

Cerebral blood flow measurements with ^{15}O -water PET using a non-invasive machine-learning-derived arterial input function

Journal of Cerebral Blood Flow & Metabolism
0(0) 1–13
© The Author(s) 2021



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0271678X21991393
journals.sagepub.com/home/jcbfm



Samuel Kuttner^{1,2,3} , Kristoffer Knutsen Wickstrøm², Mark Lubberink⁴, Andreas Tolf⁵, Joachim Burman⁵, Rune Sundset^{1,3}, Robert Jenssen², Lieuwe Appel⁴ and Jan Axelsson⁶ 

Abstract

Cerebral blood flow (CBF) can be measured with dynamic positron emission tomography (PET) of ^{15}O -labeled water by using tracer kinetic modelling. However, for quantification of regional CBF, an arterial input function (AIF), obtained from arterial blood sampling, is required. In this work we evaluated a novel, non-invasive approach for input function prediction based on machine learning (MLIF), against AIF for CBF PET measurements in human subjects.

Twenty-five subjects underwent two 10 min dynamic ^{15}O -water brain PET scans with continuous arterial blood sampling, before (baseline) and following acetazolamide medication. Three different image-derived time-activity curves were automatically segmented from the carotid arteries and used as input into a Gaussian process-based AIF prediction model, considering both baseline and acetazolamide scans as training data. The MLIF approach was evaluated by comparing AIF and MLIF curves, as well as whole-brain grey matter CBF values estimated by kinetic modelling derived with either AIF or MLIF.

The results showed that AIF and MLIF curves were similar and that corresponding CBF values were highly correlated and successfully differentiated before and after acetazolamide medication. In conclusion, our non-invasive MLIF method shows potential to replace the AIF obtained from blood sampling for CBF measurements using ^{15}O -water PET and kinetic modelling.

Keywords

Arterial input function, cerebral blood flow, Gaussian processes, kinetic modelling, machine learning, ^{15}O -water PET

Received 23 June 2020; Revised 30 November 2020; Accepted 3 January 2021

Introduction

Measurements of cerebral blood flow (CBF) can be used to separate pathological and healthy brain tissue as well as for functional brain research. Tracer kinetic modelling following dynamic ^{15}O -water positron emission tomography (PET) imaging with arterial blood sampling is considered the reference standard for CBF measurements.^{1–6} However, arterial cannulation is an invasive, laborious and time-consuming procedure, and may, due to induction of pain and risk for complications, discourage patients and volunteers from participating in research studies. Furthermore, a useful arterial input function (AIF) curve cannot be obtained

¹Department of Clinical Medicine, UiT The Arctic University of Norway, Tromsø, Norway

²Department of Physics and Technology, UiT The Arctic University of Norway, Tromsø, Norway

³The PET Imaging Center, University Hospital of North Norway, Tromsø, Norway

⁴Department of Surgical Sciences, Radiology, Uppsala University, Uppsala, Sweden

⁵Department of Neuroscience, Uppsala University, Uppsala, Sweden

⁶Department of Radiation Sciences, Umeå University, Umeå, Sweden

Corresponding author:

Samuel Kuttner, Department of Clinical Medicine, UiT The Arctic University of Norway, 9037 Tromsø, Norway.

Email: samuel.kuttner@uit.no

without careful cross calibration of the blood measurement detector and the PET scanner. In addition, because the blood is most commonly sampled from the radial artery, additional corrections for dispersion and delay of the tracer must be applied, to obtain the true AIF for the brain.^{3,7-9}

The use of an image-derived input function (IDIF) has been proposed as an alternative approach to overcome the challenges with the AIF.¹⁰ In brain PET imaging, an IDIF can be measured inside a suitable intracranial blood vessel directly in the reconstructed PET images, for instance in the intracranial carotid arteries.¹¹⁻¹³ Due to the limited spatial resolution of the PET system, and the need for short time-frames during the first pass of the bolus, image-derived methods suffer from both partial volume effects and image noise. These limitations require complex and standardized methods for partial volume correction and artery delineation, which may be difficult to achieve in practice. Recently, a few clinical studies have suggested the use of integrated PET/magnetic resonance imaging (MRI) for deriving an IDIF in brain, where the latter modality is used for artery delineation or even motion correction.¹⁴⁻¹⁸ However these methods are sensitive to registration errors between the modalities and require detailed knowledge of the scanner resolution. Another recent study overcame potential misregistration problems and formed a corrected IDIF by deriving total number of counts and artery volume from the two modalities separately.¹⁹ However, this method was not yet validated with arterial blood sampling and as hybrid PET/MRI is still an emerging modality, to date, image-derived methods are rarely used in larger clinical or research studies.^{10,20}

Alternatively, a standardized, population-based AIF can be calculated as an average AIF from a group of subjects acquired with the same tracer, injection protocol and population, and scaled to the specific subject.^{21,22} However, this method requires at least one blood sample for curve scaling while individual physiological differences and scan-dependent variations are neglected.

An approach on image data with simultaneous estimation of AIF and kinetic parameters has also been reported.²³⁻²⁵ This method, however, assumes a known mathematical AIF model and requires at least one late blood sample for parameter estimation. Recently, non-invasive simultaneous estimation methods were developed that obviate the need for the single late blood sample by using additional input variables from electronic health records into the models.^{26,27} The limitation of such an approach is that a large set of clinical variables must be collected and handled for each patient. These variables may not necessarily be available in the health records for all patients and may even

complicate inclusion of healthy volunteers in research studies.

In this study we use a machine learning-based approach for AIF estimation. Machine learning-methods are especially useful for function estimation and regression.²⁸ Briefly, one seeks to determine a function, f , that predicts the machine-learning-derived input function (MLIF), based on an input vector, \mathbf{x} , composed of multiple image-derived tissue curves, such that $\text{MLIF} = f(\mathbf{x})$. The function f is determined by optimizing hidden parameters to find the best mapping, $\text{AIF} = f(\mathbf{x})$, for a set of training data, where both the AIF and the tissue curves are known. Once the model has been trained and f is known, the MLIF can be predicted for unseen test data using only the tissue-curves extracted from the image data.²⁸

In our previous work, we developed and validated a machine-learning-based input function for ¹⁸F-fluorodeoxyglucose (FDG) in a mouse PET cohort.²⁹ In short, two learning models were evaluated that predicted an AIF from time-activity curves of up to 7 different tissue regions as input. The main limitation with our previous study was the lack of an AIF, thus the models could only be validated against a reference IDIF. However, in mouse PET scanning, the entire body of the mouse fits in the PET field-of-view, thus, time-activity curves from all organs are readily available as input data for the models. We showed that, for instance, the myocardium and liver were important for AIF prediction, because their time-activity curves closely resembled the reference IDIF. In contrast, these blood-rich organs are outside the field-of-view in clinical brain PET imaging, and thus, alternative input curves had to be derived for the MLIF model in the current study.

In the present study, we have further developed the MLIF approach for human ¹⁵O-water brain PET data and evaluated the models against an AIF obtained from continuous arterial-blood sampling. We aimed to show that an AIF could be accurately predicted by an MLIF model using multiple image-derived input data from the carotid arteries. We hypothesized that there were no significant differences in estimated CBF when using either AIF or MLIF, and investigated similarities and differences between image-derived sampling in the brain versus arterial-sampling from the arm. Further, we investigated whether the MLIF method was capable to predict a clinically relevant CBF difference between scans before and after acetazolamide medication.

Materials and methods

Subjects

Pseudonymized data from 25 subjects were retrospectively collected from a completed clinical research

study at Uppsala University Hospital. The data comprised both patients with multiple sclerosis (MS) and healthy volunteers (mean age (range) in years: 40 (23–56); F:M 15:10). In this methodological study, we did not differentiate between the two groups as we considered that the subject's health status had no impact on our evaluation of the MLIF model. Therefore, all authors were blinded for the health status of each subject and thus, no comparisons were made between healthy subjects and MS-patients. The results of the parent study will be reported elsewhere.

The parent study was approved by the Swedish Ethical Review Authority (reference 2014/453). All subjects signed written informed consent prior to inclusion. Since the present work was purely an image analysis methodology study using pseudonymized data, it was not covered by the Swedish or Norwegian regulations on medical research in humans and as such, no additional ethics approval was necessary.

Image acquisition

All subjects underwent two 10 min dynamic brain PET scans on either an ECAT Exact HR+ stand-alone PET scanner (Siemens, Knoxville, TN; $n = 9$) or a Discovery MI PET-Computed tomography (CT) scanner (GE Healthcare, Waukesha, MI; $n = 16$). The scans started simultaneously with an automated bolus injection of 5 MBq/kg ^{15}O -water (^{15}O -water in 5 ml saline at 1 ml/s followed by 35 ml saline at 2 ml/s). Each subject underwent one scan at baseline and one scan 15–30 min after intravenous administration of acetazolamide (9 mg/kg up to a maximum of 1000 mg; 5 min infusion) such that every subject was its own control. Acetazolamide medication dilates the vascular system and thereby it increases the cerebral arterial blood flow velocity.^{30–32} Attenuation correction was based on a 10 min transmission scan with rotating ^{68}Ge rod sources (ECAT) or an ultra-low-dose CT scan (Discovery MI). Images were reconstructed into 26 time-frames (1×10 , 8×5 , 4×10 , 2×15 , 3×20 , 2×30 and 6×60 s). Image reconstruction algorithms were chosen to result in a matching image resolution for the two scanners: ordered subsets expectation maximization with 6 iterations, 8 subsets and a 4 mm Hanning filter (ECAT) and 3 iterations, 16 subsets and a 5 mm Hanning filter (Discovery MI).

In addition, all subjects underwent MRI on a 3 T MRI scanner (Achieva, Philips Healthcare, Best, The Netherlands) with a 32-channel head coil. A three-dimensional T1-weighted gradient echo sequence was obtained with voxel size $1.0 \times 1.0 \times 1.0 \text{ mm}^3$, repetition time = 8.2 ms and echo time = 3.7 ms.

Blood sampling

Continuous arterial blood sampling was performed during 10 min for each scan (3 ml/min) using either an ABSS V3 (Allogg, Mariefred, Sweden; subjects scanned on ECAT) or PBS-100 (Veenstra-Comecer, Joure, The Netherlands; subjects scanned on Discovery MI). A single arterial blood sample was taken through a three-way-valve on the arterial line 5 min post injection and measured in a cross-calibrated well counter for calibration of the continuous arterial blood data.

The measured blood signal, $g(t)$, was affected by dispersion in the vessels and in the detector system tubes. This could be modelled as a convolution of the true AIF, $C_A(t)$, and a dispersion function, $d(t)$ ⁸

$$g(t) = C_A(t) \otimes d(t) \quad (1)$$

A mono-exponential dispersion model was assumed⁸

$$d(t) = \frac{1}{\tau} e^{-t/\tau} \quad (2)$$

where τ is the dispersion constant. An expression for the true AIF, $C_A(t)$, could be obtained by the Laplace transform³³

$$C_A(t) = g(t) + \tau \frac{dg}{dt} \quad (3)$$

The dispersion constant was fixed to 15 s for all subjects in this study, as suggested in the literature.⁸

The dispersion-corrected AIFs were delay-corrected, as described in the 'Image processing' Section.

A visual assessment of the AIF curves was performed to identify abnormal AIFs due to failures in tracer administration or continuous arterial blood sampling. Three subjects were excluded after the visual assessments.

Image processing

All images were corrected for inter-frame motion using in-house written software in Matlab (Mathworks, MA, USA). A simple and objective multi-VOI thresholding method, that could capture blood information from the carotid arteries, was empirically developed. First, to remove noise close to the edge slices, a PET search-volume was defined by trimming 20 voxels in the x-y-plane periphery, and by removing 5 slices in the z-direction. The algorithm for identifying the time-frame for carotid-VOI thresholding was based on a frame-wise graph of whole-brain gray-matter total intensity. The time-frame used for VOI thresholding was the first frame where the total

intensity was larger than 25% of the maximum value in this graph. Three VOIs were generated, comprising the 10, 100 and 1000 highest-intensity voxels (Figure 1a to c). The median voxel value was derived for each time-frame and VOI, resulting in three IDIF time-activity curves, named $IDIF_{10}$, $IDIF_{100}$ and $IDIF_{1000}$ (Figure 1d), which could be interpreted as three different image-derived blood-curves with different amounts of partial volume effect.

Subsequently, to match the AIF sampling, all PET data were interpolated linearly to one second time framing. To correct for delay between the AIF and the PET data, the dispersion corrected AIF was shifted to provide the best overlap with the $IDIF_{10}$ curve, i.e. where the dot product between the two curves was maximized.

In all following analyses, only PET and AIF data from 0–6 min were used, to minimize noise from late parts of the scans. Calibration of the AIF curves with the single arterial blood sample was only conducted for the training data, and thus no AIF or blood sample was required for the test data.

To extract time-activity curves for whole-brain grey matter, T1-weighted MRI images were co-registered to PET images and segmented using SPM8 (The Wellcome Centre for Human Neuroimaging, UCL Queen Square Institute of Neurology, London, UK).

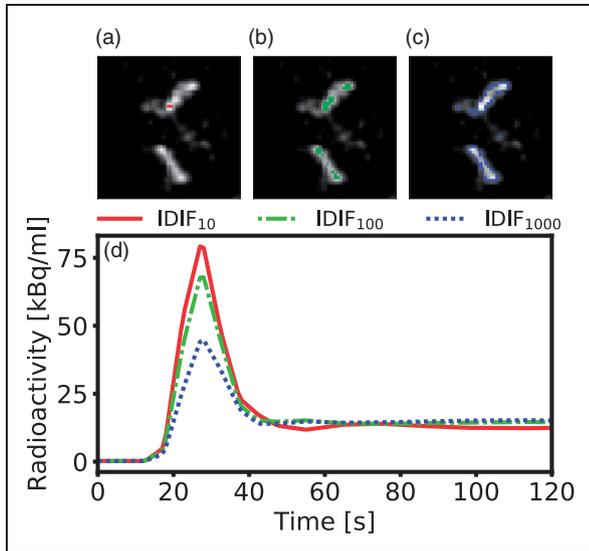


Figure 1. Outline of the VOI thresholding method implemented in this work. (a) to (c) shows an enlarged axial brain PET slice of the optimal time-frame for a representative subject. Highlighted are the parts of the $IDIF_{10}$ (a), $IDIF_{100}$ (b) and $IDIF_{1000}$ (c) comprising the 10, 100 and 1000 highest intensity voxels. (d) The resulting time-activity curves during the first two minutes for $IDIF_{10-1000}$. The $IDIF_{10}$ captures the highest activity in the carotid artery, while the $IDIF_{100}$ and $IDIF_{1000}$ shows a lower activity due to a higher number of voxels included.

All image analysis was performed in the native PET image space.

Function prediction using Gaussian processes

Gaussian processes (GP) is a well-known, non-parametric Bayesian regression method which has been frequently used within machine learning for data-driven function estimation and regression tasks.^{34,35} One advantage with GP is that it predicts not only the mean function, but also its variance, thus providing an uncertainty measure of the model.³⁶ Furthermore, GP, is known to work well with sparse training datasets, as opposed to neural networks.^{37,38}

In GP regression, the output y , is approximated by a probability distribution of functions of the input, \mathbf{x} , such that $f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'))$, which is a generalization of the multivariate Gaussian distribution to infinitely many variables. Here, $m(\mathbf{x})$ is a mean function and $k(\mathbf{x}, \mathbf{x}')$ is a covariance function.³⁶ Assuming N available input-output training samples in a dataset $\{\mathbf{x}_n, y_n\}_{n=1}^N$, each including the three IDIF time-activity curves, \mathbf{x}_n (see Figure 1), with corresponding known AIF, y_n . Then the mean value MLIF of the test sample, $\mathbb{E}[y_*]$, and the variance, $\mathbb{V}[y_*]$, can be predicted by

$$\mathbb{E}[y_*] = \mathbf{k}_*^T (\mathbf{K} + \sigma_\epsilon^2 \mathbf{I})^{-1} \mathbf{y} \quad (4)$$

$$\mathbb{V}[y_*] = k(\mathbf{x}_*, \mathbf{x}_*) - \mathbf{k}_*^T (\mathbf{K} + \sigma_\epsilon^2 \mathbf{I})^{-1} \mathbf{k}_* \quad (5)$$

Here \mathbf{k}_* is the covariance between the training samples \mathbf{x}_n and the test sample \mathbf{x}_* ; $\mathbf{K} = k(\mathbf{x}_n, \mathbf{x}_m)$ is the covariance between all training samples; $\sigma_\epsilon^2 \mathbf{I}$ 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.³⁶

Input function prediction

For input function prediction, leave-one-out cross validation was employed, which is a common validation method in machine learning with limited amounts of data.²⁸ In short, one sample was withdrawn from the dataset and assigned as test sample, while the remaining samples were allocated for training. The three time-activity curves (Figure 1) were used as input vectors into the GP framework, and the MLIF and variance of the test sample were predicted using equations (4) and (5), respectively. The process was repeated by assigning a new sample as test sample, until all samples had been tested once.

In all experiments, the Matérn covariance function was chosen, with $\nu = 5/2$, because it has been reported to generate smooth functions.³⁶ Data normalization

was applied on the input-IDIFs, which is a well-known approach to improve convergence of machine learning models.³⁹ Normalization of the IDIF₁₀ curves was performed by identifying the IDIF₁₀ curve with the highest peak value among all subjects. Subsequently, each IDIF₁₀ curve was normalized by dividing with this peak value. Similarly, the IDIF₁₀₀ and IDIF₁₀₀₀ curves were normalized independently with their respectively found highest peak value among the subjects. Thus, the model was trained with three different normalized IDIFs with values ranging between 0 and 1. The normalization was only a scale factor, meaning that relative amplitudes between subjects remained. GP regression was implemented in Python 3.6.8, using Gpflow 1.2.0, in which the matrix inversion of equation (4) was approximated by Cholesky decomposition. The hyperparameters of the covariance function were optimized by maximizing the logarithm of the marginal likelihood of the training data.⁴⁰

Kinetic modelling

Quantification of CBF was performed on the whole-brain grey-matter region. A single-tissue compartment model was used to generate CBF values. This model assumes that water can diffuse freely between vascular and tissue space, with activity concentrations C_A and C_T , respectively, and is described by the following equation^{41,42}

$$\frac{dC_T(t)}{dt} = K_1 \cdot C_A(t) - k_2 \cdot C_T(t) \quad (6)$$

where $C_A(t)$ is the whole-blood arterial time-activity curve, also known as the AIF. The solution to equation (6) is given by⁴³

$$C_T(t) = K_1 \cdot C_A(t) \otimes e^{-k_2 \cdot t} \quad (7)$$

where \otimes denotes mathematical convolution. The activity concentration measured with PET, C_{PET} , is modelled as the sum of the tissue compartment, $C_T(t)$, and the fractional arterial blood volume in the tissue, V_A ,

such that

$$C_{PET}(t) = (1 - V_A) \cdot C_T(t) + V_A \cdot C_A(t) \quad (8)$$

For tracers with high extraction fraction relative to the blood flow, such as ¹⁵O-water, CBF equals K_1 .^{43,44}

Evaluation design

The aim of the current work was to investigate whether an AIF could be accurately predicted using an MLIF model in baseline and acetazolamide scans of the same subject. In a first case, a GP model named MLIF₁ was trained and subsequently tested on baseline scans using leave-one-out cross validation. Similarly, another GP model, named MLIF₂, was trained and subsequently tested on acetazolamide scans using leave-one-out cross validation.

In a routine setting, it is of interest to train a predictive model on normal subjects and apply the same model on a disease, or medicated group. Therefore, in a second case, an additional GP model named MLIF^{*}₁, was trained on all baseline scans, and subsequently tested on all acetazolamide scans. Here, the asterisk (*) in MLIF^{*}₁ emphasises that all subjects from the baseline scan were included in the training data for this model, as opposed to the leave-one-out training for MLIF₁. Essentially, one might see MLIF^{*}₁ and MLIF₁ as two models trained on the same dataset, as the difference is only one subject.

Finally, we hypothesized that MLIF^{*}₁ might be more representative for a local AIF in the brain, compared to an AIF sampled in the wrist. Therefore, in a third case, we aimed to evaluate the CBF increase from baseline to acetazolamide scans obtained by the two different methods in *case 1* and *case 2*, by using the carotid arteries with the ten highest-intensity voxels (IDF₁₀) as input function during kinetic modelling.

The evaluation design is summarized in Table 1.

Evaluation methods

The GP predicted MLIF curves were first compared point-by-point to AIF using orthogonal regression.

Table 1. The evaluation design of the MLIF method.

Case	Training data	Testing data	Input function	Procedure
1	Baseline	Baseline	MLIF ₁	Leave-one-out
	Acetazolamide	Acetazolamide	MLIF ₂	Leave-one-out
2	Baseline	Baseline	MLIF ₁	Leave-one-out
	Baseline	Acetazolamide	MLIF [*] ₁	All baseline scans
3	^a	Baseline	IDF ₁₀ ^a	Only used to calculated change between baseline and acetazolamide CBF
	^a	Acetazolamide	IDF ₁₀ ^a	

^aIn case 3, no GP prediction was used, but instead, kinetic modelling was based on the 10 highest-intensity carotid artery voxels (IDF₁₀) as input function.

Subject scans with regression slopes that were more than three scaled-median absolute-deviations away from the median slope, were considered outliers, and removed from further model comparisons.⁴⁵ Time-activity curves averaged over subjects were calculated for: whole-brain grey matter, AIF, the MLIF models and IDIF₁₀. CBF and V_A were estimated for whole-brain grey matter using kinetic modelling with both AIF and the GP-predicted MLIF (*case 1* and *case 2*) as well as with the IDIF₁₀ time-activity curve as input function (*case 3*).

MLIF-based CBF estimates were compared with the one based on AIF by paired t-test ($\alpha=0.05$), ratio calculation, orthogonal regression and Bland-Altman plots.⁴⁶ Normality was assessed using quantile-quantile plots.

The GP variance (equation (5)) was considered as a measure of prediction error. This measure was evaluated by the relationship between the magnitude of the CBF-ratio (CBF_{MLIF}/CBF_{AIF}) and GP variance for *case 1*.

Results

CBF values based on AIF (CBF_{AIF}) and MLIF (CBF_{MLIF}) are shown in Supplementary Table S1. The average CBF_{AIF} and CBF_{MLIF} were similar for both baseline and acetazolamide scans, about 0.45 and 0.60 ml·min⁻¹·g⁻¹, respectively. The mean CBF_{MLIF}/CBF_{AIF} ratio was 1.04 and 1.03 for baseline and acetazolamide, respectively. No significant

differences were found between the average CBF_{AIF} and CBF_{MLIF} for either scan.

Individual data points for MLIF and AIF from *case 1* are shown as a scatter plot in Figure 2(a). Based on the outlier removal criterion, four scans were removed from model comparisons. There is a strong overall linear relationship between AIF and MLIF curves for both baseline (slope: 0.8) and acetazolamide scans (slope: 0.9). Individual r^2 values were high for both baseline (mean: 0.90, SD: 0.1) and acetazolamide (mean: 0.93, SD: 0.07). Histograms of slope values indicate slopes close to unity for most subjects for both baseline (median slope: 0.86, Figure 2b) and acetazolamide (median slope: 0.87, Figure 2c).

Displaying the CBF data from *case 1* as a scatter plot (Figure 3a), a strong linear relationship and high overall correlation ($r^2>0.9$) between CBF_{AIF} and CBF_{MLIF} was obtained. Bland-Altman analysis (Figure 3b) exhibited a prediction bias close to zero.

The fractional arterial blood volume, V_A , from equation (8), was found to be near zero for all scans (0.001 ± 0.003).

As a visual illustration of the effect of prediction errors on estimation of CBF, the AIF and GP-predicted MLIF from the baseline scan for four representative subjects are shown in Figure 4. Figure 4(a) and (b) comprises two examples with less than 3% difference between CBF_{AIF} and CBF_{MLIF} , while Figure 4 (c) and (d) display cases with substantial differences between both methods. Based on the CBF ratios (CBF_{MLIF}/CBF_{AIF}), it can be observed that an over-prediction of the AIF peak (Figure 4c) results in an

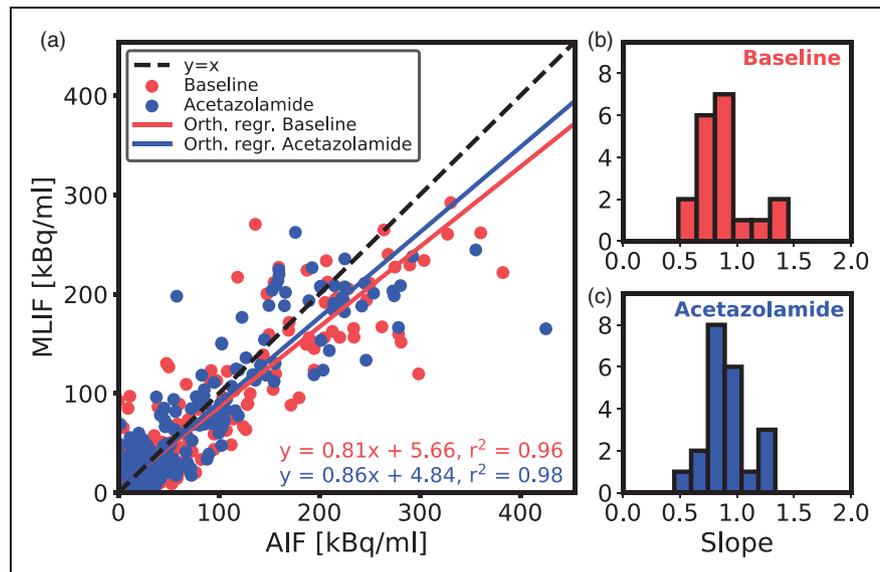


Figure 2. (a) Scatter plot of MLIF and AIF data points for all subjects for baseline (blue) and acetazolamide (red) scans based on *case 1*. The 1 s time frames were interpolated back to the original PET time framing (26 time-frames). The solid lines are the orthogonal regression fits. (b and c) Histogram of orthogonal regression slopes for individual subjects for baseline (b) and acetazolamide (c) scans.

underestimation of CBF, while an underpredicted AIF peak (Figure 4d) ends in an overestimated CBF.

Figure 5 displays box plots of whole-brain grey matter CBF for baseline and acetazolamide scans, when using different prediction models for MLIF (see Table 1).

The GP variance was evaluated as prediction error measure in *case 1* (Supplementary Figure S1). No relationship was found between the magnitude of the CBF_{MLIF}/CBF_{AIF} ratios and the GP variance values. The predicted variance was not further considered in this work.

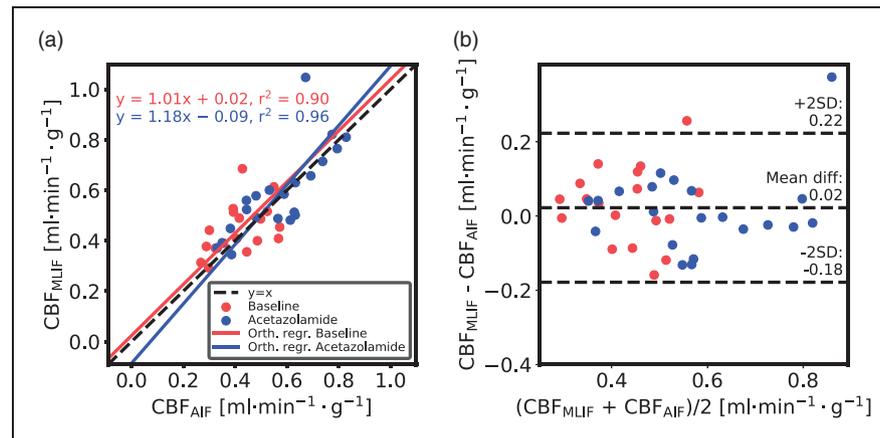


Figure 3. Evaluation of the GP-predicted MLIF for baseline (blue) and acetazolamide (red) scans based on *case 1*. (a) Scatter plot of MLIF-based and AIF-based CBF. The solid lines are the orthogonal regression fits. (b) Bland-Altman plot of *case 1*.

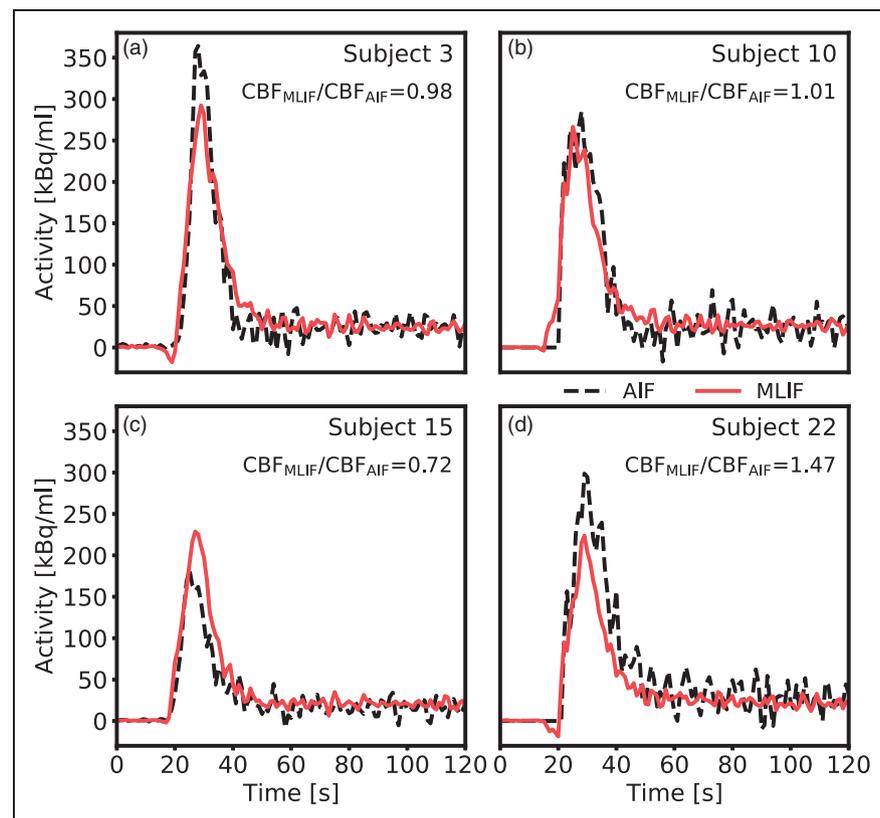


Figure 4. Comparison of AIF (black dashed line) and GP-predicted MLIF (red line) for the baseline scan of four representative subjects. (a and b) Two scans with less than 3% difference between CBF_{AIF} and CBF_{MLIF} . (c) Representative example of a scan where the MLIF overpredicts the AIF peak, resulting in an underestimation of the calculated CBF. (d) Representative example of a scan where the MLIF underpredicts the AIF peak, resulting in an overestimation of the calculated CBF.

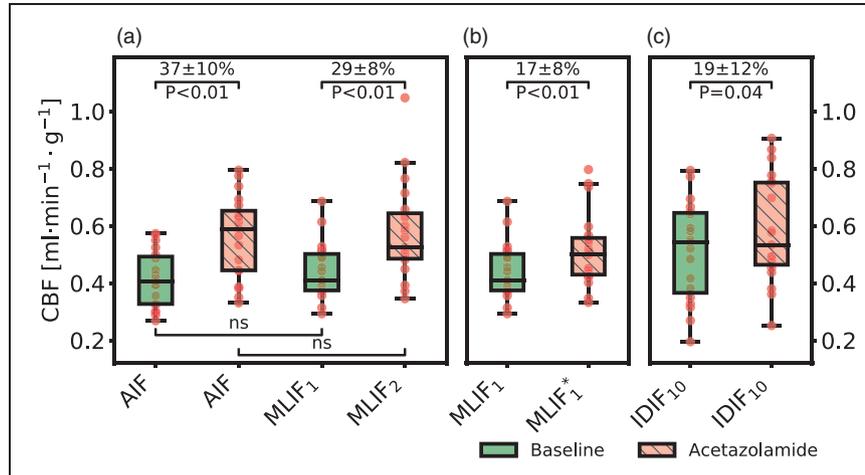


Figure 5. Box plot of estimated whole-brain grey matter CBF for baseline and acetazolamide scans, when using different input functions for case 1 (a), case 2 (b) and case 3 (c). For an explanation of the cases, see Table 1. For visual purposes, the data in panel (c) was scaled to match the range of the AIF-based CBF values in panel (a). Percentage differences are shown as mean \pm 95% confidence interval, and P values are based on paired t-test. In the box plots, red points indicate the data points; the horizontal line and the black box represent median and interquartile range (25th to 75th percentile), respectively; the whiskers indicate the maximum and minimum data point up to $1.5 \times$ interquartile range.

In *case 1* we evaluated the MLIF method by training and applying separate GP models for baseline (MLIF₁) and acetazolamide (MLIF₂) data. There were no significant differences between mean CBF_{AIF} and mean CBF_{MLIF} in neither baseline, nor acetazolamide scans. Furthermore, following acetazolamide medication, we found on average 37% increase ($P < 0.01$) in mean CBF_{AIF} and 29% increase ($P < 0.01$) in mean CBF_{MLIF} (Figure 5a). For *case 1*, both the AIF and the MLIF based methods resulted in similar CBF values, while the CBF increase after acetazolamide medication was lower for the MLIF method. The correlation coefficient between AIF and MLIF-based CBF changes for *case 1* was 0.62.

In *case 2*, a predictive model (MLIF*₁) was trained on baseline scans and then applied on acetazolamide scans. Similar to MLIF₁, the average CBF_{MLIF*1} was non-significantly different from the corresponding CBF_{AIF} after acetazolamide medication. Also, there was still a significant increase (17%, $P < 0.01$) in CBF between baseline and acetazolamide scans (Figure 5b), but notably smaller than the 29% increase observed in *case 1*. The correlation coefficient between AIF and MLIF-based CBF changes for *case 2* was 0.14.

In a final case, the relative CBF increase from baseline to acetazolamide scans was investigated by using the IDIF₁₀ as input function for each scan. A significant CBF increase of 19% ($P = 0.04$) was found after acetazolamide medication (Figure 5c), which was comparable to the change observed in *case 2*.

As CBF is based on underlying time-activity curves, we proceeded to investigate the differences in CBF

after acetazolamide medication in the first two cases by inspecting the mean time-activity curves across subjects in Figure 6. We observed that the local brain input function (IDIF₁₀) (Figure 6e) showed a shape-dependence on acetazolamide medication which was not reflected in the AIF measured in the wrist (Figure 6b). In the AIF, the baseline and acetazolamide curves were similar, while for IDIF₁₀, the acetazolamide curve had a larger area-under-curve compared to baseline. Also, for IDIF₁₀, there was a slight shift in the mean time-activity curves between baseline and acetazolamide, which was not visible for the AIF. Thus, the AIF measured in the wrist does not reflect physiological changes due to acetazolamide, which are apparent in the local brain input (IDIF₁₀).

Discussion

Tracer kinetic modelling of dynamic PET data requires accurate knowledge of an AIF, conventionally measured through arterial blood sampling. Our aim was to investigate whether an AIF could be predicted as accurately by an MLIF model using solely image-derived input data from the carotid arteries.

AIF and predicted MLIF curves were found to be similar, with no significant difference between whole-brain grey matter CBF_{AIF} and CBF_{MLIF} estimates. Furthermore, the correlation between the CBF estimates was $r^2 > 0.9$ and the mean differences were close to zero. The MLIF model was also able to accurately predict an increased CBF after acetazolamide medication, when trained on post-acetazolamide data.

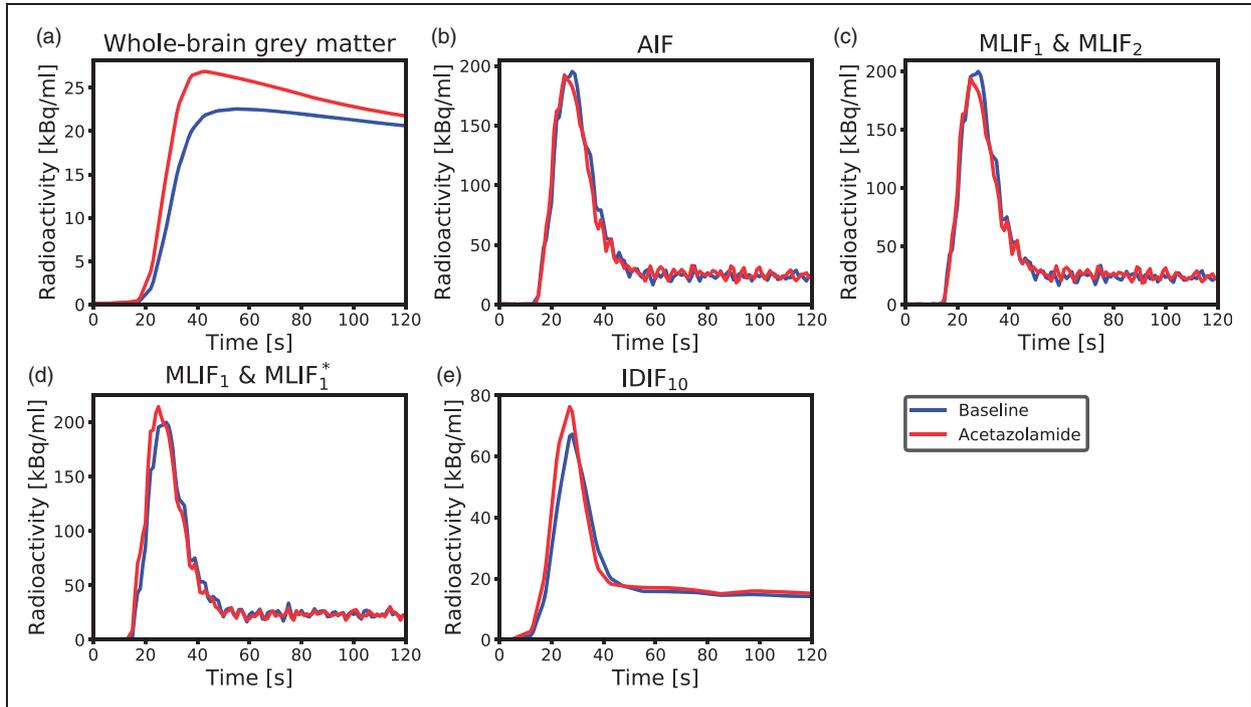


Figure 6. Mean time-activity curves across subjects for baseline (blue) and acetazolamide (red) scans during the first 2 min of PET scanning. (a) Measured radioactivity uptake in whole-brain grey matter. (b) AIF. (c) MLIF₁ model for baseline, and MLIF₂ for acetazolamide scans. (d) MLIF₁ model for baseline, and MLIF₁* for acetazolamide scans. (e) IDIF₁₀ input function based on the 10 highest voxels in the carotid arteries.

Altogether the results indicate that the MLIF method has potential as an alternative AIF for generation of CBF values using ¹⁵O-water PET and kinetic modelling, which in clinical practice implies evading of arterial cannulation.

Initially, we evaluated an MLIF model trained on pooled baseline and acetazolamide data. However, this resulted in inferior generalization to new samples compared to when training was done separately for baseline and acetazolamide scans (data not shown). The reason for this difference between the approaches may be because the input IDIFs under the two conditions vary in amplitude (Figure 6e), while the AIF (Figure 6b) does not. Therefore, in a first case, input functions of baseline scans were predicted using an MLIF model trained on baseline data (MLIF₁), and, similarly, input functions of acetazolamide scans were predicted using an MLIF model trained on acetazolamide data (MLIF₂). There was a strong linear relationship between the data points from AIF and predicted MLIF curves, although the peak values were systematically underestimated (slope <1 in Figure 2). In the time-activity curves, the number of high data values, acquired during the first pass peak, is low compared to the number of low values, acquired during the rest of the scan. Also, the standard deviation around the peaks was observed to be larger than at the

tails (data not shown). We speculate that this imbalance affects the GP models and results in the systematically underestimated peak values. Previous work has shown that when data is limited, the GP model may underestimate the mean function.⁴⁷

For both baseline and acetazolamide, the mean time-activity curves for AIF (Figure 6b), and MLIF₁, MLIF₂ (Figure 6c) appeared similar in shape. Following kinetic modelling, the average CBF values obtained using an AIF are in line with previously published work before^{48,49} and after administration of acetazolamide.⁴⁸ No significant differences were found between AIF-based and MLIF-based CBF estimates in whole-brain grey matter for neither baseline nor acetazolamide scans (Figure 5a). Also, a slope close to unity, an $r^2 > 0.9$ and low bias between CBF_{AIF} and CBF_{MLIF} estimates pointed towards an acceptable agreement of both methods (Figure 3a). Across both baseline and acetazolamide scans, differences between CBF_{AIF} and CBF_{MLIF} estimates were relatively small (Figure 3b), although several subjects had relative CBF errors of > 20% for the baseline scans while there was a somewhat lower spread for the acetazolamide scans. Our hypothesis is that a GP model trained on acetazolamide data (MLIF₂) generalize better to new samples, compared to a model trained on baseline data (MLIF₁). We speculate that the reason for this is that

a larger range of CBF values is a result from a larger range of input values ($IDIF_{10-1000}$), for acetazolamide, compared to baseline scans. This may have resulted in $MLIF_2$ being valid for a wider range of unseen samples, compared to $MLIF_1$, as mentioned and illustrated in Supplementary material. Nevertheless, the CBF increase between baseline and acetazolamide scans was highly correlated to that of the AIF in *case 1*. Furthermore, it was observed that the shape of the input function had an impact on the accuracy of the MLIF-based CBF estimates (Figure 4). Evidently, an underpredicted AIF peak resulted in an overestimated CBF while an overpredicted AIF peak caused an underestimation of CBF. This can be explained by the inverse relationship between K_1 and C_A in equation (7).

In a second case, we trained an MLIF model on all baseline scans and then applied that model on the acetazolamide scans ($MLIF_1^*$). The idea was to resemble a situation where the baseline scans reflected a database comprising healthy subject with a normal CBF whereas the acetazolamide scans reflected clinical data from patients with an altered CBF. Even in this scenario significant differences were found between the whole-brain grey matter CBF from the baseline and acetazolamide scan. However, the CBF increase between baseline and acetazolamide scans in this case displayed small between-subject variation (data not shown) and, maybe because of that, low correlation to the AIF CBF changes. Furthermore, unexpectedly, the difference in CBF was only 17% (Figure 5b) compared to 29% in the previous case (Figure 5a).

We found this difference between baseline and acetazolamide response striking. The local brain input function ($IDIF_{10}$) (Figure 6e) showed a shape-dependence on acetazolamide which was not reflected in the AIF measured in the wrist (Figure 6b). Consequently, CBF calculated with a local brain input would by necessity be different from AIF-based CBF. This effect remained also for the $MLIF_1$ and $MLIF_1^*$ models (Figure 6d) and could possibly explain the observed differences in CBF change.

In order to quantify the differences in CBF after acetazolamide medication in the first two cases, we attempted using a carotid artery region ($IDIF_{10}$) as input function for both baseline and acetazolamide scans (*case 3*). In *case 3* a significant difference of 19% was found between baseline and acetazolamide scans which was similar to the observed difference found in *case 2*. This supports that the difference in acetazolamide provocation results were caused by the different input functions in brain (*case 2*, Figure 6d) and in the wrist (*case 1*, Figure 6c). Note that, *case 3* was used only to investigate the relative increase in CBF between baseline and acetazolamide scans

found in *case 1* and *case 2*. $IDIF_{10}$ cannot be used as a substitute for AIF, due to the limitations of image derived methods, as described in the Introduction.

We suggest that the above discussed difference in input function curve shape might be explained in part by effects on the vascular system after acetazolamide medication. Acetazolamide dilates the vascular system and increases the cerebral blood flow velocity,^{30–32} which explains the increased mean time–activity curve in whole-brain grey matter after acetazolamide medication (Figure 6a). A lower back-pressure due to dilated vasculature together with increased blood velocity^{30,31} could possibly also explain the observed effect on $MLIF_1^*$ (Figure 6d) and on the $IDIF$ (Figure 6e). Figure 6e also indicated that the ^{15}O -water tracer arrived earlier to the brain after acetazolamide medication compared to the baseline scan, resulting in a slight shift of the mean time–activity curves. This observation might also support that the differences in the $IDIF_{10}$ peaks between scans were partly due to an enhanced cerebral blood-flow velocity in the acetazolamide scan. An additional contributing effect to the increased amplitude of the $IDIF$ after acetazolamide may be an increased spill-over from tissue due to increased brain uptake.

In summary, when training and testing on the same scans, similar CBF estimates for whole-brain grey matter are obtained when using AIF and MLIF (*case 1*). However, when using baseline scans for training followed by applying the model to the acetazolamide scans (*case 2*) the blood input curves are higher for the MLIF model, possibly due to the increase in blood flow velocity after acetazolamide medication. Consequently, CBF_{MLIF} for whole-brain grey matter was lower for acetazolamide, compared to baseline, and the difference in CBF before and after acetazolamide medication was reduced from 29% to 17%. Although these relative changes were different for *case 1* and *case 2*, both were significant and hence suggesting that the MLIF method has a clinical potential to differentiate baseline from acetazolamide scans.

A prerequisite for the MLIF approach is that representative training data have been collected for the specific tracer and imaging system, including both images and blood AIFs. Once an MLIF model has been trained, it offers several advantages, compared to various other image-derived and population-based methods.^{11–13,20–25} A trained MLIF model is a non-invasive method describing both the shape and the amplitude of an AIF, without any need for calibration blood samples. The MLIF models represents a learned transformation, that directly maps the image-derived input data, to a ready-to-use AIF, by inherently

correcting for partial volume effect, with no predefined assumption of the model function. Furthermore, the input data required by the MLIF approach consists of only three carotid artery regions which can be objectively and automatically segmented in the PET images. This makes the MLIF approach simple and convenient to use, without the need for MRI-based artery segmentation.^{14–19}

One limitation of this study is that the GP models were trained on at most 22 samples (MLIF*₁), which might have resulted in an inferior generalization of the model to new samples which were dissimilar from the training data. To investigate the robustness of the MLIF model to unseen test data, new IDIF curves were created by scaling the existing input IDIF time–activity curves for each subject during the leave-one-out testing (Supplementary material). When the input data was scaled, the MLIF model was stable for input ranges encountered in the training data. For higher and lower scale factors, the model performance was gradually degraded (Supplementary Figure S2A). The drop of performance of machine learning models outside the range of training data, so called *domain shift*, is expected and well known.^{50–52}

In a clinical setting where an existing MLIF model is applied to a patient, it is important that the resulting CBF is reliable. We have reported four outliers based on abnormal regression slopes between the AIF and MLIF data points. However, we cannot know if it was the AIF or the MLIF curves that were abnormal. The corresponding CBF values for these outlier scans all indicate that the AIF was abnormally high as indicated by $CBF_{MLIF}/CBF_{AIF} \ll 1$ (Table S1). This suggests that the outlier was caused by the AIF and not by the MLIF model. Future research should investigate methods for quality control of predicted MLIF curves and CBF values from a trained model.

Another limitation of the study is the combination of healthy and MS patients in the data set, as well as data from two different scanners. This increased the heterogeneity in the input data to the MLIF models, and in combination with the effects of acetazolamide, this might have affected the results. However, the limited number of subjects does not allow to study these effects further in detail. Further, the evaluation of our MLIF model was mainly based on differences in whole-brain gray matter CBF. We expect similar results to be obtained for CBF in smaller brain regions, but this aspect should be investigated in future studies. Also, the test-retest variability of the MLIF method should be investigated and compared to that of the blood AIF,⁵³ as well as the evaluation of the model sensitivity in relation to aspects such as injected dose, time-framing, reconstruction settings and different type of scanners. Finally, the generalizability of the MLIF

method to other diseases should be investigated. For example, it is known that carotid stenosis alters the temporal shape of the AIF,⁵⁴ which might have implications for MLIF models trained on baseline scans and applied to patients with pathological arterial vasculature.

In this study, different MLIF models were evaluated with ¹⁵O-labeled water. In our previous research²⁹ a machine learning approach was also feasible for AIF prediction using FDG, although not yet evaluated in clinical data. Thus, we suggest that the method can be adopted to other tracers by merely training similar MLIF models. With proper validation, it may also be conceivable that tracers requiring metabolite correction of the AIF can be included in the prediction model. As for all data-driven models, the accuracy of the MLIF approach for a particular PET application will depend on the quality and quantity of the available training data. Nevertheless, MLIF opens for simplified and non-invasive input function measurements, and thereby potentially eliminating the need for extensive arterial blood sampling in future PET studies.

In conclusion, we demonstrated that our non-invasive MLIF prediction method may be a viable alternative for CBF measurements using ¹⁵O-water PET and kinetic modelling, which in clinical practice implies evading of arterial cannulation. The MLIF method successfully differentiated CBF values before and after acetazolamide medication.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Northern Norway Regional Health Authority [grant number HNF1349-17]. The parent study was supported by BiogenIdec Sweden AB and The Swedish Foundation for MS Research.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Authors' contributions

All authors contributed substantially to the scientific process leading to this manuscript. JA, KKW, ML, RJ, RS and SK contributed in conception and design. AT, JB, LA and ML were responsible for acquisition of clinical PET data (parent study). KKW, RJ and SK implemented the machine learning models. JA and SK analysed the data. JA, LA, ML and SK interpreted the results and drafted the manuscript. All authors reviewed and approved the final manuscript.

ORCID iDs

Samuel Kuttner  <https://orcid.org/0000-0001-7747-9003>

Jan Axelsson  <https://orcid.org/0000-0002-3731-3612>

Supplementary material

Supplemental material for this article is available online.

References

1. Frackowiak RSJ, Jones T, Lenzi GL, et al. Regional cerebral oxygen utilization and blood flow in normal man using oxygen-15 and positron emission tomography. *Acta Neurol Scand* 1980; 62: 336–344.
2. Lassen NA. Regional cerebral blood flow measurement in man. *Arch Neurol* 1963; 9: 615.
3. Herscovitch P, Markham J and Raichle ME. Brain blood flow measured with intravenous H₂(15)O. I. Theory and error analysis. *J Nucl Med* 1983; 24: 782–789.
4. Raichle ME, Martin WR, Herscovitch P, et al. Brain blood flow measured with intravenous H₂(15)O. II. Implementation and validation. *J Nucl Med* 1983; 24: 790–798.
5. Mintun MA, Raichle ME, Martin WR, et al. Brain oxygen utilization measured with O-15 radiotracers and positron emission tomography. *J Nucl Med* 1984; 25: 177–187.
6. Wintermark M, Sesay M, Barbier E, et al. Comparative overview of brain perfusion imaging techniques. *Stroke* 2005; 36: e83–e99.
7. Dhawan V, Conti J, Mernyk M, et al. Accuracy of PET RCBF measurements: effective of time shift between blood and brain radioactivity curves. *Phys Med Biol* 1986; 31: 507–514.
8. Iida H, Kanno I, Miura S, et al. Error analysis of a quantitative cerebral blood flow measurement using H₂15O autoradiography and positron emission tomography, with respect to the dispersion of the input function. *J Cereb Blood Flow Metab* 1986; 6: 536–545.
9. Koeppe RA, Hutchins GD, Rothley JM, et al. Examination of assumptions for local cerebral blood flow studies in PET. *J Nucl Med* 1987; 28: 1695–1703.
10. Muzi M, O’Sullivan F, Mankoff DA, et al. Quantitative assessment of dynamic PET imaging data in cancer imaging. *Magn Reson Imaging* 2012; 30: 1203–1215.
11. Mourik JEM, van Velden FHP, Lubberink M, et al. Image derived input functions for dynamic high resolution research tomograph PET brain studies. *Neuroimage* 2008; 43: 676–686.
12. Mourik JEM, Lubberink M, Schuitemaker A, et al. Image-derived input functions for PET brain studies. *Eur J Nucl Med Mol Imaging* 2009; 36: 463–471.
13. Sari H, Erlandsson K, Law I, et al. Estimation of an image derived input function with MR-defined carotid arteries in FDG-PET human studies using a novel partial volume correction method. *J Cereb Blood Flow Metab* 2017; 37: 1398–1409.
14. Islam MM, Tsujikawa T, Mori T, et al. Estimation of arterial input by a noninvasive image derived method in brain H₂ 15 O PET study: confirmation of arterial location using MR angiography. *Phys Med Biol* 2017; 62: 4514–4524.
15. Okazawa H, Higashino Y, Tsujikawa T, et al. Noninvasive method for measurement of cerebral blood flow using O-15 water PET/MRI with ASL correlation. *Eur J Radiol* 2018; 105: 102–109.
16. Sundar LKS, Muzik O, Rischka L, et al. Towards quantitative [18F]FDG-PET/MRI of the brain: automated MR-driven calculation of an image-derived input function for the non-invasive determination of cerebral glucose metabolic rates. *J Cereb Blood Flow Metab* 2019; 39: 1516–1530.
17. Su Y, Arbelaez AM, Benzinger TLS, et al. Noninvasive estimation of the arterial input function in positron emission tomography imaging of cerebral blood flow. *J Cereb Blood Flow Metab* 2013; 33: 115–121.
18. Su Y, Vlassenko AG, Couture LE, et al. Quantitative hemodynamic PET imaging using image-derived arterial input function and a PET/MR hybrid scanner. *J Cereb Blood Flow Metab* 2017; 37: 1435–1446.
19. Khalighi MM, Deller TW, Fan AP, et al. Image-derived input function estimation on a TOF-enabled PET/MR for cerebral blood flow mapping. *J Cereb Blood Flow Metab* 2018; 38: 126–135.
20. Zanotti-Fregonara P, Chen K, Liow J-S, et al. Image-derived input function for brain PET studies: many challenges and few opportunities. *J Cereb Blood Flow Metab* 2011; 31: 1986–1998.
21. Takikawa S, Dhawan V, Spetsieris P, et al. Noninvasive quantitative fluorodeoxyglucose PET studies with an estimated input function derived from a population-based arterial blood curve. *Radiology* 1993; 188: 131–136.
22. Eberl S, Anayat AR, Fulton RR, et al. Evaluation of two population-based input functions for quantitative neurological FDG PET studies. *Eur J Nucl Med* 1997; 24: 299–304.
23. Feng D, Wong K-P, Wu C-M, et al. A technique for extracting physiological parameters and the required input function simultaneously from PET image measurements: theory and simulation study. *IEEE Trans Inf Technol Biomed* 1997; 1: 243–254.
24. Wong K-P, Feng D, Meikle SR, et al. Simultaneous estimation of physiological parameters and the input function - in vivo PET data. *IEEE Trans Inf Technol Biomed* 2001; 5: 67–76.
25. Bartlett EA, Ananth M, Rossano S, et al. Quantification of positron emission tomography data using simultaneous estimation of the input function: validation with venous blood and replication of clinical studies. *Mol Imaging Biol* 2019; 21: 926–934.
26. Mikhno A, Zanderigo F, Ogden RT, et al. Toward non-invasive quantification of brain radioligand binding by combining electronic health records and dynamic PET imaging data. *IEEE J Biomed Health Inform* 2015; 19: 1271–1282.
27. Rocca E, Mikhno A, Ogden RT, et al. Quantifying brain [18 F]FDG uptake noninvasively by combining medical

- health records and dynamic PET imaging data. *IEEE J Biomed Health Inform* 2019; 23: 2576–2582.
28. Theodoridis S and Koutroumbas K. *Pattern recognition*. 4th ed. New York: Academic Press, Inc., 2009.
 29. Kuttner S, Wickstrøm KK, Kalda G, et al. Machine learning derived input-function in a dynamic 18 F-FDG PET study of mice. *Biomed Phys Eng Express* 2020; 6: 015020.
 30. Dahl A, Russell D, Rootwelt K, et al. Cerebral vasoreactivity assessed with transcranial Doppler and regional cerebral blood flow measurements. *Stroke* 1995; 26: 2302–2306.
 31. Démolis P, Chalon S and Giudicelli JF. Acetazolamide-induced vasodilation in the carotid vascular bed in healthy volunteers. *J Cardiovasc Pharmacol* 1995; 26: 841–844.
 32. Hubner KF. PET imaging in neurology. *J Nucl Med Technol* 1990; 18: 229–234.
 33. Meyer E. Simultaneous correction for tracer arrival delay and dispersion in CBF measurements by the H215O autoradiographic method and dynamic PET. *J Nucl Med* 1989; 30: 1069–1078.
 34. Roberts S, Osborne M, Ebdon M, et al. Gaussian processes for time-series modelling. *Philos Trans R Soc A Math Phys Eng Sci* 2012; 371: 20110550.
 35. Dürichen R, Pimentel MAF, Clifton L, et al. Multitask Gaussian processes for multivariate physiological time-series analysis. *IEEE Trans Biomed Eng* 2015; 62: 314–322.
 36. Rasmussen CE and Williams CKI. *Gaussian processes for machine learning*. Cambridge: MIT Press, 2004.
 37. Cutajar K. Broadening the scope of gaussian processes for large-scale learning, <https://www.eurecom.fr/publication/5852> (2019, accessed 11 February 2020).
 38. McNeish D. On using Bayesian methods to address small sample problems. *Struct Equ Model A Multidiscip J* 2016; 23: 750–773.
 39. Ioffe S and Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. *arXiv*. Epub ahead of print 10 February 2015. DOI: 10.1007/s13398-014-0173-7.2.
 40. Matthews AGDG, van der Wilk M, Nickson T, et al. GPflow: a Gaussian process library using TensorFlow. *J Mach Learn Res* 2016; 18: 1–6.
 41. Watabe H, Ikoma Y, Kimura Y, et al. PET kinetic analysis—compartmental model. *Ann Nucl Med* 2006; 20: 583–588.
 42. Svarer C, Madsen K, Hasselbalch SG, et al. MR-based automatic delineation of volumes of interest in human brain PET images using probability maps. *Neuroimage* 2005; 24: 969–979.
 43. Morris ED, Endres CJ, Schmidt KC, et al. Kinetic modeling in positron emission tomography. In: *Emission tomography*. Amsterdam: Elsevier, 2004, chapter 23.
 44. Quarles RP, Mintun MA, Larson KB, et al. Measurement of regional cerebral blood flow with positron emission tomography: a comparison of [15 O]water to [11 C]butanol with distributed-parameter and compartmental models. *J Cereb Blood Flow Metab* 1993; 13: 733–747.
 45. Hubert M and Van der Veecken S. Outlier detection for skewed data. *J Chemometrics* 2008; 22: 235–246.
 46. Martin Bland J and Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; 327: 307–310.
 47. Richardson RR, Osborne MA and Hovey DA. Gaussian process regression for forecasting battery state of health. *J Power Sources* 2017; 357: 209–219.
 48. Grandin C, Bol A, Smith A, et al. Absolute CBF and CBV measurements by MRI bolus tracking before and after acetazolamide challenge: repeatability and comparison with PET in humans. *Neuroimage* 2005; 26: 525–535.
 49. Law I, Iida H, Holm S, et al. Quantitation of regional cerebral blood flow corrected for partial volume effect using 0-15 water and PET: II. Normal values and gray matter blood flow response to visual activation. *J Cereb Blood Flow Metab* 2000; 20: 1252–1263.
 50. Kelly CJ, Karthikesalingam A, Suleyman M, et al. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med* 2019; 17: 195.
 51. Qayyum A, Qadir J, Bilal M, et al. Secure and robust machine learning for healthcare: a survey. *arXiv* 2020; 1–22.
 52. Thiagarajan JJ, Rajan D and Sattigeri P. Understanding behavior of clinical models under domain shifts. *arXiv* <http://arxiv.org/abs/1809.07806> (2018).
 53. Bremmer JP, van Berckel BNM, Persoon S, et al. Day-to-day test–retest variability of CBF, CMRO₂, and OEF measurements using dynamic 15O PET studies. *Mol Imaging Biol* 2011; 13: 759–768.
 54. Bokkers RP, Bremmer JP, van Berckel BN, et al. Arterial spin labeling perfusion MRI at multiple delay times: a correlative study with H 2 15 O positron emission tomography in patients with symptomatic carotid artery occlusion. *J Cereb Blood Flow Metab* 2010; 30: 222–229.