Short Acquisition Time PET/MR Pharmacokinetic Modelling Using CNNs

Catherine J. Scott[1(✉)], Jieqing Jiao[1], M. Jorge Cardoso[1,2], Kerstin Kläser[1], Andrew Melbourne[2,3], Pawel J. Markiewicz[1], Jonathan M. Schott[4], Brian F. Hutton[5], and Sébastien Ourselin[2]

[1] Centre for Medical Image Computing, University College London, London, UK
catherine.scott.14@ucl.ac.uk
[2] School of Biomedical Engineering and Imaging Sciences,
King's College London, London, UK
[3] Department of Medical Physics and Biomedical Engineering,
University College London, London, UK
[4] Dementia Research Centre, Institute of Neurology,
University College London, London, UK
[5] Institute of Nuclear Medicine, University College London, London, UK

Abstract. Standard quantification of Positron Emission Tomography (PET) data requires a long acquisition time to enable pharmacokinetic (PK) model fitting, however blood flow information from Arterial Spin Labelling (ASL) Magnetic Resonance Imaging (MRI) can be combined with simultaneous dynamic PET data to reduce the acquisition time. Due the difficulty of fitting a PK model to noisy PET data with limited time points, such 'fixed-R_1' techniques are constrained to a 30 min minimum acquisition, which is intolerable for many patients. In this work we apply a deep convolutional neural network (CNN) approach to combine the PET and MRI data. This permits shorter acquisition times as it avoids the noise sensitive voxelwise PK modelling and facilitates the full modelling of the relationship between blood flow and the dynamic PET data. This method is compared to three fixed-R_1 PK methods, and the clinically used standardised uptake value ratio (SUVR), using 60 min dynamic PET PK modelling as the gold standard. Testing on 11 subjects participating in a study of pre-clinical Alzheimer's Disease showed that, for 30 min acquisitions, all methods which combine the PET and MRI data have comparable performance, however at shorter acquisition times the CNN approach has a significantly lower mean square error (MSE) compared to fixed-R_1 PK modelling ($p = 0.001$). For both acquisition windows, SUVR had a significantly higher MSE than the CNN method ($p \leq 0.003$). This demonstrates that combining simultaneous PET and MRI data using a CNN can result in robust PET quantification within a scan time which is tolerable to patients with dementia.

1 Introduction

For the accurate quantification of tracer target density (BP_{ND}), such as amyloid-β burden in Alzheimer's disease, phamacokinetic (PK) modelling of dynamic

Positron Emission Tomography (PET) data requires the acquisition to cover the delivery, binding and washout of the injected radiotracer. This may take 60 min or more which is not clinically feasible, due to patient discomfort, scanner availability, and increased motion. A framework which reduces the PET acquisition time by incorporating simultaneously acquired arterial spin labelled (ASL) MRI data into the PK model has been proposed [1]. This involves three steps; conversion of ASL cerebral blood flow (CBF) maps into the relative PET tracer delivery parameter (R_1), extrapolation of the PET input function (C_R) to account for the missing time-points, and PK model fitting to the measured PET data using fixed R_1 and extrapolated C_R. We refer to this as the 'fixed R_1 method'. Unlike the clinically used standardised uptake value ratio (SUVR), this method can account for changes in blood flow, which can confound estimates of target density in longitudinal studies [2]. However, due the difficulty of fitting a PK model to noisy PET data with limited time points, this technique is constrained to a 30 min minimum acquisition time, which may still be intolerable for some patients.

The 'fixed-R_1' approach estimates R_1 from ASL-CBF independently from dynamic PET fitting for target density (BP_{ND}) and washout rate (k_2). This implementation cannot explicitly model the known influence of CBF on washout, due to high uncertainty in washout estimation, and the complex relationship which is dependent on the local tissue tracer kinetics [3]. Furthermore, the extrapolation of the input function, C_R, uses scaled population data under the assumption that tracer washout in this region is equal to the average population value. This assumption is violated in the case of disease or blood flow changes.

In this work we propose a deep learning (DL) framework to achieve PET quantification for a short acquisition time in a single step. We avoid the noise sensitive voxelwise PK curve fitting step, through the use of deep convolutional neural networks which enforce spatial regularisation across the receptive field. Our approach negates the need for explicit modelling between CBF, tracer delivery and tracer washout, as these relationships are learnt from the data and modelled in conjunction with the dynamic PET data. This approach also avoids C_R extrapolation, overcoming the limitation of a population tracer washout rate.

To our knowledge, this is the first time in which DL has been applied to PET PK modelling. This is due to the availability of robust models to describe standard data and the lack of one-to-one mapping between model parameters and dynamic PET data. However, the standard models are not sufficient to describe PET data with missing time-points. Furthermore, the incorporation of ASL-CBF constrains the parameter estimation. DL was chosen for its ability to model the underlying relationship between ASL-CBF and the delivery, binding and washout of the PET tracer without explicit feature extraction. By exploiting all of the PET and MRI information, and avoiding voxelwise fitting, this framework provides more robust estimates of target density with a shorter acquisition.

2   Methods

2.1   Deep Learning Framework for BP_{ND} Estimation

The framework performs regression of PET target density (BP_{ND}) from PET and MRI data directly. The network was implemented in NiftyNet [4] using the 'highresnet' convolutional neural network with 20 convolutional layers [5], which uses a stack of residual dilated convolutions with increasingly large dilation factors. For training we used adaptive moment estimation (Adam) with an initial learning rate of 10^{-3}, and a root mean square error loss function. The networks were initialised randomly and trained for a maximum of 50,000 iterations. The training patch size was $56 \times 56 \times 56$ voxels and a smoothed brain mask was used for adaptive sampling. Random rotation and scaling transformations of $\pm 10\%$ were used for training data augmentation. All inputs were 3D image volumes: the ASL-CBF maps, the structural T1 weighted MRI, and the dynamic PET data, which were entered as one frame per channel, see Fig. 1.

2.2   Gold Standard PK Modelling

The linearised simplified reference tissue model (SRTM) is used for gold standard PET quantification (1). Basis functions for $C_R(t) \otimes e^{-\theta t}$ are pre-calculated over a physiologically plausible range of θ [6], where $C_R(t)$ is the tracer concentration in the reference region. $C_R(t)$ is used as an input function since the reference region, cerebellar grey matter, is considered to be devoid of the imaging target. $C_T(t)$ is the measured tracer concentration in the target tissue. The model parameters are: R_1 (the delivery rate constant in the target tissue relative to reference tissue), k_2 (the transfer rate constant from target tissue to blood), and the parameter of interest BP_{ND} (the binding potential which is related to target density and consequently amyloid-β burden). The parameters are estimated via curve fitting to $C_R(t)$ and $C_T(t)$ acquired over $t = 0{:}60$ min.

$$C_T(t) = R_1 C_R(t) + \phi C_R(t) \otimes e^{-\theta t}$$
$$\text{where} \quad \phi = k_2 - R_1 k_2/(1 + BP_{ND}), \quad \theta = k_2/(1 + BP_{ND}) \tag{1}$$

2.3   Acquisition Window Definition

For gold standard PK modelling the scan starts at tracer injection, $t_s = 0$, with a duration of $t_d = 60$ min. However, for the short acquisition methods $t_s > 0$, and t_d is chosen to fit clinical requirements. We optimise the timing window, $t = t_s{:}t_s + t_d$, for each method at different t_d's. This was performed over $t = 30{:}60$ min, as this period is recommended for routine clinical scans using this tracer.
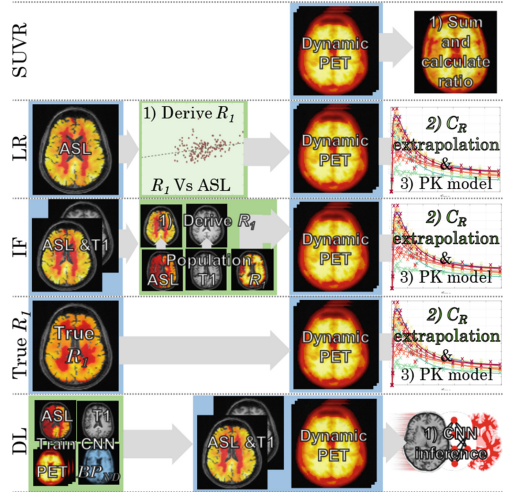
2.4 Comparison Methods for Short Acquisition PET Quantification

We compare the proposed technique to four short PET acquisition methods: three fixed-R_1 methods, and the clinical standard, SUVR.

Fixed R_1 Methods. Two methods are used to derive R_1 from ASL-CBF: the linear regression (LR) method [1], and the image fusion (IF) method [7]. Both methods require a database of subjects with 60 min of PET data and ASL.

The LR method performs linear regression between R_1 and ASL-CBF on the database and the relationship is applied to an unseen ASL-CBF map. For IF, the local similarity between the unseen ASL-CBF map and those in the database is used to weight the propagation of R_1 database values into the subject's space. An additional method using the gold standard R_1 (true R_1) is also included to demonstrate the upper limit of this approach, where R_1 is estimated perfectly from the ASL data.

For all three methods the estimation of BP_{ND} was carried out as previously described [1,7]. Briefly, the reference region, C_R is extrapolated to $t = 0$, at tracer injection, by scaling the mean population C_R

Fig. 1. Overview of methods tested. Blue boxes indicate input subject data and green boxes population data.

to the measured data using a linear least squares fit, then the derived R_1 value and the extrapolated C_R are used in (1) to estimate k_2 and BP_{ND} from the measured PET data.

Standardised Uptake Value Ratio (SUVR). SUVR is calculated by dividing the image (C_T) by the mean value in the reference region (C_R) to yield relative tracer uptake, which can not take blood flow into account. For comparison with BP_{ND}, one is subtracted, as $BP_{ND} \approx \frac{C_T}{C_R} - 1$. SUVR is calculated over different timing windows by first summing the relevant reconstructed frames.

2.5 Data Acquisition and Pre-processing

Database Construction. For each subject the T1 and ASL-CBF MR images were affinely registered into PET space. The subjects were randomly split between training (38, ~70%), validation (6, ~10%) and testing (11, ~20%). The input data used and an overview of each methodology is summarised in Fig. 1, where the dynamic PET data include the frames acquired over $t = t_s:t_s + t_d$.

PET Data. 60 min of PET data were acquired following intravenous injection of an amyloid-β targeting radiotracer, $[^{18}F]$florbetapir. Dynamic PET data were binned into $15s \times 4$, $30s \times 8$, $60s \times 9$, $180s \times 2$, $300s \times 8$ time frames, such that all frames for $t \geq 20$ min were 5 min long. The data were reconstructed into $2 \times 2 \times 2$ mm voxels, accounting for dead-time, attenuation (using synthetic CT), scatter, randoms and normalisation [8].
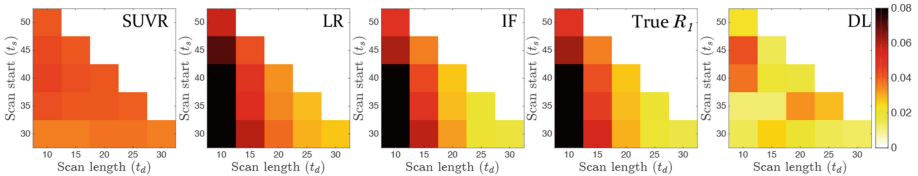
Fig. 2. MSE averaged across subjects for different timing windows and methods

ASL Data. Pseudo-continuous ASL data were acquired at $t = 55:60$ min, using a 3D GRASE readout at $3.75 \times 3.75 \times 4$ mm and reconstructed to $1.88 \times 1.88 \times 4$ mm voxels. 10 control-label pairs were acquired with a pulse duration and post labelling delay of 1800 ms. Proton density, S_0, was estimated by fitting saturation recovery images, at three recovery times (1, 2, 4s), for $[T1, S_0]$. Cerebral blood flow (CBF) maps were then estimated from the ASL and saturation recovery images [9]. The parameter values were 0.9 ml/g for the plasma/tissue partition coefficient, 1650 ms for blood T1, and 0.85 for labelling efficiency.
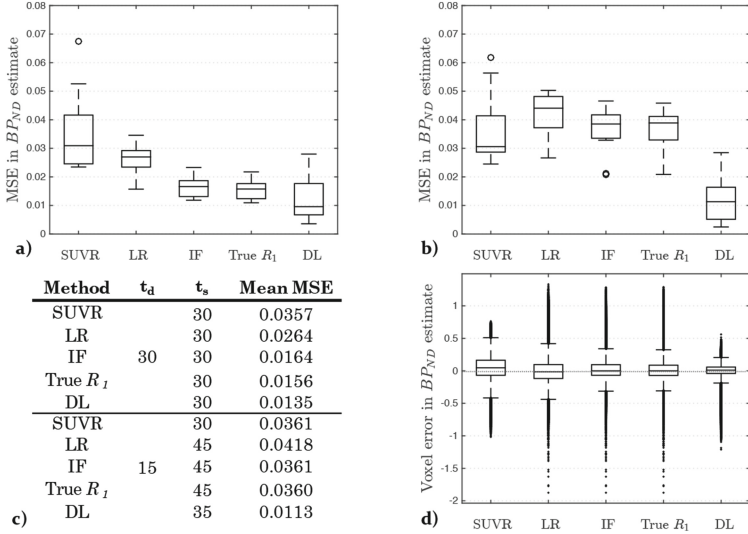
3   Experiments and Results

Data. Imaging data were collected from 55 cognitively normal subjects participating in Insight 46, a neuroimaging sub-study of the MRC National Survey of Health and Development [10], who underwent simultaneous PET and multimodal MRI on a Siemens Biograph mMR 3T PET/MRI scanner. 11 subjects were used for testing, with the remaining subjects used in the database for training and validation.

Validation. The proposed deep learning (DL) method was compared to three fixed-R_1 methods (LR, IF and true R_1), and SUVR. BP_{ND} estimation accuracy was assessed using the mean square error: MSE $= 1/v \sum_v (I_v^{est} - I_v^{GS})^2$, where I is intensity, v is the number of voxels, GS is the gold standard and est is the estimate. Statistical tests were performed using the Wilcoxon signed rank test.

3.1   Method Comparison over Different Timing Windows

Figure 2 shows the average MSE across subjects for different data acquisition windows for each method. Here, SUVR shows minimal influence from acquisition

timing and length due to the simplicity of the technique. However, since it is not able to account for tracer delivery or washout it has a consistently high MSE. For the three methods which perform kinetic modelling with a fixed R_1 there is a strong time dependence, with the error increasing greatly as the scan time is reduced. Consequently they outperform SUVR for 30 min acquisitions, but for scans of less than 20 min, where the number of datapoints is ≤ 4, they produce a higher error. This is due to the difficulty in fitting the PK model to a few noisy datapoints, and the increased uncertainty in the extrapolation of C_R.

Method	t_d	t_s	Mean MSE
SUVR		30	0.0357
LR		30	0.0264
IF	30	30	0.0164
True R_1		30	0.0156
DL		30	0.0135
SUVR		30	0.0361
LR		45	0.0418
IF	15	45	0.0361
True R_1		45	0.0360
DL		35	0.0113

Fig. 3. Subject MSE for (a) 30 and (b) 15 min scans, summarised in (c). The voxel-wise error for all subjects for a 15 min scan is shown in (d).

The deep learning based method (DL) shows a consistently low MSE across timing windows compared to the other techniques. This is because voxel based PK modelling, and C_R extrapolation, which are acquisition length dependent, are avoided. Furthermore, blood flow information is leveraged to inform both tracer delivery and washout, reducing the acquisition time required relative to the fixed-R_1 PK modelling techniques. For DL there is no clear trend to the acquisition window dependence which makes it more flexible for clinical implementation.

3.2 Optimised Timing Window Method Comparison

Based on the mean MSE, shown in Fig. 2, the best timing window for each method was selected for a 30 min (6 frames) and 15 min (3 frames) scan, representing a long clinical scan and a tolerable scan duration respectively, see Fig. 3c.

30 min Optimised Acquisition. Figure 3(a) shows the MSE across subjects for the 30 min acquisition window. As expected the fixed-R_1 methods have a lower MSE than SUVR due to the more accurate modelling of tracer delivery and washout. For this acquisition length the benefit of using the deep learning approach is limited compared to the fixed R_1 methods, and the difference in MSE did not reach statistical significance. However, DL has a significantly lower MSE compared to SUVR ($p = 0.003$). Figure 4(a) shows an example subject, which highlights the good performance of the IF, true R_1 and DL methods, while SUVR shows a large over estimation. The LR method shows corruption due to artefacts in the ASL-CBF map which propagate directly into the BP_{ND} estimation.

15 min Optimised Acquisition. When the scan time is reduced to 15 min the MSE in the fixed-R_1 methods increases, even when using the true R_1 parameter. By contrast, the DL and SUVR methods maintain their performance levels. Now DL has a significantly lower MSE than both the fixed-R_1 methods ($p \leq 0.001$) and SUVR ($p = 0.001$). The DL method also has a lower bias than all other methods, see Fig. 3(d), but this does not reach statistical significance.

Figure 4(b) shows the estimated BP_{ND} images using a 15 min acquisition for the different methods for an example subject. Here, the noise in the fixed-R_1 methods is a result of the limited timepoints for the fit. SUVR gives a plausible estimate of BP_{ND}, however the image demonstrates a general overestimation of the target density compared to the true image. By contrast, the DL technique yields a low noise image due to the spatial regularisation inherent in the technique, with high accuracy as the model is able to combine the dynamic PET data with the blood flow information from the ASL to accurately estimate BP_{ND}.
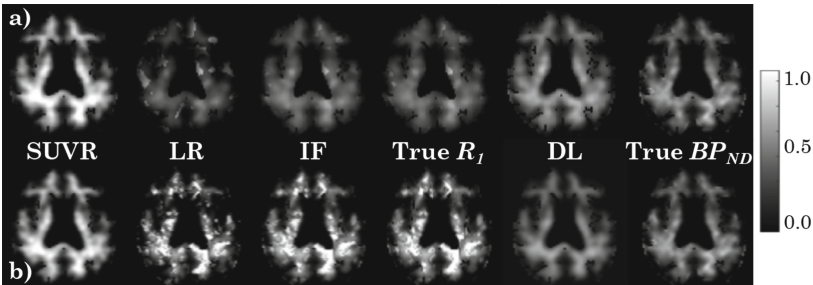
Fig. 4. Example subject BP_{ND} maps optimised for (a) 30 and (b) 15 min acquisition, where the true BP_{ND} is calculated over 60 min for both.

4   Discussion and Conclusion

In this paper we present a deep learning approach to PET target density estimation, by combining dynamic PET data with MRI blood flow and structural

images to significantly reduce the acquisition time to just 15 min, compared to the gold standard 60 min. This is applied to amyloid PET data which is used in the diagnosis and monitoring of Alzheimer's disease, as the symptoms of the disease necessitate short scans. This method was compared to the clinical standard, SUVR, as well as previously proposed techniques which fix the tracer delivery parameter R_1 using MRI blood flow data in the PET PK modelling to reduce the acquisition time. This demonstrated that, for a 30 min acquisition, the proposed technique performed comparably to the previously proposed fixed-R_1 techniques, and significantly better than SUVR ($p = 0.003$). When the acquisition window was reduced to 15 min, the fixed-R_1 methods had insufficient data to fit the PK model. However, the deep learning method maintained its low MSE, which was significantly lower than the clinically used SUVR ($p = 0.001$).

This initial work proves the benefit of using deep learning to perform PET quantification where limited PET data means that the standard model fails. In the future we intend to build on this approach by explicitly encoding the PET frame timing information into the model. This would not only give the model more information, but also the potential to cope with discontinuous scans.

References

1. Scott, C.J., et al.: ASL-incorporated pharmacokinetic modelling of PET data with reduced acquisition time: application to amyloid imaging. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9902, pp. 406–413. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46726-9_47

2. van Berckel, B.N.M., Ossenkoppele, R., Tolboom, N., et al.: Longitudinal amyloid imaging using 11C-PiB: methodologic considerations. J. Cereb. Blood Flow Metab. **54**(9), 1570–6 (2013)

3. Wu, Y., Carson, R.E.: Noise reduction in the simplified reference tissue model for neuroreceptor functional imaging. J. Cereb. Blood Flow Metab. **22**(12), 1440–1452 (2002)

4. Gibson, E., Li, W., Sudre, C., et al.: Niftynet: a deep-learning platform for medical imaging. arXiv preprint arXiv:1709.03485 (2017)

5. Li, W., Wang, G., Fidon, L., et al.: On the compactness, efficiency, and representation of 3D convolutional networks: brain parcellation as a pretext task. In: Information Processing in Medical Imaging (2017)

6. Gunn, R.N., Lammertsma, A.A., Hume, S.P., et al.: Parametric imaging of ligand-receptor binding in PET using a simplified reference region model. NeuroImage **6**(4), 279–287 (1997)
7. Scott, C., et al.: Short acquisition time PET quantification using MRI-based pharmacokinetic parameter synthesis. In: Descoteaux, M., et al. (eds.) MICCAI 2017. LNCS, vol. 10434, pp. 737–744. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66185-8_83
8. Markiewicz, P.J., Ehrhardt, M.J., Erlandsson, K., et al.: NiftyPET: a High-throughput Software Platform for High Quantitative Accuracy and Precision PET Imaging and Analysis. Neuroinformatics **16**(1), 95–115 (2017)
9. Melbourne, A., Toussaint, N., Owen, D., et al.: Niftyfit: a software package for multi-parametric model-fitting of 4D magnetic resonance imaging data. Neuroinformatics **14**(3), 319–337 (2016)
10. Lane, C.A., Parker, T.D., Cash, D.M., et al.: Study protocol: Insight 46 - a neuroscience sub-study of the MRC National Survey of Health and Development. BMC Neurol. **17**(1), 75 (2017)