Physica Medica 99 (2022) 85–93

Contents lists available at ScienceDirect

Physica Medica

journal homepage: www.elsevier.com/locate/ejmp

Validation of deep learning-based nonspecific estimates for amyloid burden quantification with longitudinal data

Ying-Hwey Nai [a,2,*], Haohui Liu [b,3], Anthonin Reilhac [a,4], for the Alzheimer's Disease Neuroimaging Initiative[1]

[a] Clinical Imaging Research Centre, Yong Loo Lin School of Medicine, National University of Singapore, Singapore
[b] Carnegie Mellon University, Pennsylvania, United States

ARTICLE INFO

Keywords:
Alzheimer's disease
Amyloid-Beta Positron Emission Tomography (Aβ-PET)
Deep learning
Nonspecific uptake

ABSTRACT

Purpose: To validate our previously proposed method of quantifying amyloid-beta (Aβ) load using nonspecific (NS) estimates generated with convolutional neural networks (CNNs) using [^{18}F]Florbetapir scans from longitudinal and multicenter ADNI data.

Methods: 188 paired MR (T1-weighted and T2-weighted) and PET images were downloaded from the ADNI3 dataset, of which 49 subjects had 2 time-point scans. 40 Aβ- subjects with low specific uptake were selected for training. Multimodal ScaleNet (SN) and monomodal HighRes3DNet (HRN), using either T1-weighted or T2-weighted MR images as inputs) were trained to map structural MR to NS-PET images. The optimized SN and HRN networks were used to estimate the NS for all scans and then subtracted from SUVr images to determine the specific amyloid load ($SA\beta_L$) images. The association of $SA\beta_L$ with various cognitive and functional test scores was evaluated using Spearman analysis, as well as the differences in $SA\beta_L$ with cognitive test scores for 49 subjects with 2 time-point scans and sensitivity analysis.

Results: $SA\beta_L$ derived from both SN and HRN showed higher association with memory-related cognitive test scores compared to SUVr. However, for longitudinal scans, only $SA\beta_L$ estimated from multimodal SN consistently performed better than SUVr for all memory-related cognitive test scores.

Conclusions: Our proposed method of quantifying Aβ load using NS estimated from CNN correlated better than SUVr with cognitive decline for both static and longitudinal data, and was able to estimate NS of [^{18}F]Florbetapir. We suggest employing multimodal networks with both T1-weighted and T2-weighted MR images for better NS estimation.

Introduction

Dementia is the fifth leading cause of death worldwide with increasing mortality rates for all age groups due to the ageing population [1]. In the US, deaths from Alzheimer's Disease (AD) has increased by 146.2% between 2000 and 2008, with an estimated US$305 billion of healthcare burden and 18.6 billion hours of care placed on family members and caregivers. The increasing burden of AD means that methods for early detection and prevention or delay of AD are important [2]. In June 2021, the food and drug administration (FDA, US) has approved Aducanumab (Aduhelm, Biogen Inc. Cambridge, Massachusetts US), an amyloid-beta (Aβ)-directed antibody, to treat AD [3]. Measuring amyloid plaque burden is a key capability of positron emission tomography (PET), with Aβ-targeting radiotracers, and will play a

* Corresponding author at: Clinical Imaging Research Centre, Yong Loo Lin School of Medicine, National University of Singapore, Centre for Translational Medicine (MD6), 14 Medical Drive, #B1-01, Singapore 117599, Singapore.
E-mail addresses: mednyh@nus.edu.sg, yinghweynai@yahoo.com (Y.-H. Nai).

[1] Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

[2] ORCID: 0000-0002-6634-8785.
[3] ORCID: 0000-0002-8172-6502.
[4] ORCID: 0000-0002-9997-9540.

https://doi.org/10.1016/j.ejmp.2022.05.016
Received 29 October 2021; Received in revised form 26 May 2022; Accepted 27 May 2022
Available online 2 June 2022

central role in identifying patients and monitoring treatment efficiency.

The common method to quantify the Aβ burden from static PET scans is the standardized uptake value ratio (SUVr). This Aβ-biomarker is derived by normalizing the Aβ-PET tracer uptake in AD-targeting regions with a reference region known to be devoid of Aβ [4]. Target and reference regions are usually derived from the structural magnetic resonance (MR) images of the subject or using generic ROIs derived using linear regression [5], principal component analysis (PCA) [6,7] or deep learning (DL) [8]. A cut-point is then computed and used to classify subjects into Aβ+/-. This cut-point strongly depends on the data samples, the cohort characteristics, Aβ-PET radiotracers, image acquisition PET systems and methods employed to derive the cut-points [9]. For these reasons, cut-points are hardly comparable between Aβ-PET studies. To overcome the issues of applying different Aβ-positivity cut-points for different Aβ-PET radiotracers and scanners, the Centiloid Working Group developed the Centiloid scale to map the measured SUVr values to a standard 0 to 100 scale but it requires normalization of images to Montreal neurological institute (MNI) space [10].

To overcome the reliance on MR images to derive SUVr, several authors have proposed to derive SUVr without structural MR images by creating an adaptive template in MNI space using linear regression [5], principal component analysis (PCA) [6] or deep learning (DL) [8]. Whittington et al. proposed a new biomarker $A\beta_L$ to quantify Aβ burden by modelling SUVr as a linear combination of nonspecific (NS) and specific binding (combination of maximum carrying capacity and Aβ load) [11]. Other authors relied on PCA to create adaptive templates in MNI space to quantify Aβ burden in terms of Aβ index [12], $SA\beta_L$ [7] and AMYQ index [13]. However, all these methods require the transformation of PET images to MNI space.

Convolutional neural networks (CNNs) showed potential in various medical imaging applications [14–17], such as improving image quantification by correcting for attenuation of PET images [15,17,18] or recovering image quality from low-quality or low-dose PET or CT images [14,15,19]. We have previously proposed a new method of quantifying Aβ burden by using CNN, or more specifically DL, to estimate the NS uptake from structural MR images and calculating specific Aβ load ($SA\beta_L$) by subtracting NS from SUVr images. This method does not require the transformation of PET images to MNI space and the new biomarker, $SA\beta_L$ showed an increased association with cognitive and functional test scores by up to 67% compared to SUVr using [^{11}C]PiB [4]. However, different PET radiotracers exhibit different NS binding to myelin and accuracy may be dependent on the quality of the MR images. Moreover, it has not been validated for longitudinal data. As such, we further validated the method using [^{18}F]Florbetapir and multicentre study data with longitudinal scans. Sensitivity analysis was also performed to determine the effect of the choice of inputs on the NS estimated from CNN.

Materials and methods

Image selection and processing

Data used in this study were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). T1-weighted Magnetization Prepared - RApid Gradient Echo (MPRAGE) and T2-weighted Fluid-Attenuated Inversion Recovery (FLAIR) MR images and the corresponding [^{18}F]Florbetapir PET images, acquired within one year from the MR images were downloaded from the ADNI3 database (2016 to 2021). All images were first visually assessed to ensure consistency in textural information across scans, particularly between time-points scans of the same subjects. MR images with obvious inhomogeneity, image artefacts, and failed segmentation, and PET images with motion artifacts or containing too much noise were removed from the study (Refer to Supplementary Fig. 1 for some examples). A total of 188 scans (139 subjects) with 35 ± 53 days between PET and MR scans, were selected for this study and the subjects' demographics are shown in Table 1. Among the 139 subjects, 49 subjects had 2 scans, acquired 2.1 ± 0.3 years apart.

The PET images were post-processed to a standard matrix of 160 × 160 × 96 with a voxel size of 1.5 mm^3, with intensity normalization and post-filtering to ensure uniform resolution. Both T1-weighted MPRAGE and T2-weighted FLAIR MR images were registered to the PET images, and segmented to obtain the brain masks and regions of interest (ROIs) using statistical parametric mapping (SPM12, https://www.fil.ion.ucl.ac.uk/spm/). Intensity normalization was then performed on both MR images using subject-specific brain mask such that the average of voxels within the mask is exactly one. The PET images were then converted to standardized uptake value ratio (SUVr) images using the whole cerebellum as a reference region. The ROI mask included the main AD-specific regions of the frontal lobe, parietal lobe, temporal lobe, and precuneus and was used to derive the global mean SUVr value in the whole brain (WB) and gray matter (GM).

Clinical diagnosis, neuropsychological and functional assessments

The clinical diagnosis was assessed at each visit: 101 cognitive normal (CN), 17 significant memory concern (SMC), 31 early mild cognitive impairment (EMCI), 29 MCI, 8 late MCI (LMCI) and 2 AD, with no change in clinical diagnosis on the second scan for all 49 subjects with 2 scans (Table 1). A total of 17 key variables were extracted from the 7 neuropsychological and functional assessment tests: (1) Alzheimer's Disease Assessment Scale (ADAS): **TOTSCORE** (11 items score) and **TOTAL13** (13 items score), (2) Clinical Dementia Rating Scale (CDR): **CDMEMORY** (memory), **CDORIENT** (orientation), **CDJUDGE** (judgement & problem solving), **CDCOMMUN** (community affairs), **CDHOME** (home and hobbies), **CDCARE** (personal care), **CDGLOBAL** (Global CDR), and **CDSOB** (CDR sum of boxes), (3) Functional Activities Questionnaires (FAQ): **FAQTOTAL**, (4) Mini-Mental State Examination (MMSE): **MMSCORE**, (5) Montreal Cognitive Assessment (MoCA): **MoCA**, (6) Neuropsychological Battery: **LIMMTOTAL** (immediate recall total score) and **LDELTOTAL** (delayed recall total), and (7) Neuropsychological Summary Scores: **ADNI_MEM** (Memory function composite score) and **ADNI_EF** (Executive function composite score).

Nonspecific Aβ uptake image estimation using deep learning

Our previous results showed that multimodal ScaleNet [20] performed best in estimating the NS uptake from both T1-weighted and T2-weighted FLAIR images, compared to monomodal HighResNet [21] and conditional Generative adversarial networks (cGAN) [2]. In this study,

Table 1
Subject clinical diagnosis distribution, age and their MMSCORE (Average ± Stdev) for all scans, and 49 subjects at baseline and second scans. CN = Cognitive normal, SMC = Significant Memory Concern, MCI = Mild cognitive impaired, EMCI = Early MCI, LMCI = Late MCI, AD = Alzheimer's disease.

	CN	SMC	EMCI	MCI	LMCI	AD
All Scans (n)	101	17	31	8	29	2
Age (years)	73.1 ± 6.9	73.2 ± 3.9	71.1 ± 4.8	72.1 ± 8.5	69.9 ± 8.2	80.5 ± 2.1
MMSE	29.0 ± 1.3	29.4 ± 0.8	28.8 ± 1.9	28.8 ± 2.0	27.6 ± 1.7	21.0 ± 0.0
1 TP (n)	26	6	10	4	3	0
Age (years)	73.2 ± 6.5	71.8 ± 4.7	68.9 ± 4.1	74.8 ± 12.1	69.3 ± 10.4	–
MMSE	28.2 ± 1.1	29.0 ± 1.1	28.2 ± 2.3	30.0 ± 0.0	28.0 ± 1.7	–
2 TP (n)	26	6	10	4	3	0
Age (years)	75.4 ± 6.5	74.0 ± 4.3	70.9 ± 4.1	76.8 ± 12.1	71.3 ± 10.4	–
MMSE	29.4 ± 0.9	29.5 ± 0.5	28.7 ± 2.3	29.8 ± 0.5	27.3 ± 2.5	–

we continued to use both ScaleNet (SN) and HighResNet (HRN) networks, which can be easily implemented on the NiftyNet platform (Version 0.5.0, https://niftynet.io/) [22]. The optimized configurations of ScaleNet and HighResNet networks were taken from our previous study with adjustment to the spatial window sampling size due to differences in image matrix size (Refer to Supplementary Table 2) [4]. HighRes3DNet was trained independently using either T1-weighted MPRAGE or T2-weighted FLAIR images, while multimodal ScaleNet was trained using both MR images to map the structural MR images to NS-PET images. The subject-specific brain masks were also input during training to provide weighting to the networks. Subject-specific SAβL image was then derived by subtracting the estimated NS from SUVr PET image (Fig. 1).

All 188 scans, including second time-point scans, were ranked based on the global mean SUVr, with stronger emphasis in GM, as the amyloid accumulation in cortical GM is more important in the diagnosing patients or in quantifying their amyloid burden. The network should be trained using preferably subjects with no /low amyloid burden (very low SUVr values) to estimate the nonspecific uptake in subjects. Therefore, 40 subjects with the lowest global mean SUVr in GM were selected for training and the next 20 were selected for network validation. These 60 scans were visually assessed as Aβ- by a research fellow with 3 years of experience in interpreting Aβ-PET scans. However, subjects with substantial differences in global mean SUVr in GM between the 2 time-points were excluded from the training dataset, but included in the validation dataset. The optimised HighResNet networks, with either T1-weighted MPRAGE or T2-weighted FLAIR MR images as input, were compared and the best HighRes3DNet and ScaleNet networks were then used to generate the NS images for all 188 scans.

Network performance evaluation

For direct comparison with our previous results, the same three metrics were used to compare the performance of the models in estimating the NS uptakes in the WB and cortical GM of 20 evaluation scans: mean squared error (MSE), absolute mean relative error (MRE) and structural similarity (SSIM) [2,4]:

$$MSE(X,Y) = \frac{1}{n}\sum_{i=1}^{n}(X_i - Y_i)^2 \tag{1}$$

$$MRE(\%) = \frac{\left|\frac{1}{n}\sum_{i=1}^{n} Y_i - \frac{1}{n}\sum_{i=1}^{n} X_i\right|}{\frac{1}{n}\sum_{i=1}^{n} Y_i} \times 100 \tag{2}$$

$$SSIM(X,Y) = \frac{(2\mu_x\mu_y + c_1) \bullet (2\sigma_{xy} + c_2)}{\left(\mu_x^2 + \mu_y^2 + c_1\right) \bullet \left(\sigma_x^2 + \sigma_y^2 + c_2\right)} \tag{3}$$

where X and Y refer to the estimated image and ground truth (Y), and μ and σ refer to mean and standard deviation. Two constant parameters, c_1 $(=0.01L)^2$ and c_2 $(=0.03L)^2$ are included to avoid division with a very small denominator; L is the dynamic range of the pixel values based on the image class, for example the default dynamic range is 255 for images of data type uint8.

Evaluation of Aβ-PET quantification

The degree of association of global mean $SA\beta_L$, derived from HighResNet and ScaleNet, and SUVr, with 17 key neuropsychological and functional assessment variables, was evaluated using Spearman's correlation analysis. We hypothesized that the proposed biomarker, $SA\beta_L$, would have stronger association with cognitive tests than SUVr (with contamination of NS), especially in subjects with memory concerns and/or cognitive impairment. Hence, association with cognitive tests was performed using all subjects, including CN (n = 188), and subjects with memory concerns and/or cognitive impairment (n = 87). To investigate the use of SAβL estimated using CNN in quantifying Aβ burden for

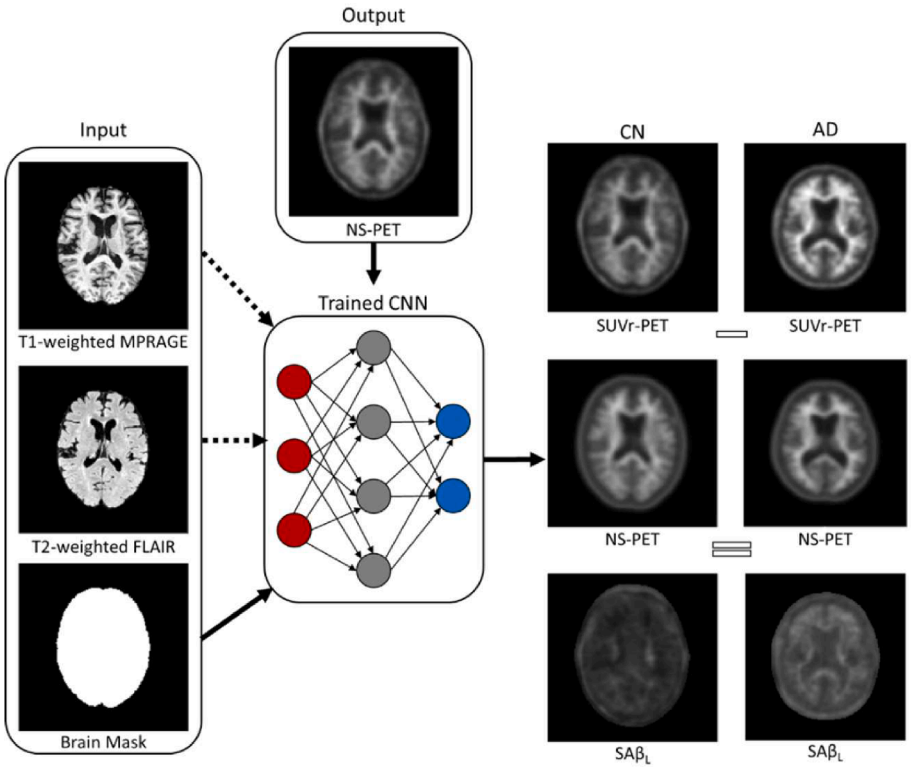

Fig. 1. Overview of NS estimation and $SA\beta_L$ derivation workflow: T1-weighted MPRAGE or T2-weighted FLAIR MR images or both are trained to map to NS-PET image with subject-specific brain mask as weighting using convolutional networks (ScaleNet or HighResNet) using supervised training. The trained network is then used to estimate NS-PET from subject-specific MR images. $SA\beta_L$ image is then derived by subtracting the NS-PET from SUVr PET image.

longitudinal studies, the differences in SAβL (HRN), SAβL (SN) and SUVr were evaluated using Spearman's correlation analysis with the differences in cognitive test scores between the 2 time-points scans.

Sensitivity analysis of methodology

Sensitivity analysis was performed by training the optimized networks of HighResNet (T1-weighted MPRAGE as input) and ScaleNet 5 times and randomly replacing 5 subjects from the training dataset with that in the validation dataset. The variations in NS measurements and the changes in association with the 17 neuropsychological and functional variables were then investigated.

Results

Network evaluation

Table 2 shows the performance evaluation of HighResNet, with T1-weighted MPRAGE and T2 weighted FLAIR images used independently as input, and ScaleNet, with both MR images as inputs. Performance are reported using MSE, SSIM and MRE averaged over the 20 validation scans (Refer to Supplementary Table 2 for that of 40 training scans). HighResNet (T1), yielded the highest SSIM and lowest MSE, while HighResNet (T2) yielded the lowest MRE, but performed the poorest in terms of SSIM and MSE. ScaleNet, with both MR images as input, ranked second in terms of MSE and SSIM, the last for MRE, thus showing possibilities of conflicting information between T1-weighted MPRAGE and T2 weighted FLAIR images.

Fig. 2 shows the NS in GM measured from SN (top), HRN with MPRAGE as input (middle) and HRN with FLAIR as input (bottom), with the corresponding SUVr in GM for all 188 scans. A dip in the SUVr was observed at the dashed line separating the scans used for training and evaluation. This was due to the selection criterion where subjects with big differences in global mean SUVr between the 2 time-points were excluded from the training dataset, but included in the validation dataset. The cut-point of global mean SUVr in GM, based on ADNI protocol is 1.11, derived using the target regions of frontal, anterior/posterior cingulate, lateral parietal, lateral temporal and whole cerebellum as reference region [23]. The range of global mean SUVr in GM of scans used for training is 0.923 to 1.136, with only 5 out of 40 scans with SUVr slightly $>$ 1.11, while that for validation is 0.945 to 1.166, with only 4 out of 20 scans with SUVr slightly $>$ 1.11. Although we included a few subjects with SUVr values greater than the cut-point, our SUVr values were obtained using target regions of the frontal lobe, parietal lobe, temporal lobe, and precuneus and whole cerebellum as the reference region. Moreover, all PET scans used for training were visually assessed as Aβ- so there were only a few scans with NS greater than the SUVr uptakes, mainly scans used for training and validation, which were assumed to contain no or little Aβ. Similar to our previous results, NS estimates in GM obtained by ScaleNet and HighResNet for the unseen scans showed the same levels and magnitude of intersubject variations as with the training scans [4]. This demonstrated the feasibility and generalization of the models to estimate NS from structural MR images for both [^{11}C]PiB and [^{18}F]Florbetapir.

Table 2
Comparison of network performance of HighResNet, with T1 and T2 images as inputs, and ScaleNet. (mean ± stdev) using Mean Square Error (MSE), Structural Similarity (SSIM) and Mean Relative Error in SUVr (MRE) within the whole brain (WB) and gray matter (GM) masks.

Metric	MSE		SSIM		MRE (%)	
	WB	GM	WB	GM	WB	GM
HighResNet (T1)	0.034 ± 0.016	0.027 ± 0.012	0.826 ± 0.041	0.824 ± 0.034	5.492 ± 4.182	5.598 ± 3.999
HighResNet (T2)	0.044 ± 0.023	0.035 ± 0.013	0.788 ± 0.054	0.793 ± 0.035	5.353 ± 4.848	4.746 ± 4.091
ScaleNet	0.041 ± 0.027	0.029 ± 0.015	0.806 ± 0.041	0.808 ± 0.037	6.524 ± 5.472	5.308 ± 5.049

Although HighResNet (T2) yielded the lowest MRE in both WB and GM, it overestimated the NS for all scans, particularly those used for training (Fig. 2). ScaleNet performed the best in our previous study [4], thus HighResNet (T1) and ScaleNet were selected for further evaluation. The distributions of SAβ_L (SN), SAβ_L (HRN) and SUVr across the different clinical diagnosis groups can be found in Supplementary Fig. 2.

Strong Pearson correlations of 0.97 and 0.96 were obtained between SAβ_L (HRN) and SAβ_L (SN) with SUVr for all 188 scans (Fig. 3**a**). Similarly, strong correlations of 0.97 and 0.92 were obtained between the differences in SAβ_L (HRN) and SAβ_L (SN) with the differences in SUVr values of 2 time-point scans of 49 subjects (Fig. 3**b**).

Association of PET biomarkers with cognition decline and neurodegeneration

SAβ_L (HRN) and SAβ_L (SN) showed higher association than SUVr for ADNI-MEM, LIMMTOTAL, MOCA, MMSCORE, CDJUDGE, CDORIENT, TOTAL13 and TOTSCORE for scans of subjects with memory concerns and/or cognitive impairment (Fig. 4). SAβ_L (HRN) generally yielded a higher association than SAβ_L (SN) except for LIMMTOTAL and CDJUDGE. Poorer associations of SAβ_L (HRN) and SAβ_L (SN) than SUVr were observed with the executive functions-related cognitive test scores, namely ADNI_EF, FAQTOTAL, CDHOME and CDCOMMUN. The association ($|\rho|$) and confidence (P-values) of our proposed biomarker SAβL was higher than that of SUVr, and also increased when we focused on subjects with memory concerns and/or cognitive impairment (Fig. 4). Lower association was observed using the whole study cohort of 188 scans for SAβ_L and SUVr (see Supplementary Fig. 3). This was expected as amyloid load accumulation is linked to cognition decline. SAβL showed higher association than SUVr with ADNI-MEM, LIMMTOTAL, MOCA, MMSCORE, CDJUDGE, CDORIENT, TOTAL13, TOTSCORE, which better reflects early-AD symptoms of memory/cognitive decline.

Association of PET biomarkers with cognition decline and neurodegeneration for longitudinal data

To investigate the use of SAβ_L estimated using CNN in measuring Aβ burden changes in longitudinal studies, the differences in SAβ_L (HRN), SAβ_L (SN) and SUVr were compared with the differences in cognitive test scores between the 2 time-points scans as shown in Fig. 5. In general, SAβ_L (SN) showed higher association than SUVr with cognitive decline as measured by cognitive test scores of ADNI_EF, ADNI_MEM, LDELTOTAL, FAQTOTAL, CDSOB, CDORIENT, CDMEMORY, TOTAL13 and TOTSCORE. However, SAβ_L (HRN) showed higher association than SUVr for only a few variables, namely LDELTOTAL, CDGLOBAL, CDJUDGE, CDMEMORY, and TOTSCORE. This may be due to the added information from T2-weighted FLAIR images in estimating NS. The better biomarker should have higher association between changes in the biomarker and changes in cognitive assessments scores. However, the association of SAβ_L and SUVr with cognition decline and neurodegeneration appeared poor, which may be due to the larger number of cognitive normal subjects (53%) with 2 time-point scans and these subjects had slightly improved MMSCORE (Table 1). Only SAβ_L (SN) yielded higher associations ($|\rho| > 0.2$) and confidence (P-values < 0.2) with ADNI_EF, FAQTOTAL and TOTAL13.

Sensitivity analysis

The error bars (top and middle) and errors (bottom) in Fig. 6 shows the variations in NS of 5 repeated measurements in sensitivity analysis and the original estimates. ScaleNet generally yielded higher variations

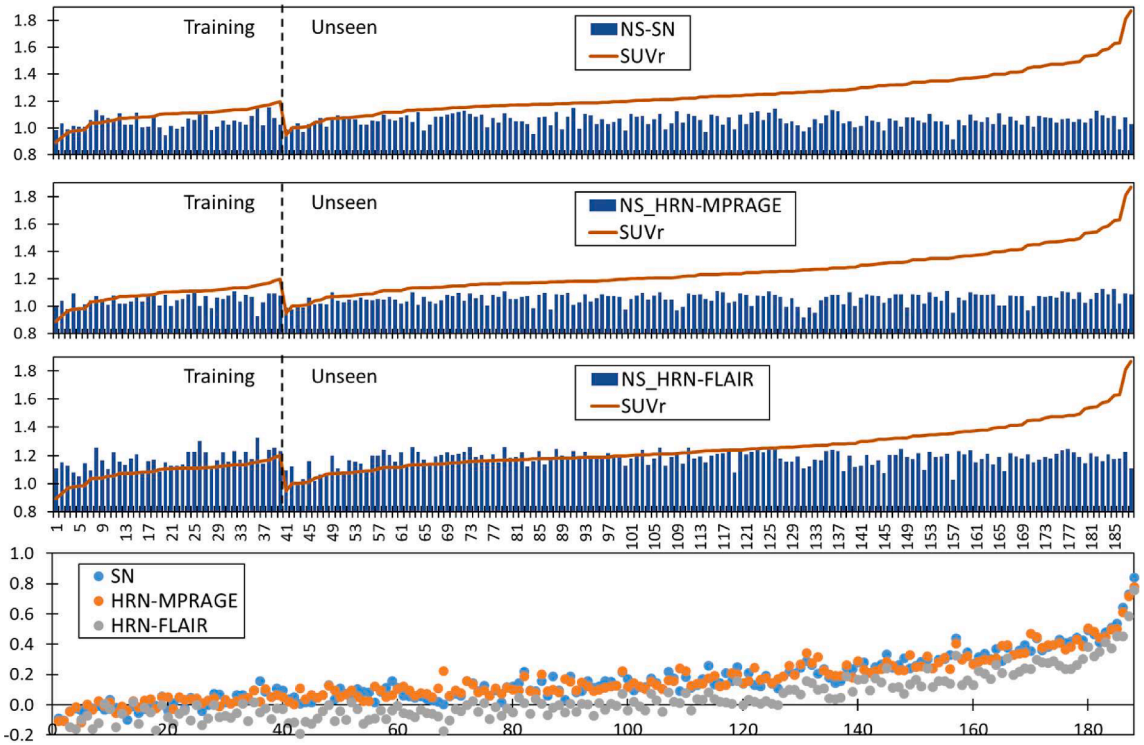

Fig. 2. Nonspecific uptake (NS) estimated using ScaleNet (SN) and HighResNet (HRN) with T1-weighted MPRAGE and T2-weighted FLAIR as inputs (Top, middle and bottom bar charts), for the whole cohort of 188 scans (40 training and 148 unseen). The red line gives the corresponding SUVr values measured from the actual PET scans. Subjects were ordered on the X-axis, on the left if used for training and otherwise on the right of the dashed line, and with increasing SUVr values. Scatter Plot shows the specific Aβ load ($SA\beta_L$) computed for the whole cohort using the NS estimates generated by ScaleNet and HighResNet. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

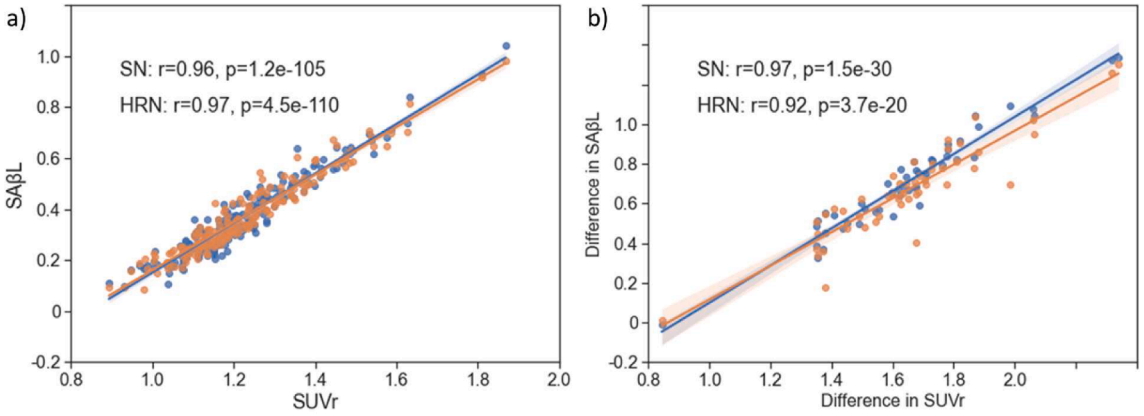

Fig. 3. Pearson correlation of (a) $SA\beta_L$ and SUVr of all 188 scans and (b) differences in SAβL and SUVr between the 2 time-point scans of 49 subjects.

but without any major outliers, as observed for HighResNet demonstrating that ScaleNet is more consistent in estimating NS due to additional information from T2-weighted FLAIR. However, the error was generally small (<0.08) even though greater amount of NS-overestimation was observed for both ScaleNet and HighResNet.

Fig. 7 shows the average association and the standard deviation of 5 repeated measurements in sensitivity analysis, as well as the original measurement for 87 subjects with memory concerns and/or cognitive impairment (For the full cohort, refer to Supplementary Fig. 3). Similar to the original results (Fig. 4), both $SA\beta_L$ (HRN) and $SA\beta_L$ (SN) yielded higher association than SUVr with ADNI-MEM, LIMMTOTAL, MOCA, MMSCORE, CDJUDGE, CDORIENT, TOTAL13 and TOTSCORE, with higher variations generally observed for $SA\beta_L$ (HRN) than $SA\beta_L$ (SN). Similarly, similar trend was observed for the whole cohort with $SA\beta_L$ (HRN) showing greater variations (Supplementary Fig. 3 vs. 4). For sensitivity analysis, the better biomarker should have generally higher association and confidence than SUVr, despite the changes in the training data. Higher associations were observed for the changes in $SA\beta_L$ (HRN) and $SA\beta_L$ (SN) for 49 subjects with 2 time-points scans with the 17 variables measured from neuropsychological and functional assessment tests, but with greater variations observed for $SA\beta_L$ (SN) than $SA\beta_L$ (HRN) (Supplementary Fig. 5).

Discussion

In this study, we validated our previously proposed method of quantifying Aβ burden with NS estimated using DL with multicentre and longitudinal [^{18}F]Florbetapir PET scans. Compared to our previous

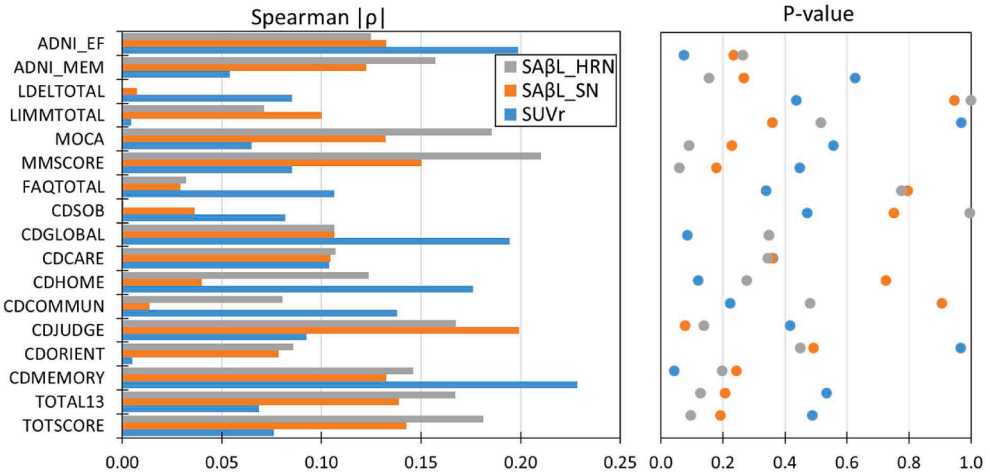

Fig. 4. Association (Spearman's ρ) and confidence (p-values) of 3 PET biomarkers of the brain Aβ burden with cognition and neurodegeneration for 87 subjects with memory concerns and/or cognitive impairment.

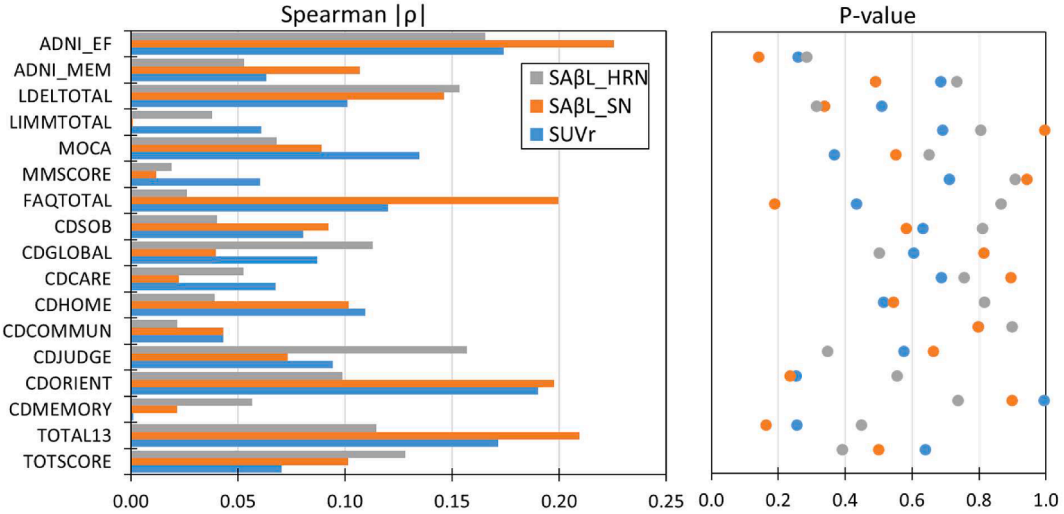

Fig. 5. Association (Spearman's ρ) and confidence (p-values) of changes in 3 PET biomarkers of the brain Aβ burden with changes in neuropsychological and functional assessments scores for 49 subjects with 2 time-points scans.

results, the association of $SA\beta_L$ with cognitive and functional test scores was lower but still better than SUVr, particularly for CDORIENT, CDJUDGE, TOTSCORE and TOTAL13 test scores for different cohorts (subjects with memory concerns and/or cognitive impairment only or including normal controls) and longitudinal data, and including MMSCORE and MoCA for the different cohorts. CDORIENT was shown to be an excellent predictor for the progression from MCI to AD [24] and MoCA reflects cognitive reserve better than MMSCORE and was said to be a better variable in assessing early cognitive decline [25]. This showed that $SA\beta_L$ showed good correlation with memory decline and is a better predictor of early stages than SUVr. The lower association obtained in this study with [^{18}F]Florbetapir than our previous study using [^{11}C]PiB may be due to lower binding affinity of [^{18}F]Florbetapir to Aβ40/42, the use of multicenter data or the study distribution, with about 50% cognitive normal subjects (Table 1). However, our results showed that our proposed biomarker $SA\beta_L$ yielded higher association than SUVr with cognitive scores, particularly those associated with early-AD symptoms of memory/cognitive decline. Thus, the proposed method can be applied to [^{18}F]Florbetapir and other fluorinated Aβ-PET radiotracers, though the network has to be trained for the individual Aβ-PET radiotracers independently.

Our results showed that monomodal HighResNet did not correlate as well as multimodal ScaleNet in estimating the NS for the longitudinal data. Similar to our previous results, ScaleNet yielded the best associations with cognitive and functional test scores for [^{11}C]PiB scans of our local cohorts with CeVD [4]. This may suggest that T2-weighted FLAIR adds new information into the DL network, which helps to improve the estimation of NS. The current MRE of less than 7% was greater than our previous results of less than 2% [4], but we used more subjects for validation in this study (20 vs. 4) and an error of less than 7% in GM is still considered small. In addition, sensitivity analysis demonstrated the reliability of $SA\beta_L$ (SN) in showing consistent associations with the 17 variables, measured from neuropsychological and functional assessment tests, for the whole cohort, cohorts with memory concerns and/or cognitive impairment and in longitudinal data despite the greater variation observed for longitudinal data. Although $SA\beta_L$ (HRN) generally showed the same trend, the variations were huge for different cohorts and longitudinal data. We thus suggest using multi-modal ScaleNet to estimate the NS of the Aβ-PET radiotracers.

Study limitations

For ADNI3 imaging data, attempts have been made to harmonize the PET images for more accurate quantification. However, the MR images

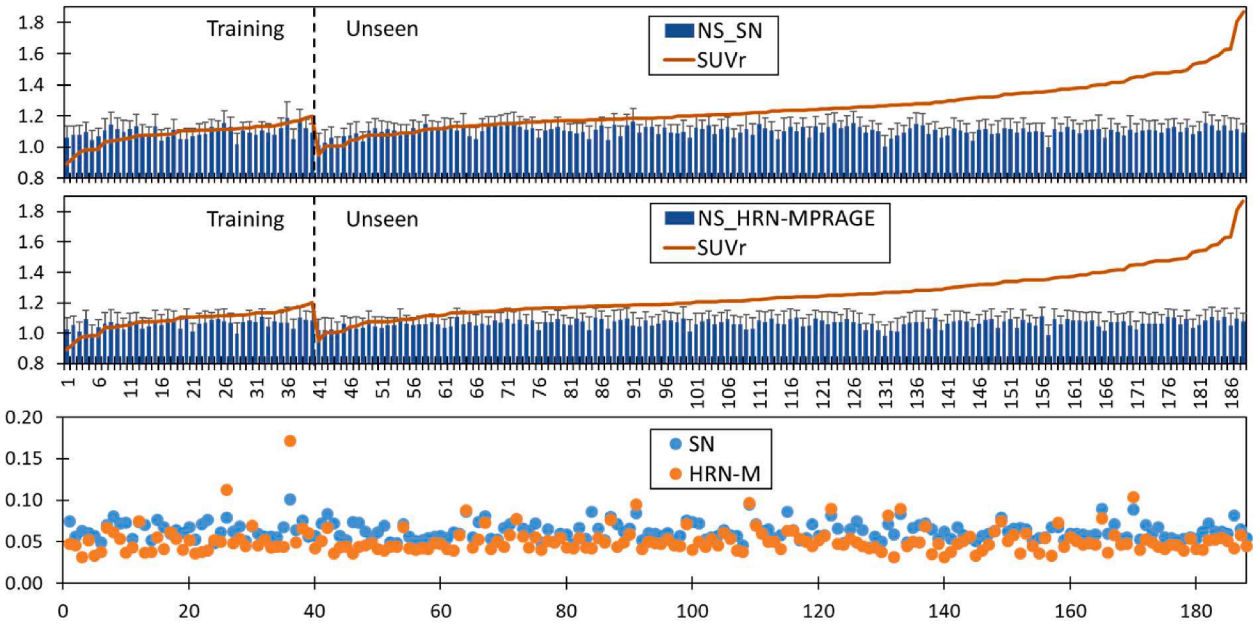

Fig. 6. Averaged nonspecific uptake (NS) estimated using ScaleNet (SN) and HighResNet (HRN) with T1-weighted MPRAGE as inputs (Top and bottom bar charts), of the whole cohort of 188 scans (40 training and 148 unseen). NS was averaged over 5 repeated networks and the original network. The red line gives the corresponding SUVr values measured from the actual PET scans. Scatter plots shows the standard deviation of NS over 5 repeated measurements and the original NS estimates. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

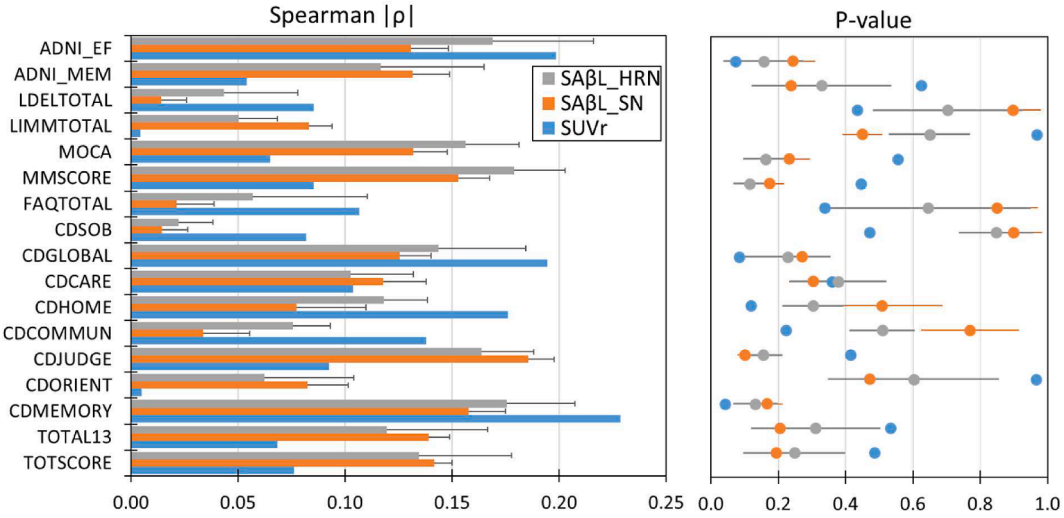

Fig. 7. Average Association (Spearman's ρ) and confidence (p-values) of 3 PET biomarkers of the brain Aβ burden with cognition and neurodegeneration for 87 subjects with memory concerns and/or cognitive impairment averaged over 5 repeated measurements in sensitivity analysis, as well as the original measurement.

acquired are not always similar in quality both across centres and within the centre for longitudinal data. The MR images were visually assessed and images that did not meet the selection criteria were removed leading to a small sample size of only 188 scans, mainly due to huge variations observed in T2-weighted FLAIR images. This limited our evaluation using longitudinal data as most of the subjects were cognitive normal and the performance of our proposed biomarker $SA\beta_L$ over SUVr in detecting changes in Aβ load with cognition decline was not as obvious as expected. Although we included scans with SUVr higher than the recommended cut-points for training, the networks (SN and HRN-MPRAGE) estimated the NS decently well with only a few subjects with NS greater than SUVr uptake. For more accurate NS estimation, we suggest using PET scans acquired in subjects with very low SUVr values, though ideally young healthy volunteers, to ensure that the NS-PET truly did not contain any Aβ. MR image post-processing was performed with normalization and/or without bias-correction and to ensure consistency in data input into the networks for training as we suspect the accuracy of our proposed method depends on the quality of the MR images. However, further bias correction did not improve the network performance or NS estimation.

Conclusions

Our proposed $SA\beta_L$ derived using NS estimated from DL network showed improvement in quantifying the Aβ burden and has stronger association with cognitive and functional test scores than SUVr for both single or longitudinal data. It also showed good reliability with consistent associations with neuropsychological and functional assessment tests in sensitivity analysis. Above all, SAβL yielded higher associations than SUVr with cognitive scores that better reflects early-AD symptoms

of memory/cognitive decline of ADNI-MEM, LIMMTOTAL, MOCA, MMSCORE, CDJUDGE, CDORIENT, TOTAL13 and TOTSCORE. Although our method requires structural MR images, it does not require any transformation of PET images to MNI space and can generate subject-specific NS and SAB_L images for better visualization and flexible quantification using either population-based or subject-specific atlas. Moreover, it can be easily implemented using the open-source NiftyNet platform (https://niftynet.io/) and can be set up specifically for different datasets or different PET-radiotracers. However, we suggest using multimodal networks such as ScaleNet with both T1-weighted and T2-weighted images for better NS estimation.

Compliance with Ethical Standards

Research involving human participants and/or animals

All ADNI procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. The study was approved by the institutional review boards of all the participating institutions and informed written consent was obtained from all participants at each site. More details can be found at adni.loni.usc.edu. This article does not contain any studies with human participants performed by any of the authors.

Informed consent

ADNI administrators have granted the authors consent for publication.

Availability of data and material

The data will not be available.

Code availability

The code will not be available.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.;Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.;Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ejmp.2022.05.016.

References

[1] Nichols E, Vos T. Estimating the global mortality from Alzheimer's disease and other dementias: a new method and results from the Global Burden of Disease study 2019: Epidemiology / Prevalence, incidence, and outcomes of MCI and dementia. Alzheimer's Dementia 2020;16(S10).
[2] Liu H, Nai YH, Chen C, Reilhac A. Deep learning-based estimation of non-specific uptake in amyloid-pet images from structural mri for improved quantification of amyloid load in alzheimer's disease. Proc. - IEEE Symp. Comput. Med. Syst., vol. 2020- July, Institute of Electrical and Electronics Engineers Inc.; 2020, p. 573–8. 10.1109/CBMS49503.2020.00114.
[3] Asher Mullard. Landmark Alzhemier's Drug Approval Confounds Research Community. Springer Nat Ltd 2021. https://media.nature.com/original/magazine-assets/d41586-021-01546-2/d41586-021-01546-2.pdf.
[4] Liu H, Nai Y-H, Saridin F, Tanaka T, O' Doherty J, Hilal S, et al. Improved amyloid burden quantification with nonspecific estimates using deep learning. Eur J Nucl Med Mol Imaging 2021;48(6):1842–53.
[5] Lundqvist R, Lilja J, Thomas BA, Lötjönen J, Villemagne VL, Rowe CC, et al. Implementation and validation of an adaptive template registration method for 18F-flutemetamol imaging data. J Nucl Med 2013;54(8):1472–8.
[6] Lilja J, Leuzy A, Chiotis K, Savitcheva I, Sörensen J, Nordberg A. Spatial normalization of 18 F-flutemetamol PET images using an adaptive principal-component template. J Nucl Med 2019;60:285–91. https://doi.org/10.2967/jnumed.118.207811.
[7] Tanaka T, Stephenson MC, Nai Y-H, Khor D, Saridin FN, Hilal S, et al. Improved quantification of amyloid burden and associated biomarker cut-off points: results from the first amyloid Singaporean cohort with overlapping cerebrovascular disease. Eur J Nucl Med Mol Imaging 2020;47(2):319–31.
[8] Kang SK, Seo S, Shin SA, Byun MS, Lee DY, Kim YK, et al. Adaptive template generation for amyloid PET using a deep learning approach. Hum Brain Mapp 2018;39(9):3769–78.
[9] Jack CR, Wiste HJ, Weigand SD, Therneau TM, Lowe VJ, Knopman DS, et al. Defining imaging biomarker cut points for brain aging and Alzheimer's disease. Alzheimer's Dement 2017;13(3):205–16.
[10] Klunk WE, Koeppe RA, Price JC, Benzinger TL, Devous MD, Jagust WJ, et al. The Centiloid Project: Standardizing quantitative amyloid plaque estimation by PET. Alzheimer's Dement 2015;11(1):1.
[11] Whittington A, Gunn RN. Amyloid load: A more sensitive biomarker for amyloid imaging. J Nucl Med 2019;60:536–40. https://doi.org/10.2967/jnumed.118.210518.
[12] Leuzy A, Lilja J, Buckley CJ, Ossenkoppele R, Palmqvist S, Battle M, et al. Derivation and utility of an Aβ-PET pathology accumulation index to estimate Aβ load. Neurology 2020;95:e2834–44. 10.1212/WNL.0000000000011031.
[13] Pegueroles J, Montal V, Bejanin A, Vilaplana E, Aranha M, Santos-Santos MA, et al. AMYQ: An index to standardize quantitative amyloid load across PET tracers. Alzheimer's Dement 2021;17(9):1499–508.
[14] Castiglioni I, Rundo L, Codari M, Di Leo G, Salvatore C, Interlenghi M, et al. AI applications to medical images: from machine learning to deep learning. Physica Medica. Associazione Italiana di Fisica. Medica 2021;83:9–24.
[15] Arabi H, AkhavanAllaf A, Sanaat A, Shiri I, Zaidi H. The promise of artificial intelligence and deep learning in PET and SPECT imaging. Phys Medica 2021 Mar;1(83):122–37.
[16] Manco L, Maffei N, Strolin S, Vichi S, Bottazzi L, Strigari L. Basic of machine learning and deep learning in imaging for medical physicists. Phys Medica 2021 Mar;1(83):194–205.
[17] Barragán-Montero A, Javaid U, Valdés G, Nguyen D, Desbordes P, Macq B, et al. Artificial intelligence and machine learning for medical imaging: A technology review. Vol. 83, Physica Medica. Associazione Italiana di Fisica Medica; 2021. p. 242–56.
[18] Wolterink JM, Dinkla AM, Savenije MHF, Seevinck PR, Berg CAT van den, Isgum I. Deep MR to CT Synthesis using Unpaired Data. Lect Notes Comput Sci (Including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) 2017;10557 LNCS:14–23. Available from: https://arxiv.org/abs/1708.01155v1.
[19] Lei Y, Dong X, Wang T, Higgins K, Liu T, Curran WJ, et al. Whole-body PET estimation from low count statistics using cycle-consistent generative adversarial networks. ab4891 Phys Med Biol 2019;64(21). https://doi.org/10.1088/1361-6560/ab4891.

[20] Fidon L, Li W, Garcia-Peraza-Herrera LC, Ekanayake J, Kitchen N, Ourselin S, et al. Scalable multimodal convolutional networks for brain tumour segmentation. Lect Notes Comput Sci (Including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) 2017;10435 LNCS:285–93. 10.1007/978-3-319-66179-7_33.

[21] Li W, Wang G, Fidon L, Ourselin S, Cardoso MJ, Vercauteren T. On the compactness, efficiency, and representation of 3D convolutional networks: Brain parcellation as a pretext task. Lect Notes Comput Sci (Including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) 2017;10265 LNCS:348–60. 10.1007/978-3-319-59050-9_28.

[22] Gibson E, Li W, Sudre C, Fidon L, Shakir DI, Wang G, et al. NiftyNet: a deep-learning platform for medical imaging. Comput Methods Programs Biomed 2018;158:113–22.

[23] Landau S, Jagust W. Florbetapir processing methods. 2015; Available from: https://adni.bitbucket.io/reference/docs/UCBERKELEYAV45/ADNI_AV45_Methods_JagustLab_06.25.15.pdf.

[24] Kim JW, Byun MS, Sohn BK, Yi D, Seo EH, Choe YM, et al. Clinical dementia rating orientation score as an excellent predictor of the progression to Alzheimer's disease in mild cognitive impairment. Psychiatry Investig 2017;14(4):420.

[25] Kang JM, Cho Y-S, Park S, Lee BH, Sohn BK, Choi CH, et al. Montreal cognitive assessment reflects cognitive reserve. BMC Geriatr 2018;18(1). https://doi.org/10.1186/s12877-018-0951-8.