Machine learning derived input-function in a dynamic ^{18}F-FDG PET study of mice

Samuel Kuttner[1,2,3*], Kristoffer Knutsen Wickstrøm[2], Gustav Kalda[1], S. Esmaeil Dorraji[4,5], Montserrat Martin-Armas[1,3], Ana Oteiza[1,3], Robert Jenssen[2], Kristin Fenton[4], Rune Sundset[1,3], Jan Axelsson[6]

[1]Nuclear Medicine and Radiation Biology Research Group, Department of Clinical Medicine, UiT The Arctic University of Norway, Tromsø, Norway.
[2]UiT Machine Learning Group, Department of Physics and Technology, UiT The Arctic University of Norway, Tromsø, Norway.
[3]The PET Imaging Center, University Hospital of North Norway, Tromsø, Norway.
[4]RNA and Molecular Pathology Research Group, Department of Medical Biology, UiT The Arctic University of Norway, Tromsø, Norway.
[5]Translational cancer immunotherapy research group, Department of Cancer Immunology, Norwegian Radium Hospital, Oslo University Hospital, Oslo, Norway.
[6]Department of Radiation Sciences, Umeå University, Umeå, Sweden.

[*]Corresponding and first author:
Samuel Kuttner, Nuclear Medicine and Radiation Biology Research Group, Department of Clinical Medicine, UiT The Arctic University of Norway, 9037 Tromsø, Norway, Phone: +47-77 66 99 53, e-mail: samuel.kuttner@uit.no

Abstract. Tracer kinetic modelling, based on dynamic ^{18}F-fluorodeoxyglucose (FDG) positron emission tomography (PET) is used to quantify glucose metabolism in humans and animals. Knowledge of the arterial input-function (AIF) is required for such measurements. Our aim was to explore two non-invasive machine learning-based models, for AIF prediction in a small-animal dynamic FDG PET study.

7 tissue regions were delineated in images from 68 FDG PET/computed tomography mouse scans. Two machine learning-based models were trained for AIF prediction, based on Gaussian processes (GP) and a long short-term memory (LSTM) recurrent neural network, respectively. Because blood data were unavailable, a reference AIF was formed by fitting an established AIF model to vena cava and left ventricle image data. The predicted and reference AIFs were compared by the area under curve (AUC) and root mean square error (RMSE). Net-influx rate constants, K_i, were calculated with a two-tissue compartment model, using both predicted and reference AIFs for three tissue regions in each mouse scan, and compared by means of error, ratio, correlation coefficient, P value and Bland-Altman analysis. The impact of different tissue regions on AIF prediction was evaluated by training a GP and an LSTM model on subsets of tissue regions, and calculating the RMSE between the reference and the predicted AIF curve.

Both models generated AIFs with AUCs similar to reference. The LSTM models resulted in lower AIF RMSE, compared to GP. K_i from both models agreed well with reference values, with no significant differences. Myocardium was highlighted as important for AIF prediction, but AIFs with similar RMSE were obtained also without myocardium in the input data.

Machine learning can be used for accurate and non-invasive prediction of an image-derived reference AIF in FDG studies of mice. We recommend the LSTM approach, as this model predicts AIFs with lower errors, compared to GP.

Keywords: PET, input-function, machine learning, compartment modelling
Submitted to: *Biomedical Physics & Engineering Express*

Introduction

Positron emission tomography (PET) is a widely used method for imaging in vivo biological processes in humans and animals. In particular, dynamic PET imaging of ^{18}F-fluorodeoxyglucose (FDG), combined with tracer kinetic modelling, can be used to quantify glucose metabolism (R. N. Gunn, S. R. Gunn, and Cunningham 2001). Compartment modelling requires accurate determination of an arterial input-function (AIF), i.e. the FDG time-activity curve in whole blood and plasma. The gold-standard AIF is obtained by measuring the time-dependent FDG radioactivity concentration in arterial blood through invasive blood sampling. In small-animal PET imaging of rodents, such a procedure is hampered by the limited blood volume that can be withdrawn without altering animal physiology, the complex surgery required for inserting an arterial catheter into the blood vessel, and the terminal end-point of the procedure. Several methods have been proposed to overcome these limitations, which we describe in the following:

A population-based AIF template, obtained from a large dataset acquired with the same tracer, injection protocol and population, can be calibrated to the specific subject (Takikawa et al. 1993). However, this method neglects individual physiological differences and scan-dependent variations, and requires at least one blood sample for curve scaling.

An image-derived input-function can be extracted from a large blood pool, visible in the images, such as the ascending or descending aorta, left ventricle (LV) or vena cava (VC) (Weerdt et al. 2001; Wu et al. 2007; Green et al. 1998; Lanz, Poitry-Yamate, and Gruetter 2014). This method is restricted by the limited spatial and temporal resolution of the PET imaging system, image noise, and cardiac and respiratory motion (Laforest et al. 2005). Specifically, the spatial resolution limitation introduces partial-volume effects, including signal spill-in and spill-over, which must be accounted for (Frouin et al. 2002; E. Kim et al. 2013; Y.-H. D. Fang and Muzic 2008).

Simultaneous estimation can be applied on image data to estimate both the AIF and kinetic parameters (Feng et al. 1997; Wong et al. 2001; Bartlett et al. 2018; Roccia et al. 2019), however, the method is complex, assumes a known mathematical AIF model and requires at least one late blood sample for parameter estimation.

Factor analysis can separate blood and myocardial signals from whole heart images (J. Kim et al. 2006), yet, the obtained factors may not necessarily represent truly corrected blood and tissue signals, and the method still requires one blood

sample for curve scaling.

In this study, we take a different approach to AIF estimation, based on machine learning (ML) (Theodoridis and Koutroumbas 2009). These methods are especially useful for function estimation and regression (Sapankevych and Sankar 2009), and have been actively used within medicine (Miles N. Wernick et al. 2014; Erickson et al. 2017). Briefly, one seeks to predict an output variable y, based on an input vector, $\mathbf{x}$, composed of one or multiple variables. An underlying functional relationship between the input and output is assumed, such that $y = f(\mathbf{x})$. This mapping is learned through available training data, for which both the input and output is known. Once the model has been trained, the potentially non-linear function, $f(\mathbf{x})$, can be applied on unseen samples to make predictions (Miles N. Wernick et al. 2014).

Although, ML techniques have not previously been applied for input-function estimation, attempts to use related statistical methods, such as multiple linear regression, and Bayesian models, have shown potential for AIF estimation in human brain (Y. H. Fang et al. 2004) and breast cancer studies (O'Sullivan et al. 2017). Gaussian process (GP) regression is a well-known statistical ML method for data driven function estimation (Roberts et al. 2013), and has been used to predict time series within health care (Dürichen et al. 2015). One advantage with GP is that it estimates not only the mean function, but also its variance, thus providing an uncertainty measure directly from the input training data (Rasmussen and Williams 2004). In contrast, neural networks, which have been applied within medicine for the past 25 years (Baxt 1995), build on learning mappings of high-dimensional input data, into a representation where linear regression can take place. Particularly, recurrent neural networks (RNN), were designed to handle time series data. However, while RNN models struggle to learn long-term dependencies, so called long short-term memory (LSTM) networks were introduced to efficiently incorporate long-term time-dependent information (Hochreiter and Schmidhuber 1997). LSTMs, have had successful applications within medicine, for prediction of electrocardiograms (Chauhan and Vig 2015), and blood glucose levels (Sun et al. 2018).

In this work we compare a machine learning-derived input-function (MLDIF) with an image-derived AIF estimated from vena cava and left ventricle. Our hypothesis is that this AIF can be accurately predicted by an MLDIF model using multiple tissue time-activity curves, not necessarily including the myocardium wall, as input.

Methods

The PET/CT images, volume delineations and time-activity curves used in this work were collected in retrospect from a completed study at our institution, focusing on PET imaging of Tertiary Lymphoid Structures (TLS) in two different mouse strains (Dorraji et al. 2016). Relevant details from the TLS study are given in the following.

Animals

All animal studies were approved by the Competent Authority on Animal Research, the Norwegian Food Safety Authority; FOTS id 6676/2015. 36 female mice from two strains (NZBWF1, Jax stock #10008 (n=24) and BALB/ cAnNCrl (n=12)), purchased from The Jackson Laboratory and Charles River Laboratories, respectively, were included in the TLS study (Dorraji et al. 2016). To minimize the effect of dietary state and anaesthesia on the FDG uptake in the mice (Spangler-Bickell et al. 2016; Fueger et al. 2006), the following strict fasting and anaesthesia protocol was followed prior to PET imaging: The mice were fasted for 3 h 50 min ± 20 min, weighed and anesthetized for 1 h 17 min ± 19 min prior to FDG injection, in an oxygen-isoflurane mixture (4% and 2% isoflurane for induction and maintenance, respectively). Blood glucose was measured in venous blood to 6.9 mmol/l ± 1.6 mmol/l at 56 min ± 20 min prior to tracer administration, using a glucose meter (FreeStyle Lite, Abott Laboratories). A catheter, made from polyethylene tubing and a 30 gauge needle, was placed into the lateral tail-vein to allow FDG injection.

PET/CT imaging

PET/computed tomography (CT) imaging of totally 68 mouse scans was performed using a Triumph^{TM} LabPET-8^{TM} small-animal PET/CT scanner (TriFoil Imaging Inc.). Each mouse was scanned between 1-5 times at different ages (range 7-37 weeks), weighing 33 ± 8 g at imaging time. 20 mice were scanned one time, 6 mice were scanned two times, 6 mice were scanned three times, two mice were scanned four times and two mice were scanned five times. The anesthetized mice were centered in the field-of-view of the PET/CT scanner, lying on a 35°C heated bed inside an animal imaging cell (Equipment Veterinaire Minerve), with sensors monitoring heart and breathing rate. 10.5 ± 1.8 MBq of FDG (MAP Medical Technologies) in 100 μl sterile saline was injected through the tail-vein

catheter during 30s, with an infusion pump (56 scans), or by manual injection followed by 20 μl flush of sterile saline (12 scans). A 60 minute list-mode PET acquisition was started at injection time.

Immediately following PET imaging, a CT scan was performed for PET attenuation correction. The following settings were used: 80 kVp, 2x2 binning, 512 projections and 1.3x magnification.

Image reconstruction

The list-mode PET data were binned into 44 time steps (24×5s, 9×20s and 11×300s) and reconstructed to 0.5 × 0.5 × 0.6 mm^3 voxel size, using a 3-dimensional maximum-likelihood estimator algorithm with 50 iterations. Corrections for detector efficiency, radioactive decay, random coincidences, dead time, attenuation and scatter were applied. The voxels were normalized into standardized uptake value (SUV) [g/ml] (Keyes 1995).

The CT data were reconstructed using filtered back projection, to images with 0.177 mm isotropic voxel size.

Image analysis

Volumes of interests (VOI) were delineated in either CT, dynamic PET or static PET space, the latter which was formed by averaging the last 20 minutes of the dynamic PET acquisition. The image modality in which each VOI could be defined in a standardized and reproducible way was chosen (Table 1). From among the tissue regions delineated in the TLS study (Dorraji et al. 2016), the following 7 were selected, hypothesized to be relevant for this study: vena cava, left ventricle, myocardium, brain, liver, muscle and brown fat (Figure 1). These regions were systematically delineated using the same method for all mouse scans by either of two experienced imaging researchers. Researcher 1 and 2 delineated 52 and 16 mouse scans, respectively. Subsequently, all delineations were quality assured by Researcher 1. The CT VOIs were down-sampled to the resolution of the PET images, and co-registered with these using rigid transformation. All VOIs were applied to the dynamic PET images, and the mean time-activity curve was extracted from each VOI.

Table 1. Delineation details of the VOIs. (a) Tissue regions used for reference AIF estimation. (b) Tissue regions used for training and testing the MLDIF models.

	Tissue region	Volume [mm^3]	Image space	Delineation method
(a) Regions used for reference AIF estimation				
1	Vena cava	0.9 ± 0.1	dPET	0.6 mm radius sphere centered on peak voxel in early time step
2	Left ventricle	9.7 ± 3.1	sPET	Region inside myocardium uptake
(b) Regions used for training and testing the regression models				
1	Myocardium	114.5 ± 36.4	sPET	Voxels > 40-60%* of max voxel value above background† in whole heart
2	Brain	33.5 ± 0.2	CT	2 mm radius sphere in the dorsal region of the skull
3	Liver	32.7 ± 4.0	CT	2 mm radius sphere in upper right liver lobe
4	Muscle	33.5 ± 0.2	CT	2 mm radius sphere in skeletal muscle of right front leg
5	Brown fat	96.2 ± 42.9	sPET	Voxels > 40-60%‡ of max voxel value above background† around BF

dPET/sPET = Dynamic/static positron emission tomography; VC = Vena cava; VOI = Volume of interest; CT = Computed tomography; BF = Brown fat.
*The threshold was determined with manual interaction for each mouse scan, such that the myocardium VOI formed an approximate torus.
†The maximum voxel value was subtracted from minimum before thresholding, i.e. I_{max} - I_{min}, where I is the voxel value.
‡The threshold was determined with manual interaction for each mouse scan, such that the BF VOI was confined to a single FDG active region.

FDG compartment model

An irreversible two-tissue compartment model (2TCM) was used to calculate the rate constants K_1, k_2 and k_3, while k_4=0 for FDG (R. N. Gunn, S. R. Gunn, and Cunningham 2001). This model assumes FDG to be either free, or phosphorylated (FDG-6P) and trapped in tissue, with activity concentrations C_1 and C_2, respectively. The two state equations are:

$$\frac{dC_1(t)}{dt} = K_1 \cdot C_p(t) - (k_2 + k_3) \cdot C_1(t) \quad (1)$$

$$\frac{dC_2(t)}{dt} = k_3 \cdot C_1(t) \quad (2)$$

where $C_p(t)$ is the arterial plasma time-activity curve, also known as the AIF. Although it has been shown that the ratio of FDG concentration in whole blood, $C_a(t)$, and plasma, $C_p(t)$, varies over time (Wu et al. 2007; Weber et al. 2002), such a correction would require blood sampling, and was therefore not possible in this study. Consequently, equality between $C_a(t)$ and $C_p(t)$ was assumed.

Ignoring the blood fraction in tissue, the tissue concentration measured with PET, $C_t(t)$, is the sum of the two states, $C_t(t) = C_1(t)+C_2(t)$. By using Equations 1-2, $C_t(t)$ can be expressed as (Phelps, Hoffman, and Kuhl 1980):

$$C_t(t) = \frac{K_1}{k_2 + k_3}\left[k_3 + k_2 \cdot e^{-(k_2+k_3)\cdot t}\right] \otimes C_p(t) \quad (3)$$

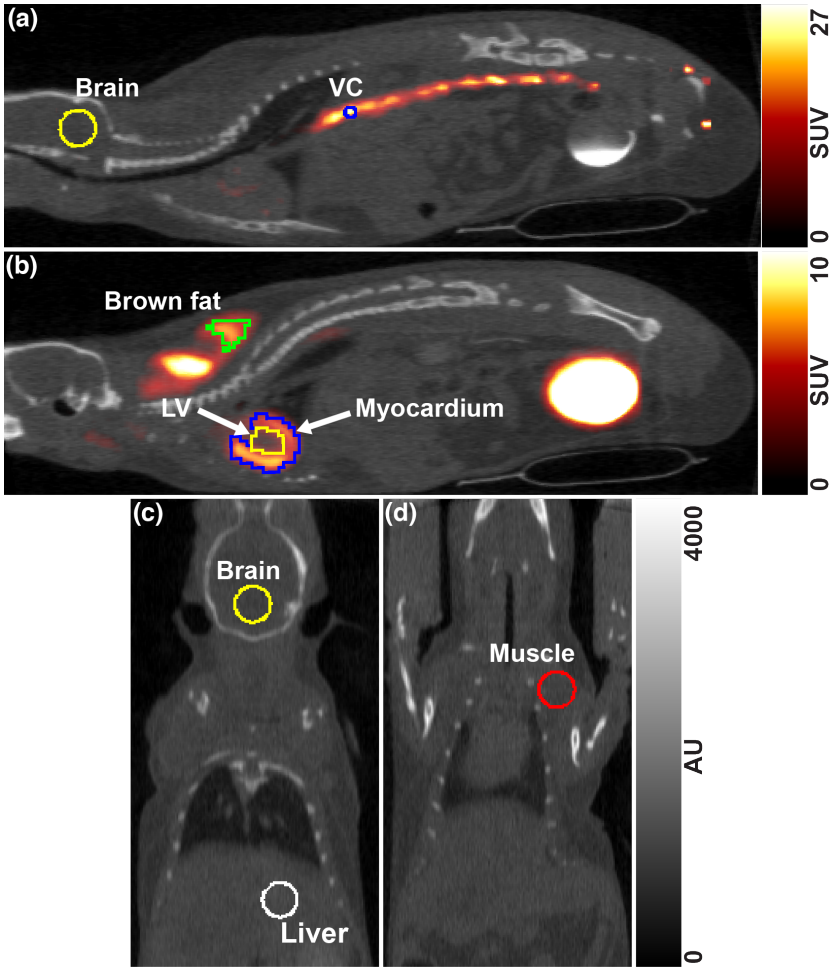

Figure 1. VOI delineations in one representative mouse. (a) Brain (*yellow*) and VC (*blue*). (b) LV (*yellow*), myocardium (*blue*) and brown fat (*green*). (c) Brain (*yellow*) and liver (*white*). (d) Skeletal muscle (*red*). (a)-(b) in PET/CT space, (c)-(d) in CT space. AU = Arbitrary units; LV = Left ventricle; SUV = Standardized uptake value; VC = Vena cava

where $\otimes$ denotes mathematical convolution. The net-influx rate constant, K_i, is defined as (Y.-H. D. Fang and Muzic 2008):

$$K_i \equiv \frac{K_1 \cdot k_3}{k_2 + k_3} \tag{4}$$

Reference input-function estimation

In this study, no blood samples were collected. Instead, for each mouse scan, an AIF was formed from vena cava (VC) and left ventricle (LV) VOI data. After tail-vein FDG injection, the tracer flows through VC before reaching the heart. Therefore, the initial VC peak consists mostly of FDG prior to mixing with blood, thus overestimating the true AIF peak in early time steps (Lanz, Poitry-Yamate, and Gruetter 2014). Furthermore, the large ($\sim$10 mm^3) LV VOI is less affected by spill-over effects than the small (0.9 mm^3) VC VOI. Therefore, LV yields a more correct representation of the AIF in early time steps compared to VC. However, LV is significantly affected by spill-in from myocardium (Y.-H. D. Fang and Muzic 2008), hence in later time steps, the VC curve is more representative of the AIF. This knowledge was implemented by forming a measured, image-derived AIF, $C_p^{VC,LV}(t)$, for each time step, t, by:

$$C_{p,t}^{VC,LV} = min(C_{VC,t}, C_{LV,t}) \qquad t = 1, 2, ..., 44 \tag{5}$$

where $C_{VC,t}$ and $C_{LV,t}$ are the mean SUVs in each time step, t, in the VC and LV VOIs, respectively (Vesa Oikonen, personal communication, June 12, 2018).

To reduce noise among the discrete AIF data points, a well-known parametric model was used to describe the AIF (Dagan Feng, Huang, and Wang 1993):

$$C_p(t) = \begin{cases} 0 & \text{if } t < \tau \\ (A_1(t-\tau) - A_2 - A_3)e^{L_1(t-\tau)} + A_2 e^{L_2(t-\tau)} + A_3 e^{L_3(t-\tau)} & \text{otherwise} \end{cases} \tag{6}$$

where A_1 through A_3 and L_1 through L_3 are model constants, and τ is a timing delay constant. Although this model has limitations, such as assuming bolus tracer injections, recently improved models have not shown significantly improved AIF fits for FDG (Tonietto et al. 2015). Therefore, the parametrized model of the input-function (Equation 6) was fitted to the image-derived data points, $C_p^{VC,LV}(t)$, and used as reference AIF for each mouse scan. Linear interpolation to 1 second uniform time steps was performed for the AIF fit, before the obtained reference AIFs were interpolated back to the original, non-uniform time steps of the dynamic PET data.

Gaussian processes

GP can be used to solve non-linear regression tasks, where the output, y_n, is approximated by a probability distribution over functions of the input, $\mathbf{x}_n$, such that $f(\mathbf{x}_n) \sim \mathcal{GP}(m(\mathbf{x}_n), k_{\boldsymbol{\theta}}(\mathbf{x}_n, \mathbf{x'}_m))$. Here, $m(\mathbf{x}_n)$ is a mean function, $k_{\boldsymbol{\theta}}(\mathbf{x}_n, \mathbf{x'}_m)$ is a covariance function, parameterized by $\boldsymbol{\theta}$, and σ^2 specifies the noise power (Rasmussen and Williams 2004). Having N available input-output training samples in a set $D = \{\mathbf{x}_n, y_n\}_{n=1}^{N}$, each including the time-activity curves of the tissues from Table 1(b), with corresponding known reference AIF, the mean value AIF of the test sample, $\mathbb{E}[y_*]$, and the variance, $\mathbb{V}[y_*]$, can be calculated by:

$$\mathbb{E}[y_*] = \mathbf{k}_*^\top (\mathbf{K} + \sigma_n^2 \mathbf{I})^{-1} \mathbf{y} \tag{7}$$

$$\mathbb{V}[y_*] = k(\mathbf{x}_*, \mathbf{x}_*) - \mathbf{k}_*^\top (\mathbf{K} + \sigma_n^2 \mathbf{I})^{-1} \mathbf{k}_*. \tag{8}$$

Here $\mathbf{k}_*$ is the covariance between the training samples $\mathbf{x}$ and the test sample $\mathbf{x}_*$; $[\mathbf{K}]_{ij} = k_{\boldsymbol{\theta}}(\mathbf{x}_i, \mathbf{x}_j)$ is the covariance between all training samples; $\sigma_n^2 \mathbf{I}$ is a scalar matrix with diagonal elements equal to the noise level; $k(\mathbf{x}_*, \mathbf{x}_*)$ is the covariance between the test sample and itself (Rasmussen and Williams 2004).

Long short-term memory network

RNNs are designed to process sequential data and learn time-dependencies (Lipton, Berkowitz, and Elkan 2015). They take time series as input, processes it element-wise, and outputs a vector, named the hidden state, that contains information from previous time steps. For each time step, t, the prediction, y_t is modelled as $y_t = f(\mathbf{x}_t, \mathbf{h}_{t-1})$, where $\mathbf{x}_t$ is the current time step input, $\mathbf{h}_{t-1}$ is the previous time step hidden state, and f is parametrized by a neural network. Unfortunately, as a result of vanishing or exploding gradients during training, RNNs have difficulties learning long-term dependencies (Hochreiter and Frasconi 2009). To overcome this, a modified architecture was introduced, named LSTM network (Hochreiter and Schmidhuber 1997), that could incorporate long-term dependencies into a cell state, that passes information forward from previous time steps. Three serial gates, an input, a forget and an output gate, modifies the information that will be added to, removed from, or carried on by, the cell state, at each time step (Hochreiter and Schmidhuber 1997).

Input-function prediction

For AIF prediction, the data set ($N = 68$) was randomly shuffled, and divided into a training set ($N_{tr} = 56$) and test set ($N_{te} = 12$). The training set was used to calculate the parameters, while the test set was used to evaluate the performance of the MLDIF models. Subsequent model training was repeated 1000 times for both GP and LSTM, respectively, with a new shuffle and split at each repeat. The same 1000 shuffle and splits were used for both GP and LSTM experiments. This resulted in a varying number of predicted AIFs for each mouse scan ($N_{min} = 151$, $N_{max} = 206$), depending on the frequency with which it occurred in the test set in the 1000 experiments. Because the tissue regions in Table 1(a) were used for reference AIF estimation, only regions from Table 1(b) were included for training and testing the MLDIF models.

For GP, an AIF prediction, $\mathbb{E}[y_*]$ was calculated for each mouse scan in the test set, y_*, using Equation 7. With the 44 time step tissue time-activity curves as input vectors, the corresponding output was a 44 time step AIF curve. The Matérn covariance function was chosen, with $\nu = 5/2$, because this choice produces smooth function samples, as discussed in (Rasmussen and Williams 2004). To obtain an equal number of AIFs for each mouse scan, $N_{min} = 151$ predicted AIF models were randomly selected for each mouse scan. The average and standard deviation (SD) over these 151 AIFs was then calculated to represent the predicted AIF and its variation, for each mouse scan.

For LSTM, the model training was performed by fitting the weights of the network through a series of iterations (epochs). For this model, validation data was required to determine when to stop iterating to avoid over-fitting. Therefore, a validation set, N_{vl}, was formed by randomly selecting 12 mouse scans from the training set, which were not used for weight fitting. The hyperparameters of the LSTM models were empirically set to: 20 neurons in the hidden state; maximum 1000 epochs training but using early stopping with minimum delta 0.0001 and 50 epochs patience while monitoring the validation set loss; 0.001 learning rate; a mini-batch size of 12. Training was performed using the ADAM optimizer (Kingma and Ba 2014) and the mean squared error loss function. For LSTM, each of the 151 - 206 predicted AIFs, for each mouse scan, was associated with a validation loss, calculated as the sum of the mean squared errors of all samples in the validation set after LSTM training. For each mouse scan, the predicted AIF associated with the lowest validation data set loss was chosen to represent the AIF for that mouse scan. The average of $N_{min} = 151$ randomly selected AIFs for each mouse scan

and time step, including the selected AIF, as well as the corresponding SD was calculated for each mouse scan.

Input-function validation

The predicted AIFs, $\hat{\mathbf{y}}$, were compared with the reference AIF, $C_p(t)$ from Equation 6, for each mouse scan, by the area under curve (AUC) and root mean square error (RMSE):

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^{T=44}(\hat{y}_t - C_{p,t})^2}{T}} \tag{9}$$

An irreversible 2TCM (Equation 3) was used to estimate the rate constants K_1, k_2 and k_3, using the reference AIF, and the predicted AIF from GP and LSTM, respectively. Calculations were performed for brain, skeletal muscle and myocardium, which were the three tissues from Table 1 expected to follow this kinetic model. Subsequently, K_i was calculated for these three tissues using Equation 4. The error in K_i was calculated as:

$$\left(\frac{K_i^{Model}}{K_i^{Ref}} - 1\right) \times 100\% \tag{10}$$

where K_i^{Model} and K_i^{Ref} represents K_i, obtained from the predicted AIF and the reference AIF, respectively. The percent errors over mouse scans were summarized using mean and SD. Furthermore, the correlation coefficients between K_i^{Model} and K_i^{Ref} were calculated. Also, after checking for normality, a paired t test with $\alpha = 0.05$ was used assess statistical significance in K_i for each tissue region and MLDIF model. Morover, Bland-Altman plots were generated to further investigate the agreement in K_i between model-derived and reference values (Martin Bland and Altman 1986). In these diagrams, both the mean difference and the $\pm$ 2 SD interval were used for evaluation.

One mouse scan was removed from model comparisons due to failed reference AIF fit attributed to noisy input data. Two additional mouse scans for each MLDIF model were defined as outliers and also excluded from model comparisons, because their AIF RMSE was more than three scaled median absolute deviations away from the median RMSE (Hubert and Van der Veeken 2008). Furthermore, compartment modelling resulted in abnormal rate constants for four mouse scans for either heart or muscle tissue regions, and for two additional mouse scans, the brain time-activity curves were abnormally noisy due to failed normalization for

peripheral detectors. Therefore, these mouse scans were also excluded from model comparisons, for the affected tissues.

Tissue region importance

To investigate the importance of each tissue on AIF prediction, 11 different data sets were formed, using the following permutations of tissues from Table 1(b) for MLDIF model training: all, all except myocardium, all except brain, all except liver, all except muscle, all except brown fat, myocardium, brain, liver, muscle and brown fat. Briefly, the data set was shuffled and split into training and test sets, as described earlier. Subsequently, one GP and one LSTM model was trained on each of these 11 tissue permutations, and then used to obtain a predicted AIF for each of the 12 mouse scans in the test set of the current shuffle. The experiment was repeated 100 times, with a new shuffle and split at each repeat. The same 100 shuffle and splits were used for both GP and LSTM experiments. The mean RMSE over the mouse scans in the test set was used to evaluate the predictive performance of each tissue permutation.

Software and computational environments

The AIF regression models were implemented in Python 3.6.3, using GPflow 1.2.0 for the GP models (Matthews et al. 2017), and Keras 2.1.5 API for the LSTM models (Chollet 2015). The source code for these models is available at https://github.com/Kuttner/MLDIF.

Reference AIF estimation and compartment modelling was performed in Matlab R2018a (Mathworks). A constrained nonlinear multivariable optimizer (*fmincon*), minimizing the weighted sum-of-squared errors, was used for the AIF model fit and a nonlinear least-squares solver (*lsqcurvefit*) was used for compartment modelling.

The VOIs in Table 1 were delineated using PMOD 3.8 (PMOD Technologies Ltd.).

Results

Reference input-function estimation

The parameterized reference AIF curve (Equation 6) and the underlying VC and LV curves (Table 1(a), Equation 5) are shown in Figure 2(a), for one representative

mouse scan. The parameterized curve is noiseless and smooth, compared to VC and LV. The time-activity curves for the 5 tissue regions from Table 1(b) are displayed in Figure 2(b).

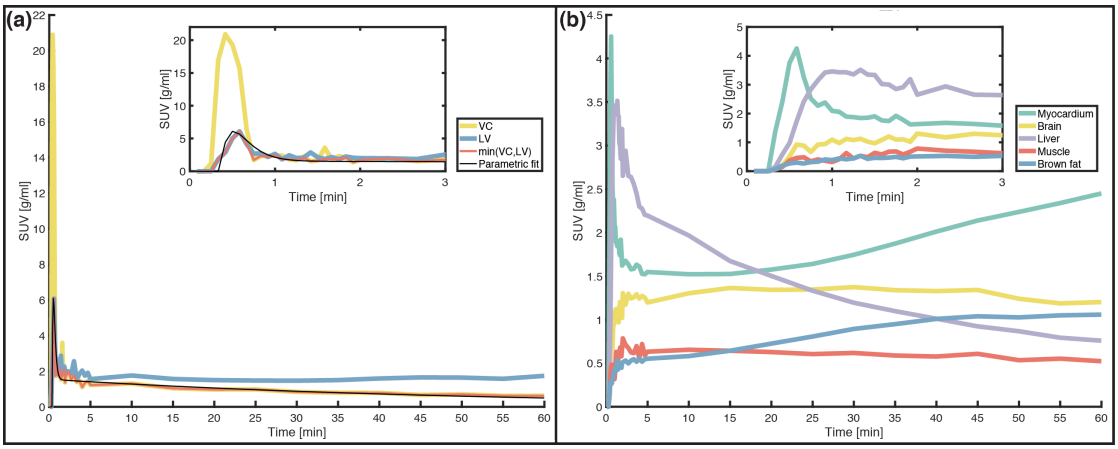

Figure 2. Time-activity curves from the VOIs delineated in one representative mouse scan. (a) The VC and LV tissue regions from Table 1(a), the minimum of these, and the parameterized reference AIF, using Equations 5-6. (b) The 5 tissue regions from Table 1(b). In (a) and (b), the inserts depict the first three minutes of the time-activity curves.

Input-function validation

Results from comparisons between the predicted and reference AIFs in terms of AUC and RMSE are shown in Table 2. Both models generated AIFs with AUCs similar to reference, with mean AUC errors $< 5\%$.

The corresponding AUC values for the two mice scanned 5 times were 80.4 $\pm$ 19.7 and 78.1 $\pm$ 14.2. The within-subject AUC was thus in the same range as the AUC calculated over all subjects. This suggests similar inter- as intra-subject variability among the AIFs. Consequently, mice that were scanned multiple times were treated as independent samples. The predicted AIFs for the three mouse scans with lowest, 50th percentile and 75th percentile RMSE, respectively, are shown in Figure 3, for the GP and LSTM model. Additionally, the RMSE histogram for each model is shown. For both the GP and LSTM models, the regression curves with the lowest RMSE ($\text{RMSE}_{\text{GP}} = 0.23$ g/ml, $\text{RMSE}_{\text{LSTM}} = 0.19$ g/ml) agrees well with the reference AIF (Figure 3, first column). The LSTM model fit generally resembles the reference AIF better and with lower variations,

compared to GP, also for the 50th percentile ($\text{RMSE}_{\text{GP}} = 0.58$ g/ml, $\text{RMSE}_{\text{LSTM}} = 0.44$ g/ml) and 75th percentile ($\text{RMSE}_{\text{GP}} = 0.84$ g/ml, $\text{RMSE}_{\text{LSTM}} = 0.54$ g/ml) scan. Furthermore, the RMSE histogram and Table 2 display lower mean RMSE for the LSTM model (0.44 ± 0.16 g/ml), compared to GP (0.65 ± 0.29 g/ml).

Table 2. Comparison of the predicted and reference AIFs. The RMSE [g/ml] was calculated with Equation 9, while the AUC error [%] was calculated as $(\text{AUC}_{\text{Model}}/\text{AUC}_{\text{Ref}} - 1) \times 100\%$.

AIF	AUC [g·min/ml]	AUC error [%]	RMSE [g/ml]
Reference	80.6 ± 15.5		
GP	79.0 ± 11.8	1.3 ± 22.6	0.652 ± 0.287
LSTM	83.2 ± 15.1	4.9 ± 14.4	0.444 ± 0.156

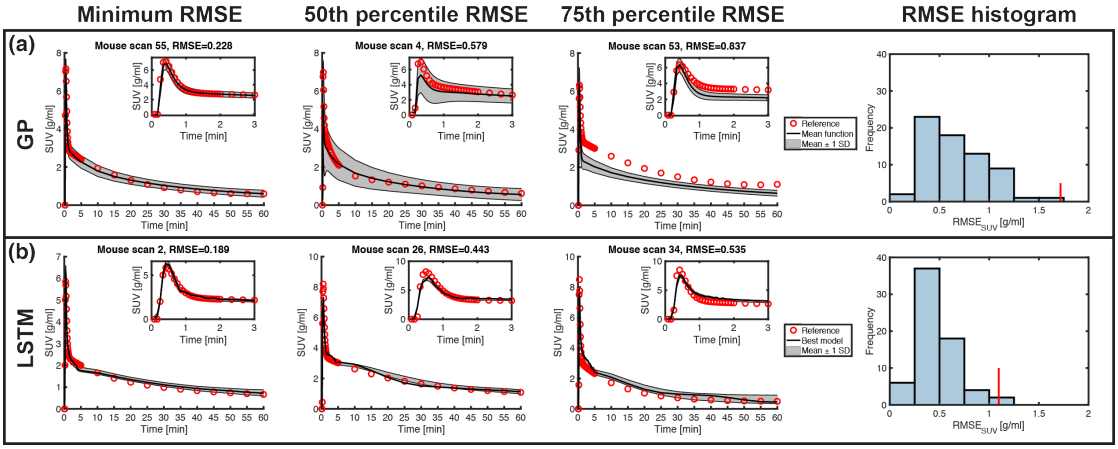

Figure 3. AIFs generated by the two evaluated MLDIF models. (a) GP. (b) LSTM. For each model, the mouse scan with lowest, 50th and 75th percentile RMSE is shown, respectively, as well as the RMSE histogram over all mouse scans. In the histograms, the mean ± SD RMSE for the LSTM and GP model is 0.44 ± 0.16 g/ml and 0.65 ± 0.29 g/ml, respectively. Red circles indicate the reference AIF. Black line represents the mean over $N_{min} = 151$ models in the test set for GP in (a) and the single LSTM model in the test set with lowest validation loss in (c). In both (a) and (b), the shaded area shows the mean AIF ± 1 SD across 151 models. The vertical red lines in the histograms indicate the three scaled median absolute deviation threshold for outlier removal (Hubert and Van der Veeken 2008).

As the aim of estimating the AIF is for its use in tracer kinetic modelling, it is important to evaluate the error induced in K_i. Table 3 shows the K_i values

obtained from the reference AIF and the two model-derived AIFs, GP and LSTM, for brain, muscle and myocardium tissue regions. Furhtermore, Figure 4 presents the ratio distribution of K_i obtained with the two MLDIF models, to K_i obtained with Reference AIF for the same three tissues. Both models yielded rate constants very similar to reference, with average errors over the three tissues of 5.5% ± 33.2% for the GP model and -0.7% ± 35.4% for the LSTM model and with correlation coefficients of 0.95 and 0.94, respectively. As shown in Figure 4, the LSTM model resulted in slightly more underestimated K_i values when compared to reference, with median K_i ratio over the three tissues of 0.934, compared to GP, with a corresponding median ratio of 0.999. The paired t test did not detect significant differences in K_i for either of the tissue regions, with $P > 0.05$ for both GP and LSTM models, when comparing to reference (Table 3).

Table 3. Comparison of K_i calculated from Reference AIF and the respective MLDIF model.

Model	Statistics of K_i^*	Brain	Muscle	Myocardium
Reference	Estimate (g/min/ml)	0.0146 ± 0.0044	0.0057 ± 0.0036	0.1041 ± 0.0696
GP	Estimate (g/min/ml)	0.0145 ± 0.0048	0.0058 ± 0.0039	0.1049 ± 0.0720
	Error (%)	6.7 ± 37.8	5.5 ± 33.4	4.2 ± 28.1
	Corr. coefficient	0.56	0.91	0.90
	t test P value	0.911	0.588	0.647
LSTM	Estimate (g/min/ml)	0.0138 ± 0.0047	0.0054 ± 0.0037	0.1023 ± 0.0709
	Error (%)	-1.4 ± 34.8	-1.8 ± 30.6	1.0 ± 40.6
	Corr. coefficient	0.68	0.91	0.86
	t test P value	0.058	0.148	0.466

*Estimate (g/min/ml) and error (%) are expressed as mean ± SD. Correlation coefficient (Corr. coefficient) and P values are calculated from (K_i^{MLDIF}, $K_i^{Ref.}$) pairs of the included mouse scans.

Figure 5 shows Bland-Altman plots of the model-derived and reference K_i values, for brain, muscle and myocardium tissue regions. Generally, the mean difference was close to zero for both MLDIF models for the three tissue regions (GP, mean difference = 0.0007. LSTM, mean difference = -0.0015), indicating that K_i from the predicted AIFs agree well with reference for the three tissues. Also, the 2 SD interval was similar in both models for brain (GP, 2 SD = 0.008. LSTM 2 SD = 0.007), muscle (GP, 2 SD = 0.003. LSTM 2 SD = 0.003) and

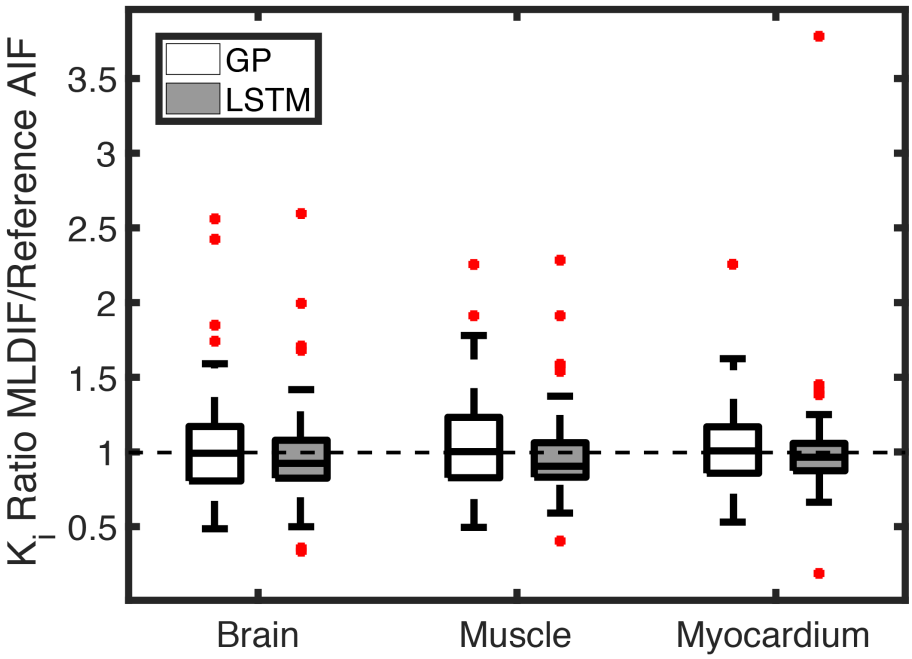

Figure 4. Box plot of ratios of K_i obtained with GP and LSTM MLDIF to K_i obtained with Reference AIF for brain, muscle and myocardium tissues. As a result of outlier removal, as described in the main text, the number of included mouse scans varied between 62-65, for the shown models and tissues. The horizontal line and the black box represents median and interquartile range (25th to 75th percentile), respectively, while the whiskers indicate the maximum and minimum data point up to 1.5 × interquartile range. Data points outside this interval are shown with red dots. The dashed black line indicates unity ratio. There was no significance difference between the two MLDIF models and the Reference AIF for the three tissues (Table 3).

myocardium (GP, 2 SD = 0.063. LSTM 2 SD = 0.073).

Tissue region importance

Training a GP and an LSTM model with each of the 11 tissue permutations resulted in 11 AIFs for each test mouse scan and model, each with an associated RMSE. Figure 6 shows the distribution of the mean RMSE over the 12 test mouse scans for the 11 tissue permutations, averaged over all 100 GP and LSTM experiments. The lowest RMSE was obtained when training an LSTM

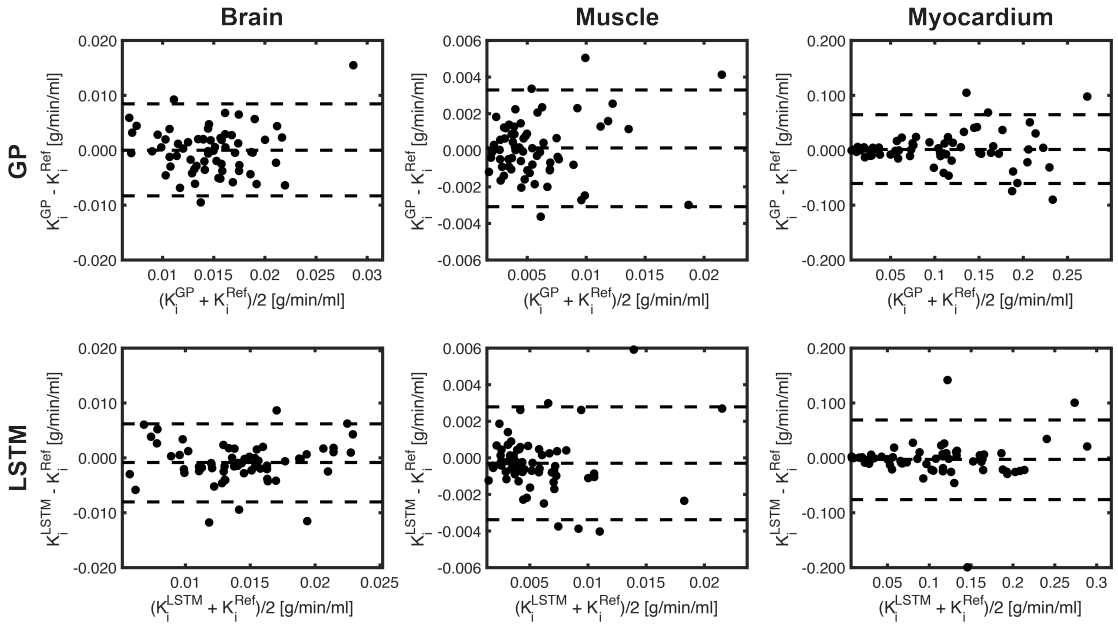

Figure 5. Bland-Altman plots of predicted and reference K_i using GP and LSTM MLDIF models for brain, muscle and myocardium tissue regions. Central and outer dashed lines indicate mean value and mean ± 1.96 SD (simply referred to as 2 SD throughout this paper), respectively (Martin Bland and Altman 1986). GP, Brain: mean difference = 0.0001, 2 SD = 0.0084. GP, Muscle: mean difference = 0.0001, 2 SD = 0.0032. GP, Myocardium: mean difference = 0.0019, 2 SD = 0.0631. LSTM, Brain: mean difference = -0.0009, 2 SD = 0.0071. LSTM, Muscle: mean difference = -0.0003, 2 SD = -0.0031. LSTM, Myocardium: mean difference = 0.0034, 2 SD = 0.0728.

model with all except brain tissue regions (median RMSE = 0.47 g/ml, max-min = 0.48 g/ml), indicating that brain was least important for AIF prediction, although this error was similar to when all regions were included for training (P = 0.06, median RMSE = 0.48 g/ml, max-min = 0.33 g/ml). Furthermore, a similar error with only slightly higher variability was obtained when including only myocardium (P = 0.16, median RMSE = 0.50 g/ml, max-min = 0.44 g/ml), suggesting that myocardium is important for AIF prediction. Training on all regions except myocardium, or on all regions except liver resulted in significantly larger errors ($P < 0.05$, median RMSE = 0.65 g/ml and median median RMSE = 0.55 g/ml, respectively), compared to when all regions were included. Furthermore, for LSTM, single-tissue permutations resulted in larger RMSE (overall mean RMSE = 0.70 g/ml, SD = 0.14 g/ml), compared to multi-tissue permutations (overall mean RMSE = 0.53 g/ml, SD = 0.10 g/ml). All

single-tissue errors, except myocardium, were significantly different from when all regions were used for training ($P < 0.05$). For GP, the lowest RMSE was obtained when training the model on myocardium exclusively (median RMSE = 0.66 g/ml, max-min = 0.62 g/ml), while all other investigated tissue permutations resulted in significantly larger errors ($P < 0.03$, 0.66 < median RMSE < 0.87 g/ml). All LSTM tissue permutation errors (overall mean RMSE = 0.61 g/ml, SD = 0.15 g/ml) were significantly smaller ($P < 0.05$) compared to GP (overall mean RMSE = 0.81 g/ml, SD = 0.14 g/ml).

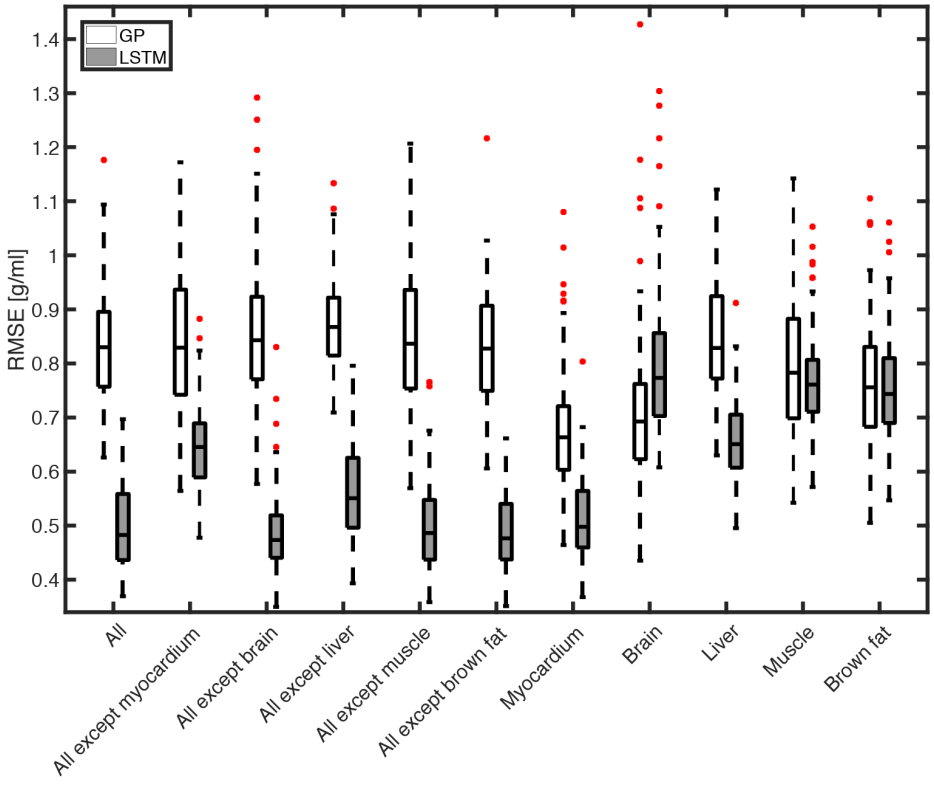

Figure 6. Box plot of the mean RMSE distribution between MLDIF-derived and reference AIFs, over the 12 test mouse scans for all 100 GP and LSTM experiments. 7 specific tissue permutations are shown, where MLDIF model training was based solely on the indicated tissues. See Figure 4 legend for description of what the box height, center line, and whiskers indicate.

Discussion

Tracer kinetic modelling from dynamic PET imaging requires accurate knowledge of the AIF, ideally determined through arterial blood sampling. In small-animal imaging, an image-derived AIF approximation is often preferred because of limited blood volume, and to avoid terminal experiments and complex surgery. Our aim was to find a non-invasive, image-derived method, for determining the AIF, without the need for surgery, and with an inherent potential to be insensitive to partial-volume effects. In this study, we proposed two machine learning-derived AIF models (MLDIFs) that, when properly trained, approximates the real AIF: a statistical method based on GP, and a deep learning-based approach based on an LSTM network. We compared the predicted AIFs with image-derived reference AIFs, because blood input data was not available.

Our results showed that both investigated MLDIF models were well-suited for this task, predicting AIFs with similar AUC compared to reference and with low average errors (Table 2). The magnitude of the errors were comparable to earlier studies (Y.-H. D. Fang and Muzic 2008). The use of AUC alone to quantify agreement between curves may, however, be misleading, because two AIFs with vastly different curve shape can have similar AUC. Therefore, we applied the RMSE, which provides a better measure of the agreement between two AIFs. Evidently, the LSTM model predicted AIFs with lower RMSE and less variation, compared to GP (Table 2, Figure 3). Since the AIF curve itself is not the interesting result in most dynamic PET studies, we evaluated the tracer kinetic output, K_i, calculated from a 2TCM with the reference AIF as input, and compared it to the corresponding K_i, when using the model-derived AIFs as input.

Compartment modelling showed that both MLDIFs resulted in similar population averaged rate constants compared to reference, with the error being lower for the LSTM model, compared to GP (Table 3, Figure 4 and Figure 5). Both the absolute values of K_i and the errors agreed well with previously published results (Y.-H. D. Fang and Muzic 2008). Correlation between model-derived and reference K_i values was strong and positive for muscle and myocardium (correlation coefficient > 0.9) for both MLDIF models, while for brain, it was somewhat lower (correlation coefficient > 0.6) (Table 3). This may be explained by the brain region being located close to the end slices of the scanner, where noise is high, and thus suggests that the MLDIF methods are sensitive to noisy input data. All P values were above the significance level of 0.05, indicating that significant differences between model-derived and reference K_i could not be detected for any

of the tissues or MLDIF models (Table 3).

The Bland-Altman analysis (Figure 5) revealed mean differences close to zero for both MLDIF models and all three tissues. Furthermore, the 2 SD intervals were very similar for GP and LSTM within each tissue, thus neither model outperformed the other in terms of K_i accuracy.

The time-consuming manual delineation of all 5 tissue regions from Table 1(b) can be minimized if only one, or few of the regions can be used for AIF prediction. Furthermore, dynamic PET acquisitions are usually restricted to a single bed position. For larger rodents, such as rats, or for human PET imaging, this implies that only a few of the regions from Table 1(b) is visible in the dynamic images. Figure 6 indicated that for the LSTM model, an AIF with similar RMSE as the AIF derived with all tissues used for training, could be predicted solely based on myocardium data. This region inevitably contains spill-in from the blood pool, thus inherently including a strong component that reflects the AIF. The importance of the myocardium for the LSTM model was also shown as an increased RMSE in the "all except myocard" permutation, compared to all other multi-tissue permutations. A similar effect was observed for the liver tissue region, which similarly to myocardium, has a high blood content. Interestingly, while myocardium was the best performing tissue for GP, training on all tissues resulted in the largest RMSE among the investigated tissue permutations. This suggests that the GP model handles single-tissue data better than multi-tissue data, showing increasing errors as the number of included tissues increase. In contrast the LSTM model was generally able to predict AIFs with lower overall errors in both single- and multi-tissue data.

Most importantly, even though the LSTM model generated AIFs with lower RMSE, thus better agreement between predicted and reference AIF curve shapes, compared to GP, the result from compartment modelling, in terms of K_i values, showed similar performance between the models. It remains to show in a future study, if this is due to K_i being robust to the AIF variations encountered in the data set, or if it is a limitation of the image-derived reference AIFs, used in this study.

A prerequisite for the MLDIF approach is that representative training data have been collected for the specific mouse strain, tracer and imaging system, including both images and reference AIFs, the latter preferably validated with blood samples. Once an MLDIF model has been trained, it offers several advantages, relative to currently available methods for AIF estimation. Compared to blood sampling, a trained MLDIF model is a non-invasive method, implying

simple and convenient use, without the need for surgery, allowing non-terminal PET experiments for mice. Similar to other image-derived methods, such as simultaneous estimation (Y.-H. D. Fang and Muzic 2008) and Bayesian statistical models (O'Sullivan et al. 2017), MLDIF is based on minimization of an objective function. However, as opposed to the former-mentioned methods, MLDIF is based on well-known ML models that do not require a predetermined function or fine-tuning parameter initialization and limits. Furthermore, as opposed to many image-derived methods, including factor analysis (J. Kim et al. 2006), our experiments indicate that a trained MLDIF model is able to describe both the shape and the amplitude of an image-derived reference AIF. The authors hypothesize that MLDIF models, in experiments with available blood data, needs no blood sample for AIF scaling during prediction, but solely image-derived input data. Lastly, multiple linear regression has shown potential for predicting the AIF in human brain studies (Y. H. Fang et al. 2004), but this method assumes identical AIF shape in all patients, differing only by magnitude. In contrast, MLDIF takes time-dependent input data, and outputs time-dependent AIFs. The model thus accounts for variations in both magnitude and shape, as shown in Figure 3. These variations originate from relative magnitude and shape variations in the image input data, as opposed to absolute AIF scaling, which is possible when blood samples are available.

Because blood data were unavailable, the reference AIF was generated by fitting a well-known AIF model (Dagan Feng, Huang, and Wang 1993) to image-derived data. However, the same reference AIF was used for both reference compartment modelling and for MLDIF model training, thus, a valid comparison can still be made between K_i^{Ref} and K_i^{Model}. The comparison to an image-derived reference AIF does not fully validate the MLDIF method, but does provide an exploratory foundation for this novel and non-invasive AIF estimation method. Nevertheless, ML have previously been successfully applied in various regression tasks (Sapankevych and Sankar 2009; Miles N. Wernick et al. 2014; Erickson et al. 2017), thus in future research, it remains to prove that a reference blood-AIF can be predicted with the MLDIF approach. Moreover, although an attempt was made to avoid the influence of signal spill-in and spill-over effects in this work (Equation 5), it remains to validate that MLDIF can explicitly account for these effects by comparing it to existing partial-volume correction methods (Frouin et al. 2002; E. Kim et al. 2013; Y.-H. D. Fang and Muzic 2008).

The MLDIF approach was verified with FDG in this study, however, based on the robustness of the investigated ML models to variations in the input data,

the authors suggest that these models could be adopted to other tracers by merely retraining the models. With comprehensive validation it is also conceivable that tracers requiring metabolite-correction may be modelled. If validated correctly, this will give a foundation for a simplified MLDIF-based approach in research subsequent to such a validation. In the end, the accuracy of the MLDIF models for a particular PET application will depend on the quality, quantity and relevance of the available training data.

Conclusion

In this study we have shown that two different machine learning-based models, GP and LSTM, can be used for non-invasive AIF prediction in an FDG study of mice. The resulting net-influx rate constants from compartment modelling agreed well with reference values for both models. We recommend the deep-learning based LSTM approach, as this model predicts AIFs with lower errors for both single- and multi-tissue input data, compared to GP.

Acknowledgements

This research was supported by grants HNF1349-17 and SFP1263-15 from the Northern Norway Regional Health Authority.

References

Bartlett, Elizabeth A et al. (2018). "Quantification of Positron Emission Tomography Data Using Simultaneous Estimation of the Input Function : Validation with Venous Blood and Replication of Clinical Studies". In: DOI: 10.1007/s11307-018-1300-1.

Baxt, W.G (1995). "Application of artificial neural networks to clinical medicine". In: *Lancet* 346.8983, pp. 1135–1138. ISSN: 01406736. DOI: 10.1016/S0140-6736(95)91804-3.

Chauhan, Sucheta and Lovekesh Vig (2015). "Anomaly detection in ECG time signals via deep long short-term memory networks". In: *Proc. 2015 IEEE Int. Conf. Data Sci. Adv. Anal. DSAA 2015*. DOI: 10.1109/DSAA.2015.7344872.

Chollet, François (2015). *Keras*. https://keras.io.

Dorraji, S Esmaeil et al. (2016). "PET imaging of Tertiary Lymphoid Structures (TLS) in development of Lupus nephritis". In: *8th Natl. Conf. Med. Imaging MedIm*.

Dürichen, Robert et al. (2015). "Multitask Gaussian processes for multivariate physiological time-series analysis". In: *IEEE Trans. Biomed. Eng.* 62.1, pp. 314–322. ISSN: 15582531. DOI: 10.1109/TBME.2014.2351376.

Erickson, Bradley J. et al. (2017). "Machine Learning for Medical Imaging". In: *RadioGraphics* 37.2, pp. 505–515. ISSN: 0271-5333. DOI: 10.1148/rg.2017160130. arXiv: 15334406.

Fang, Yu-Hua Dean and Raymond F Muzic (2008). "Spillover and partial-volume correction for image-derived input functions for small-animal ^{18}F-FDG PET studies." In: *J. Nucl. Med.* 49.4, pp. 606–14. ISSN: 0161-5505. DOI: 10.2967/jnumed.107.047613.

Fang, Yu Hua et al. (2004). "Estimating the input function non-invasively for FDG-PET quantification with multiple linear regression analysis: Simulation and verification with in vivo data". In: *Eur. J. Nucl. Med. Mol. Imaging* 31.5, pp. 692–702. ISSN: 16197070. DOI: 10.1007/s00259-003-1412-x.

Feng, Dagan, Sung Cheng Huang, and Xinmin Wang (1993). "Models for computer simulation studies of input functions for tracer kinetic modeling with positron emission tomography". In: *Int. J. Biomed. Comput.* 32.2, pp. 95–110. ISSN: 00207101. DOI: 10.1016/0020-7101(93)90049-C.

Feng, D et al. (1997). "A technique for extracting physiological parameters and the required input function simultaneously from PET image measurements:

theory and simulation study." In: *IEEE Trans. Inf. Technol. Biomed.* 1.4, pp. 243–254. ISSN: 1089-7771. DOI: 10.1109/4233.681168.

Frouin, Vincent et al. (2002). "Correction of partial-volume effect for PET striatal imaging: fast implementation and study of robustness." In: *J. Nucl. Med.* 43.12, pp. 1715–1726. ISSN: 0161-5505.

Fueger, Barbara J et al. (2006). "Impact of animal handling on the results of ^{18}F-FDG PET studies in mice." In: *J. Nucl. Med.* 47.6, pp. 999–1006. ISSN: 0161-5505.

Green, L a et al. (1998). "Noninvasive methods for quantitating blood time-activity curves from mouse PET images obtained with fluorine-18-fluorodeoxyglucose." In: *J. Nucl. Med.* 39, pp. 729–734. ISSN: 0161-5505.

Gunn, R N, S R Gunn, and V J Cunningham (2001). "Positron emission tomography compartmental models." In: *J. Cereb. Blood Flow Metab.* 21.6, pp. 635–52. ISSN: 0271-678X. DOI: 10.1097/00004647-200106000-00002.

Hochreiter, Sepp and Paolo Frasconi (2009). "Gradient Flow in Recurrent Nets: The Difficulty of Learning LongTerm Dependencies". In: *A F. Guid. to Dyn. Recurr. Networks*. IEEE. ISBN: 978-0-7803-5369-5. DOI: 10.1109/9780470544037.ch14. arXiv: arXiv:1011.1669v3.

Hochreiter, Sepp and Jürgen Schmidhuber (1997). "Long Short-Term Memory". In: *Neural Comput.* 9.8, pp. 1735–1780. ISSN: 08997667. DOI: 10.1162/neco.1997.9.8.1735. arXiv: 1206.2944.

Hubert, Mia and Stephan Van der Veeken (2008). "Outlier detection for skewed data". In: *J. Chemom.* 22.3-4, pp. 235–246. ISSN: 08869383. DOI: 10.1002/cem.1123.

Keyes, J W (1995). "SUV: standard uptake or silly useless value?" In: *J. Nucl. Med.* 36.10, pp. 1836–1839. ISSN: 0161-5505.

Kim, Euitae et al. (2013). "Partial volume correction using structural-functional synergistic resolution recovery: Comparison with geometric transfer matrix method". In: *J. Cereb. Blood Flow Metab.* 33.6, pp. 914–920. ISSN: 0271678X. DOI: 10.1038/jcbfm.2013.29.

Kim, Joonyoung et al. (2006). "Minimally invasive method of determining blood input function from PET images in rodents." In: *J. Nucl. Med.* 47.2, pp. 330–336. ISSN: 0161-5505, 2159-662X. DOI: 47/2/330[pii].

Kingma, Diederik P. and Jimmy Ba (2014). "Adam: A Method for Stochastic Optimization". In: pp. 1–15. ISSN: 09252312. DOI: http://doi.acm.org.ezproxy.lib.ucf.edu/10.1145/1830483.1830503. arXiv: 1412.6980.

Laforest, Richard et al. (2005). "Measurement of input functions in rodents: Challenges and solutions". In: *Nucl. Med. Biol.* 32.7, pp. 679–685. ISSN: 09698051. DOI: 10.1016/j.nucmedbio.2005.06.012.

Lanz, Bernard, Carole Poitry-Yamate, and Rolf Gruetter (2014). "Image-Derived Input Function from the Vena Cava for ^{18}F-FDG PET Studies in Rats and Mice." In: *J Nucl Med* 55.8, pp. 1380–1388. ISSN: 1535-5667. DOI: 10.2967/jnumed.113.127381.

Lipton, Zachary C., John Berkowitz, and Charles Elkan (2015). "A Critical Review of Recurrent Neural Networks for Sequence Learning". In: pp. 1–38. ISSN: 9781450330633. DOI: 10.1145/2647868.2654889. arXiv: 1506.00019.

Martin Bland, J. and DouglasG. Altman (1986). "Statistical methods for assessing agreement between two methods of clinical measurement". In: *Lancet* 327.8476, pp. 307–310. ISSN: 01406736. DOI: 10.1016/S0140-6736(86)90837-8. arXiv: arXiv:1011.1669v3.

Matthews, Alexander G De G et al. (2017). "GPflow: A Gaussian process library using TensorFlow". In: *J. Mach. Learn. Res.* 18.40, pp. 1–6. arXiv: 1610.08733.

Miles N. Wernick et al. (2014). "Machine Learning in Medical Imaging". In: *IEEE Signal Process Mag.* 27.4, pp. 25–38. ISSN: 1558-349X. DOI: 10.1109/MSP.2010.936730.Machine.

O'Sullivan, Finbarr et al. (2017). "Assessment of a statistical AIF extraction method for dynamic PET studies with ^{15}O water and ^{18}F fluorodeoxyglucose in locally advanced breast cancer patients". In: *J. Med. Imaging* 5.01, p. 1. ISSN: 2329-4302. DOI: 10.1117/1.JMI.5.1.011010.

Phelps, E, J Hoffman, and David E Kuhl (1980). "Noninvasive determination of local cerebral metabolic rate of glucose in man". In: *Am. J. Physiol. - Endocrinol. Metab.* 238.1, E69–E82. ISSN: 0002-9513. DOI: 10.1152/ajpendo.1980.238.1.E69.

Rasmussen, Carl E. and Christopher K. I. Williams (2004). *Gaussian processes for machine learning*. MIT Press. ISBN: 026218253X. DOI: 10.1142/S0129065704001899. arXiv: 026218253X.

Roberts, S. et al. (2013). "Gaussian processes for time-series modelling". In: *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* 371.1984. ISSN: 1364503X. DOI: 10.1098/rsta.2011.0550.

Roccia, Elisa et al. (2019). "Quantifying Brain [^{18}F]FDG Uptake Noninvasively by Combining Medical Health Records and Dynamic PET Imaging Data".

In: *IEEE J. Biomed. Heal. Informatics* 2194.c, pp. 1–7. ISSN: 2168-2194. DOI: 10.1109/JBHI.2018.2890459.

Sapankevych, Nicholas and Ravi Sankar (2009). "Time Series Prediction Using Support Vector Machines: A Survey". In: *IEEE Comput. Intell. Mag.* 4.2, pp. 24–38. ISSN: 1556-603X. DOI: 10.1109/MCI.2009.932254.

Spangler-Bickell, Matthew G. et al. (2016). "The effect of isoflurane on ^{18}F-FDG uptake in the rat brain: a fully conscious dynamic PET study using motion compensation". In: *EJNMMI Res.* 6.1, p. 86. ISSN: 2191-219X. DOI: 10.1186/s13550-016-0242-3.

Sun, Qingnan et al. (2018). "Predicting Blood Glucose with an LSTM and Bi-LSTM Based Deep Neural Network". In: pp. 1–5. arXiv: 1809.03817.

Takikawa, S et al. (1993). "Noninvasive quantitative fluorodeoxyglucose PET studies with an estimated input function derived from a population-based arterial blood curve." In: *Radiology* 188.1, pp. 131–136. ISSN: 0033-8419. DOI: 10.1148/radiology.188.1.8511286.

Theodoridis, Sergios and Konstantinos Koutroumbas (2009). *Pattern Recognition, Fourth Edition*. 4th. Orlando, FL, USA: Academic Press, Inc. ISBN: 1597492728, 9781597492720.

Tonietto, Matteo et al. (2015). "Modelling arterial input functions in positron emission tomography dynamic studies". In: *2015 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* IEEE, pp. 2247–2250. ISBN: 978-1-4244-9271-8. DOI: 10.1109/EMBC.2015.7318839.

Weber, Bruno et al. (2002). "A femoral arteriovenous shunt facilitates arterial whole blood sampling in animals". In: *Eur. J. Nucl. Med.* 29.3, pp. 319–323. ISSN: 03406997. DOI: 10.1007/s00259-001-0712-2.

Weerdt, a P van der et al. (2001). "Image-derived input functions for determination of MRGlu in cardiac ^{18}F-FDG PET scans." In: *J. Nucl. Med.* 42.11, pp. 1622–9. ISSN: 0161-5505.

Wong, K P et al. (2001). "Simultaneous estimation of physiological parameters and the input function–in vivo PET data." In: *IEEE Trans. Inf. Technol. Biomed.* 5.1, pp. 67–76. ISSN: 1089-7771. DOI: 10.1109/4233.908397.

Wu, Hsiao-Ming et al. (2007). "In vivo quantitation of glucose metabolism in mice using small-animal PET and a microfluidic device." In: *J. Nucl. Med.* 48, pp. 837–845. ISSN: 0161-5505. DOI: 10.2967/jnumed.106.038182.