172. Springer, Cham (2022). [https://doi.org/10.1007/978-3-031-16440-8\\_16](https://doi.org/10.1007/978-3-031-16440-8_16)
9. Huang, Z., et al.: Parametric image generation with the uEXPLORER total-body PET/CT system through deep learning. *Eur. J. Nucl. Med. Mol. Imaging* **49**(8), 2482–2492 (2022). <https://doi.org/10.1007/s00259-022-05731-x>
10. Küstner, T., et al.: CINENet: deep learning-based 3D cardiac CINE MRI reconstruction with multi-coil complex-valued 4D spatio-temporal convolutions. *Sci. Rep.* **10**(1), 13710 (2020)
11. Li, A., Tang, J.: Direct parametric image reconstruction for dynamic myocardial perfusion PET using artificial neural network representation (2022)
12. Li, Y., et al.: A deep neural network for parametric image reconstruction on a large axial field-of-view PET. *Eur. J. Nucl. Med. Mol. Imaging* **50**(3), 701–714 (2023). <https://doi.org/10.1007/s00259-022-06003-4>
13. Moradi, H., Vegh, V., O'Brien, K., Hammond, A., Reutens, D.: FDG-PET kinetic model identifiability and selection using machine learning (2022)
14. Pantel, A.R., Viswanath, V., Muzi, M., Doot, R.K., Mankoff, D.A.: Principles of tracer kinetic analysis in oncology, part I: principles and overview of methodology. *J. Nucl. Med.* **63**(3), 342–352 (2022)
15. Pantel, A.R., Viswanath, V., Muzi, M., Doot, R.K., Mankoff, D.A.: Principles of tracer kinetic analysis in oncology, part II: examples and future directions. *J. Nucl. Med.* **63**(4), 514–521 (2022)

16. Sari, H., et al.: First results on kinetic modelling and parametric imaging of dynamic  $^{18}\text{F}$ -FDG datasets from a long axial FOV PET scanner in oncological patients. *Eur. J. Nucl. Med. Mol. Imaging* **49**, 1997–2009 (2022). <https://doi.org/10.1007/s00259-021-05623-6>
17. Snyman, J.A., Wilke, D.N., et al.: *Practical Mathematical Optimization*. Springer, New York (2005). <https://doi.org/10.1007/b105200>
18. Surti, S., Pantel, A.R., Karp, J.S.: Total body PET: why, how, what for? *IEEE Trans. Radiat. Plasma Med. Sci.* **4**(3), 283–292 (2020)
19. Wang, G., et al.: Total-body PET multiparametric imaging of cancer using a voxelwise strategy of compartmental modeling. *J. Nucl. Med.* **63**(8), 1274–1281 (2022)
20. Watabe, H.: Compartmental modeling in PET kinetics. In: Khalil, M.M. (ed.) *Basic Science of PET Imaging*, pp. 323–352. Springer, Cham (2017). [https://doi.org/10.1007/978-3-319-40070-9\\_14](https://doi.org/10.1007/978-3-319-40070-9_14)
21. Zuo, Y., Sarkar, S., Corwin, M.T., Olson, K., Badawi, R.D., Wang, G.: Structural and practical identifiability of dual-input kinetic modeling in dynamic PET of liver inflammation. *Phys. Med. Biol.* **64**(17), 175023 (2019)