Multivariate analysis of dual-point amyloid PET intended to assist the diagnosis of Alzheimer's disease

F. Segovia[a,*], J. Ramírez[a], D. Castillo-Barnes[a], D. Salas-Gonzalez[a], M. Gómez-Río[b], P. Sopena-Novales[c], C. Phillips[d], Y. Zhang[e], J. M. Górriz[a]

[a]*Department of Signal Theory, Networking and Communications, University of Granada, Granada, Spain*
[b]*Department of Nuclear Medicine, "Virgen de las Nieves" University Hospital, Granada, Spain*
[c]*Department of Nuclear Medicine, "9 de Octubre" Hospital, Valencia, Spain*
[d]*Cyclotron Research Centre, University of Liège, Liège, Belgium*
[e]*Department of Informatics, University of Leicester, Leicester, United Kingdom*

Abstract

Several studies have recently suggested that amyloid Positron Emission Tomography (PET) data acquired immediately after the radiotracer injection provide information related to the brain metabolism, similar to that contained in ^{18}F-Fluorodeoxyglucose (FDG) PET neuroimages. If corroborated, it would allow us to acquire information about brain injury and potential brain amyloid deposits in a single examination, using a dual-point protocol.

In this work we assess the equivalence between early ^{18}F-Florbetaben (FBB) PET and ^{18}F-FDG PET data using multivariate approaches based on machine learning. In addition, we propose several systems based on data fusion that take advantage of the additional information provided by dual-point amyloid PET examinations. The proposed systems perform an initial dimensionality reduction of the data using a partial-least-square-based algorithm and then combine early and standard PET acquisitions using two approaches: multiple kernel learning (intermediate fusion) or an ensemble of two Support Vector Machine classifiers (late fusion). The proposed approaches were evaluated and compared with other fusion techniques using data from 43 subjects with cognitive impairments. They achieved a good trade-off between sensitivity and specificity and higher accuracy rates than systems based on single-modality approaches such as standard ^{18}F-FBB PET data or ^{18}F-FDG PET neuroimages.

Keywords: computer aided diagnosis, multimodal systems, amyloid PET imaging, support vector machine, multiple kernel learning, late fusion, partial least squares, Alzheimer's disease

*Corresponding author
Email address: fegovia@ugr.es (F. Segovia)

1. Introduction

According to the World Health Organization, about 70% of cases of dementia are due to Alzheimer's disease (AD), a neurodegenerative disease characterized by the presence of brain amyloid plaques and neurofibrillary tangles [1, 2]. Nowadays, different modalities of neuroimaging data are frequently used to diagnose and monitor the evolution of AD. For example, Magnetic Resonance Imaging (MRI) is commonly used to estimate the brain atrophy caused by the disease [3, 4]. Nevertheless, the most common neuroimaging modality in AD diagnosis is probably ^{18}F-Fluorodeoxyglucose (FDG) Positron Emission Tomography (PET) [5, 6, 7], a molecular modality that allows us to visualize the glucose metabolism. In addition, AD can be detected by looking for brain amyloid deposits using specific radiotracers such as the N-methyl-[^{11}C]2-(4'-methylaminophenyl)-6-hydroxybenzothiazole, commonly referred as Pittsburgh Compound B (PiB). This radiotracer is a radioactive analog of thioflavin T, which binds to amyloid plaques with high affinity. Since 2012 other fluoride labeled radiotracers with longer half-live have been successfully used with same purpose.

Recently, some studies have suggested that amyloid radiotracers can be used along with a dual-point (or dual-phase) protocol, allowing us to obtain two neuroimages with different information in a single examination [8, 9]. The first one would be acquired immediately after the radiotracer injection and would contain information about the neuronal injury, similar to that obtained using a ^{18}F-FDG PET examination. Indeed, most of amyloid radiotracers are lipophilic agents, which are commonly used to assess the regional cerebral blood flow (rCBF), thus their early acquisition could be used to estimate the regional cerebral metabolic rate of glucose (rCMRglc), closely related to rCBF. The second neuroimage would be acquired according to the kinetics of the radiotracer and would allow us to assess the potential brain amyloid deposits [10].

In this work we evaluate the equivalence between early acquisitions of amyloid PET and ^{18}F-FDG PET using multivariate approaches based on machine learning [11]. In addition, we propose several multimodal Computer-Aided Diagnosis (CAD) systems in order to take advantage of the additional information provided by amyloid PET examinations using a dual-point protocol.

2. Material and methods

2.1. Neuroimaging dataset

A dataset with neuroimaging data from 43 subjects with cognitive impairments was used for evaluation purposes. The subjects were recruited in the Cognitive Behavioral Unit of two tertiary hospitals in Spain. At the first visit, experienced neurologists evaluated each subject by means of cognitive tests (including medical interviews and standardized tests for neuropsychological examination of the domains of orientation, attention, memory, executive function, language, visual and constructive functions, and behavior) and ordered ancillary neuroimaging examinations: ^{18}F-Florbetaben (FBB) PET and ^{18}F-FDG PET. The former examination was carried out following a dual-point protocol,

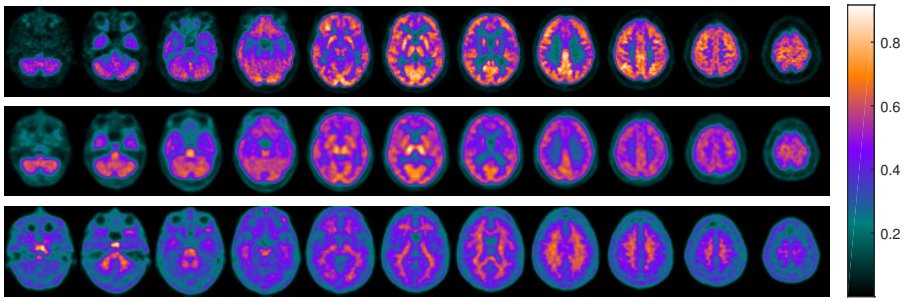

Figure 1: Comparison of the three neuroimaging modalities used in this work. Axial slices from a FDG (top), eFBB (middle) and sFBB (bottom) neuroimage are shown. The three neuroimages correspond to the same patient, who was diagnosed with AD.

resulting in two neuroimages: i) an early acquisition (eFBB) with information about the neuronal injury and ii) a standard acquisition (sFBB) intended to evaluate the brain amyloid deposits. The acquisition of eFBB data started immediately after the radiotracer injection whereas sFBB data was acquired 90 minutes after the radiotracer injection. Both acquisitions lasted for 20 minutes [10]. The ^{18}F-FDG PET examination was conducted according to internationally accepted criteria [12], resulting in one neuroimage showing the brain metabolism. The acquisition protocols were as follows (further details are given in [10]):

- **Center A**: A General Electric Discovery STE 16 camera was used. Data were reconstructed using a VUE Point Iterative algorithm (5 it, 35 sub) and filtered by a standard Z-Axis filter. The following corrections were also applied: Scatter; CT attenuation; well counter sensitivity and activity; delayed event subtraction and normalization.

- **Center B**: A Siemens Biograph 16 camera was used. Data were reconstructed using an OS-OM algorithm (6 it, 21 sub) and filtered by a standard Z-Axis filter. The following corrections were also applied: Scatter; CT attenuation; Slice coincidence location with CT.

The subjects were then monitored during at least 1 year before making a final diagnosis. At that time, two groups were defined: AD patients and subjects with disorders other than AD. Demographics details about the two groups are given in table 1

All participants gave their written informed consent to participate in the study, which complied with the principles of the Helsinki Declaration and was approved by the hospital Ethics and Investigation Committee.

2.2. Data preprocessing

After their acquisition, the PET images were spatially and intensity normalized in order to make data from different subjects comparable. Spatial normalization consisted in warping the images in a standard space, ensuring that any

Table 1: Demographic details and group distribution of subjects involved in this work. #, μ and σ stand for the number of patients, mean and standard deviation respectively.

		Sex		Age		
	#	M	F	μ	σ	range
AD	21	12	9	65.00	7.87	49–74
Non-AD	22	15	7	63.41	9.08	42–79

given voxel in different images corresponded to the same anatomical structure (SPM8, [13])

Subsequently the intensity of each voxel was normalized by dividing them by a value computed as the mean intensity of the 1% of voxels with maximum intensity. This step was carried out individually for each image and intended to eliminate the differences due to the use of different scanners for data acquisition.

2.3. Multivariate analysis of neuroimaging data based on statistical classification

Statistical classification is usually considered an instance of supervised learning where a classifier, trained with N-dimensional patterns $\mathbf{x}_i$ and their class labels $y_i \in \{0, 1, ..., L\}$, is able to determine the class label of a new pattern $\mathbf{x}$. Mathematically, it can be seen as a function $f : \mathbb{R}^N \rightarrow \{0, 1, ..., L\}$.

In this work, we are focusing on Support Vector Machine (SVM), a family of classification algorithms that builds the classification function using a hyperplane computed to maximize the margin between the classes [14]:

$$g(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + w_0 = 0 \tag{1}$$

where $\mathbf{w}$ is the weight vector, orthogonal to the decision hyperplane, and w_0 is the threshold. Once the hyperplane is defined, the class of a new unknown pattern, $\mathbf{x}_i$, is estimated according to which side of the hyperplane it is, i.e. as the sign of $\mathbf{w}^T\mathbf{x}+w_0$. Each pattern is composed of a set of *features* that defined it. The straightforward approach to classify neuroimages consists on using the intensity of all voxels as features [15]. This approach may lead to the so-called small sample size problem (small number of training patterns compared to the number of features) [16], even when using the *kernel trick* [17]. The small sample size problem can be addressed by: i) selecting the voxels in some regions of interest and discarding the remaining ones (requires previous knowledge) or ii) applying some feature extraction method that summarizes the information into a reduced set of features [6].

2.4. Feature selection based on previous knowledge

As mentioned above, the small sample size problem can be addressed by selecting a reduced set of features, leading to a reduction of the dimensionality of the classification problem. In neuroimaging studies, this can be done by selecting only the voxels in regions which are known to be affected (previous

knowledge) by the disease under study. The following regions are often used for AD [18, 8, 9]: medial temporal, lateral temporal, precuneus, posterior cingulate, anterior cingulate, frontal, occipital, striatum, thalami and cerebellum.

2.5. Feature extraction based on Partial Least Squares

Partial Least Squares (PLS) [19] is a statistical procedure to model the relation between a dataset of predictors, $\mathbf{X} \subset \mathbb{R}^{n \times N}$, and other dataset of responses, $\mathbf{Y} \subset \mathbb{R}^{n \times M}$. It decomposes the data into the form:

$$\begin{aligned} \mathbf{X} &= \mathbf{T}\mathbf{P}^T + \mathbf{E} \\ \mathbf{Y} &= \mathbf{U}\mathbf{Q}^T + \mathbf{F} \end{aligned} \tag{2}$$

where $\mathbf{T} \subset \mathbb{R}^{n \times p}$ and $\mathbf{U} \subset \mathbb{R}^{n \times p}$ contain the p extracted score vectors, $\mathbf{P} \subset \mathbb{R}^{N \times p}$ and $\mathbf{Q} \subset \mathbb{R}^{M \times p}$ are the orthogonal loadings matrices, and $\mathbf{E} \subset \mathbb{R}^{n \times N}$ and $\mathbf{F} \subset \mathbb{R}^{n \times M}$ are the matrices of residuals.

In order to extract the most relevant features from our data, we used the voxel intensity of the training neuroimages and their corresponding class labels as the datasets $\mathbf{X}$ and $\mathbf{Y}$ respectively. After performing the PLS decomposition of those matrices, the score vectors (matrix $\mathbf{T}$) summarized the training data and were used as features. The features for new patterns (not included in the training set) were obtained by projecting them onto the $\mathbf{X}$ loadings [20].

2.6. Combining multiple neuroimage modalities

Multimodal systems have been proposed to improve the performance of the classification procedures used in CAD systems when several examinations (for example, several neuroimages of different modality) are available. According to where data from different sources are combined, three different schemes can be defined [21]:

- **Early fusion.** Data from different modalities are concatenated before the classification. The pattern $\mathbf{x}_i$, representing the i-th subject, is defined as:

$$\begin{aligned} \mathbf{x}_i = \{ & x_{i_1}^{(1)}, x_{i_2}^{(1)}, ...x_{i_L}^{(1)}, \\ & x_{i_1}^{(2)}, x_{i_2}^{(2)}, ...x_{i_L}^{(2)}, ..., \\ & x_{i_1}^{(K)}, x_{i_2}^{(K)}, ...x_{i_L}^{(K)} \} \end{aligned} \tag{3}$$

  where $x_{i_j}^{(k)}$ stands for the j-th feature from neuroimage k. This approach is indeed a fusion (concatenation) of features. Therefore, it is more sensitive to the small sample size problem, especially when different neuroimaging modalities are combined (as in this work), and the application of dimensionality reduction methods is usually desired. These methods can be applied independently to each neuroimage (dimensionality reduction before fusion) or to the combined feature set (dimensionality reduction after fusion). In this work, we used the PLS-based method described above to summarize the features of each neuroimage before the combination.

Table 2: Accuracy, sensitivity and specificity achieved by a SVM classifier using i) the intensity of all voxels as feature (top), ii) a PLS approach for feature extraction (middle) or iii) the intensity of voxels in well-known AD ROIs (bottom). These measures were computed as the mean (± standard deviation) of 5 repetitions of a 10-fold CV loop.

Feature extraction	Image modality	Accuracy	Sensitivity	Specificity
Voxel as feature	FDG	68.84 % (± 2.08 %)	54.14 % (± 0.00 %)	80.00 % (± 4.07 %)
Voxel as feature	eFBB	68.37 % (± 2.08 %)	67.62 % (± 3.98 %)	69.09 % (± 3.80 %)
Voxel as feature	sFBB	72.09 % (± 2.33 %)	54.29 % (± 2.61 %)	89.09 % (± 2.49 %)
PLS	FDG	71.63 % (± 1.95 %)	60.00 % (± 5.43 %)	82.73 % (± 2.03 %)
PLS	eFBB	70.23 % (± 4.16 %)	62.86 % (± 5.22 %)	77.27 % (± 3.21 %)
PLS	sFBB	81.40 % (± 3.29 %)	79.05 % (± 2.61 %)	83.64 % (± 5.18 %)
ROIs for AD	FDG	69.77 % (± 2.33 %)	57.14 % (± 0.00 %)	81.82 % (± 4.55 %)
ROIs for AD	eFBB	69.30 % (± 2.55 %)	69.52 % (± 4.26 %)	69.09 % (± 2.03 %)
ROIs for AD	sFBB	73.95 % (± 1.95 %)	54.29 % (± 2.61 %)	92.73 % (± 2.49 %)

- **Intermediate fusion.** Data from different modalities are combined during the classification [22]. This approach involves the use of the so-called *kernel trick* [17] and the combination of the kernels obtained from each modality. To this end, we used a linear combination function:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \sum_{m=1}^{K} w_m k_m(\mathbf{x}_i^{(m)}, \mathbf{x}_j^{(m)}) \tag{4}$$

  where K is the number of kernels (i.e. the number of data sources); w_m stands for the weight for kernel k_m; $\mathbf{x}_i$, $\mathbf{x}_j$ are feature vectors for patterns i and j and $\mathbf{x}_i^{(m)}$, $\mathbf{x}_j^{(m)}$ are subset of $\mathbf{x}_i$, $\mathbf{x}_j$ with the features obtained from the data source m. More sophisticated (non-linear) functions can also be used, however they would make the model more complicated (increasing the computational burden) with unclear benefits.

- **Late fusion.** Data from different modalities are combined after the classification. Here, there is one classifier per modality. Each of them was used with the data from one source. Then, the different classification outputs are combined to obtain a single prediction per subject. Different combination options can be applied and most of them are based on a measure of classification confidence. In addition, different weights can be applied to the output of each single classifier. In this work, two methodologies to combine the classification outputs were used: i) majority voting and ii) greatest confidence, which is computed as the distance to the hyperplane.

3. Experiments and results

First, a SVM classifier was used to separate AD and non-AD patients using each of the neuroimaging modalities considered in this work independently. This

Table 3: Mean accuracy (± standard deviation) obtaining by a SVM classifier when separating AD and non-AD patients using the intensity of voxels in specific ROIs as feature.

Region	FDG	eFBB	sFBB
Medial temporal	61.86 % (± 4.82 %)	56.74 % (± 2.08 %)	75.35 % (± 1.27 %)
Lateral temporal	77.67 % (± 1.27 %)	62.79 % (± 1.64 %)	76.28 % (± 1.95 %)
Precuneus	71.63 % (± 1.04 %)	64.65 % (± 3.03 %)	81.86 % (± 1.95 %)
Posterior cingulate	69.77 % (± 3.68 %)	63.72 % (± 2.65 %)	75.35 % (± 1.27 %)
Anterior cingulate	50.23 % (± 3.53 %)	58.60 % (± 4.16 %)	76.28 % (± 1.04 %)
Frontal	62.33 % (± 1.04 %)	63.72 % (± 2.08 %)	73.02 % (± 1.27 %)
Occipital	70.70 % (± 2.65 %)	53.49 % (± 3.68 %)	73.02 % (± 2.08 %)
Striatum	49.77 % (± 7.28 %)	62.79 % (± 2.85 %)	77.21 % (± 3.03 %)
Parietal	69.77 % (± 1.64 %)	68.37 % (± 2.08 %)	71.63 % (± 3.03 %)
Thalami	51.63 % (± 7.95 %)	64.65 % (± 3.45 %)	71.16 % (± 2.08 %)
Cerebellum	53.95 % (± 5.55 %)	52.56 % (± 1.27 %)	62.33 % (± 1.04 %)

allowed us to compare FDG and eFBB neuroimages from the perspective of a CAD system and determined the ability of amyloid data (sFBB) to differentiate among AD and non-AD patients without using information about the neuronal injury. Due to the reduced number of patients in our datasets, the classification performance was estimated using a 10-fold cross-validation (CV) strategy. The accuracy, sensitivity and specificity was computed as the mean of those obtained in 5 repetitions of the CV loop. They are shown in table 2 (top).

In that initial experiment, we used the straightforward approach for features extraction, i.e. the intensity of all brain voxels was used as feature. As mentioned above, this approach often leads to poor results due to the small sample size problem. Thus, we repeated the classification including a PLS feature extraction procedure that reduced the sample size, achieving higher accuracy rates. The results are included in table 2 (middle). After the PLS decomposition, the $\mathbf{X}$ loadings (matrix $\mathbf{P}$ in equation 2) define the new reduced space where data is translated, being the first columns the dimensions gathering the most of the total covariance between the neuroimages and their labels. Thus, in order to visualize the regions focused by the PLS-based feature extraction algorithm, we represented the first column of the $\mathbf{X}$ loadings corresponding to each data modality in "brain form" [23]. They are shown in figure 2.

Alternatively, we calculated the classification performance when only the voxels in some selected ROIs are used (feature selection based on previous knowledge). In this case, the Automated Anatomical Labeling (AAL) atlas [24] was used as a mask to the select specific regions mentioned in section 2.4. The results when the voxels in all those ROIs were used together are also included in table 2 (bottom) and the accuracy rates achieved with the voxels of each individual ROI are shown in table 3.

Subsequently, we tested the 3 schemes of multimodal combination (early, intermediate and late fusion) for each pair of neuroimaging modalities. In all cases, a SVM classifier was used along with either all the voxels as feature

Table 4: Classification performance measures obtained by the multimodal CAD systems implemented in this work. Accuracy, sensitivity and specificity were calculated as the mean (± standard deviation) of 5 repetitions of a 10-fold CV loop. VaF stands for voxel as feature

Fusion approach	Feats.	Image modalities	Accuracy	Sensitivity	Specificity
Early fusion	VaF	FDG + sFBB	73.02 % (± 1.27 %)	59.05 % (± 2.61 %)	86.36 % (± 3.21 %)
Early fusion	VaF	eFBB + sFBB	71.16 % (± 2.08 %)	61.90 % (± 0.00 %)	80.00 % (± 4.07 %)
Early fusion	PLS	FDG + sFBB	81.86 % (± 1.95 %)	77.14 % (± 3.98 %)	86.36 % (± 3.21 %)
Early fusion	PLS	eFBB + sFBB	77.21 % (± 1.04 %)	71.43 % (± 0.00 %)	82.73 % (± 2.03 %)
Intermediate fusion	VaF	FDG + sFBB	85.12 % (± 3.53 %)	81.90 % (± 3.98 %)	88.18 % (± 5.18 %)
Intermediate fusion	VaF	eFBB + sFBB	81.86 % (± 3.03 %)	77.14 % (± 3.98 %)	86.36 % (± 3.21 %)
Intermediate fusion	PLS	FDG + sFBB	83.26 % (± 1.95 %)	80.95 % (± 0.00 %)	85.45 % (± 3.80 %)
Intermediate fusion	PLS	eFBB + sFBB	79.07 % (± 1.64 %)	72.38 % (± 2.13 %)	85.45 % (± 3.80 %)
Late fusion	VaF	FDG + sFBB	76.28 % (± 1.04 %)	61.90 % (± 0.00 %)	90.00 % (± 2.03 %)
Late fusion	VaF	eFBB + sFBB	72.09 % (± 1.64 %)	63.81 % (± 2.61 %)	80.00 % (± 2.49 %)
Late fusion	PLS	FDG + sFBB	85.12 % (± 2.08 %)	80.95 % (± 0.00 %)	89.09 % (± 4.07 %)
Late fusion	PLS	eFBB + sFBB	83.72 % (± 2.85 %)	80.00 % (± 3.98 %)	87.27 % (± 3.80 %)

Table 5: Accuracy, sensitivity and specificity achieved by the multimodal systems that combine the three neuroimaging modalities considered in this work. VaF stands for voxel as feature

Fusion approach	Feats.	Accuracy	Sensitivity	Specificity
Early fusion	VaF	71.63 % (± 1.95 %)	62.86 % (± 2.13 %)	80.00 % (± 2.49 %)
Early fusion	PLS	81.40 % (± 1.64 %)	79.05 % (± 2.61 %)	83.64 % (± 2.49 %)
Intermediate fusion	VaF	81.40 % (± 4.03 %)	77.14 % (± 3.98 %)	85.45 % (± 5.93 %)
Intermediate fusion	PLS	81.40 % (± 2.85 %)	79.05 % (± 4.26 %)	83.64 % (± 2.49 %)
Late fusion (greatest confidence)	VaF	72.56 % (± 1.04 %)	62.86 % (± 2.13 %)	81.82 % (± 0.00 %)
Late fusion (greatest confidence)	PLS	83.26 % (± 2.55 %)	79.05 % (± 2.61 %)	87.27 % (± 3.80 %)
Late fusion (majority voting)	VaF	75.35 % (± 2.08 %)	70.48 % (± 2.13 %)	80.00 % (± 4.07 %)
Late fusion (majority voting)	PLS	80.00 % (± 1.27 %)	74.29 % (± 2.61 %)	85.45 % (± 2.03 %)

approach or the PLS-based feature extraction method. The results are shown in table 4. As in previous experiments, the performance of each model was estimated by means of a 10-fold CV strategy that was repeated 5 times. The performance of multimodal systems achieving highest accuracy rates was also compared in terms of their Receiver Operation Curves (ROCs). The result is shown in figure 4. These representations allow us to compute the Area Under the Curve (AUC), which is considered a good estimator of the classification performance [25].

Finally, we implemented a system combining all the three available modalities. The results are shown in table 5. Three fusion approaches and two feature extraction methodologies were analyzed. In this experiment we also evaluated the two rules mentioned in the previous section to combine the classifier outputs in the late fusion scheme, i.e. majority voting and greatest confidence.

4. Discussion

The results shown in the previous section can be analyzed from two perspectives: i) equivalence between FDG and eFBB data and, ii) improvement achieved by multimodal systems compared to system using a single data modality.

On the one hand, the results shown in table 2 indicate that early acquisitions of ^{18}F-FBB PET can be used as surrogates of FDG data to separate AD and non-AD patients using machine-learning-based approaches. The accuracy achieved by a CAD system fed by FDG or eFBB neuroimages was similar, regardless of the feature extraction method used (about 68% using all brain voxels as feature, 69% using the voxels in some disease-specific regions as feature, and 70-71% using a PLS-based method for feature extraction). This suggests that the estimation of the neuronal injury provided by eFBB neuroimages is similar to that provided by FDG neuroimages. In addition, the regions highlighted in figure 2 for FDB data coincide, in many cases, with those highlighted for eFBB data. Observe that, in both cases, there are big highlighted regions located in the parietal and occipital lobes, which are known to be affected by AD. However, it is worth noting that when small specific regions are considered independently, our results were not conclusive. As shown in table 3, the accuracy rate obtained with some regions extracted from FDG data differs widely from that obtained for the same regions from eFBB data. In our opinion, those regions, when isolated, do not contain information enough to separate AD and non-AD patients and that, along with the small size of our database, led to fluctuating results. This also explains the low accuracy rates (almost random classification) obtained for some regions.

The idea of using a dual-point protocol for amyloid PET to obtain information about the neuronal injury and amyloid deposits in a single examination is a promising idea that can bring important social and economic benefits [8]. The results obtained in this work support the use of early acquisitions of ^{18}F-FBB PET as surrogates of FDG data and are in line with previous works [9, 10] that analyzed the equivalent between these data modalities using univariate (voxel-wise) approaches. Specifically, they analyzed the correlation between both data modalities at specific target regions and concluded that they could be used indistinctly for the same purpose.

The results shown in table 2 also reveal that sFBB is an useful biomarker to separate AD and non-AD patients. The systems using these data provided better accuracy rates than those using other modalities, especially when a PLS-based dimensionality reduction is performed. This is not strange if we bear in mind that brain amyloid deposits are one of the neurological hallmarks of AD and they are not present in patients with other cognitive impairments. In addition, the good performance of the PLS-based feature extraction was also expected, given the effectiveness of this technique on similar problems [26, 20]. This approach transforms the data into a set of latent variables, similar to the well-known principal component analysis method, but unlike the latter, PLS maximizes the covariance between the data and their class labels. Hence, PLS

provides the main advantages of PCA (large reduction of the data dimensionality, improved signal-to-noise ratio, reduced model overfitting) and performs a data transformation focused on class separation (maximizing covariance between data and labels rather than data variance). The main disadvantages of PLS are the loss of information (any dimensionality reduction implies the elimination of some of the initial data) and the increase of the complexity of the model (additional parameters must be optimized).

On the other hand, a dual-point protocol for amyloid PET would make more sense if we are able to take advantage of the additional information it provides. Accordingly, multimodal CAD systems are a useful approach to analyze these data. From the machine learning perspective, three general schemes have been proposed to combine data from different sources: early, intermediate and late fusion [22]. The former scheme, considered as the most straightforward approach, has become more relevant during last years due to the development of deep learning methods that are able to successfully deal with the high resulting dimensionality [27, 28]. However, deep learning approaches require large training datasets to fit the large number of parameters (related to the number of layers) they involve [29]. This is not the case of this and many other neuroimaging studies, thus we decided not to use deep learning approaches. However, given that our objective was to study in depth the possibility of combining the data provided by a dual-point protocol for amyloid PET, we analyzed all the three fusion schemes. As shown in table 4, combining FDG or eFBB and sFBB data by concatenating the features (early fusion) leads to mediocre results. This can be explained because of the large dimensionality of the input space (twice as than in the case with only a neuroimage per patient). Conversely, intermediate and late schemes provide good accuracy rates, better than those obtained using a single neuroimage, making these schemes interesting approaches to take advantage of dual-point protocols.

The performance of these two schemes was also investigated in terms of the trade-off between sensitivity and specificity (see figure 4). A well balanced trade-off is particularly important in the development of CAD systems, where both false positives and false negatives should be kept as small as possible. In addition, we reported in figure 4 the AUC of each system, which is considered by many to be a more representative measure of classification performance than accuracy [25].

The equivalence between FDG and eFBB could be also investigated through these multimodal approaches, by comparing the performance of systems using FDG+sFBB and eFBB+sFBB data. According to our experiments, systems using FDG+sFBB provide higher accuracy rates than those using eFBB+sFBB. This is consistent with results reported in [30], where the authors concluded that FDG + ^{11}C-PiB has better classification accuracy than early ^{11}C-PiB + ^{11}C-PiB when separating AD patients and controls subjects. These differences were not reflected in the comparison with single-modality systems (table 2), which could indicate a larger similarity between eFBB and sFBB than between FDG and sFBB that makes the systems combining eFBB and sFBB have less information than those combining FDG and sFBB.

Finally, the CAD systems of the last experiment, intended to assess the potential advantages of combining FDG, eFBB and sFBB data, achieved lower accuracy rates than systems using only two data modalities. In some sense, these results corroborate the idea that FDG and eFBB contain similar information and that the use of both in the same system will only increase the input space, making the separation problem more complex and impairing the classification performance.

All in all, our experiments indicate that: i) eFBB data could be used as surrogates of FDG neuroimages to distinguish between AD and non-AD patients using machine learning methods and, ii) fusion techniques based on intermediate or late fusion schemes are suitable approaches to take advantage of the additional information provided by dual-point protocols for amyloid PET. Nevertheless, this work and the confidence in its conclusions are limited by the small size of our dataset. We addressed this issue by using a k-fold CV strategy and repeating the CV loop (with different sample distribution) several times, however large datasets are always desirable. Another limitation, related with the former, is the use of SVM for classification. During the last decade, approaches based on deep learning have had a great growth due to the improvements achieved in classification tasks in different fields, including medical imaging [31, 32, 33]. However, these methods require large data sets to determine the huge amount of parameters they use and this is clearly not the case in our study.

5. Conclusion

Several experiments were performed to assess whether dual-point ^{18}F-FBB PET data could be used as biomarkers to separate AD and non-AD patients using machine learning based methods. Our experiments indicated that the data acquired during early phase of ^{18}F-FBB PET (just after the radiotracer injection) provide similar results than ^{18}F-FDG PET neuroimages when they are used in modern SVM-based CAD systems for AD. In order to take advantage of the additional information provided by dual-point protocols with amyloid PET (the data acquired during the early phase), we proposed to use a multi-modal system based on intermediate or late fusion schemes. According to our experiments, these systems are able to separate AD and non-AD patients with higher accuracy than using only standard amyloid neuroimages, which contain information about the potential brain amyloid deposits, or data about the brain metabolism, such as ^{18}F-FDG PET. Nevertheless, these results should be corroborated by future studies with larger datasets. In addition, as future work, we plan to evaluate the use of modern classification systems based on deep learning for this problem. These systems are able to perform feature extraction along with classification (using a single neural network) and could be useful in dealing with the high dimensionality of data resulting from the early fusion approach.

Acknowledgment

This work was partly supported by the MINECO/FEDER under the RTI2018-098913-B100 and A-TIC-080-UGR18 projects

References

[1] A. Burns, S. Iliffe, Alzheimer's disease, BMJ 338 (2009) b158. doi:10.1136/bmj.b158.

[2] D. Castillo-Barnes, L. Su, J. Ramírez, D. Salas-Gonzalez, F. J. Martinez-Murcia, I. A. Illan, F. Segovia, A. Ortiz, C. Cruchaga, M. R. Farlow, C. Xiong, N. R. Graff-Radford, P. R. Schofield, C. L. Masters, S. Salloway, M. Jucker, H. Mori, J. Levin, J. M. Gorriz, D. I. A. N. (dian), Autosomal dominantly inherited alzheimer disease: Analysis of genetic subgroups by machine learning, Information Fusion 58 (2020) 153–167. doi:10.1016/j.inffus.2020.01.001.

[3] G. B. Frisoni, N. C. Fox, C. R. Jack, P. Scheltens, P. M. Thompson, The clinical use of structural MRI in Alzheimer disease, Nature reviews. Neurology 6 (2) (2010) 67–77. doi:10.1038/nrneurol.2009.215.

[4] D. Castillo-Barnes, I. Peis, F. J. Martínez-Murcia, F. Segovia, I. A. Illán, J. M. Gorriz, J. Ramírez, D. Salas-Gonzalez, A heavy tailed Expectation Maximization Hidden Markov Random field model with applications to segmentation of MRI, Frontiers in Neuroinformatics 11 (Nov. 2017). doi:10.3389/fninf.2017.00066.

[5] C. Cabral, P. M. Morgado, D. Campos Costa, M. Silveira, Predicting conversion from MCI to AD with FDG-PET brain images at different prodromal stages, Computers in Biology and Medicine 58 (2015) 101–109. doi:10.1016/j.compbiomed.2015.01.003.

[6] S. Rathore, M. Habes, M. A. Iftikhar, A. Shacklett, C. Davatzikos, A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer's disease and its prodromal stages, NeuroImage 155 (Supplement C) (2017) 530–548. doi:10.1016/j.neuroimage.2017.03.057.

[7] F. J. Martinez-Murcia, J. M. Górriz, J. Ramírez, I. A. Illán, F. Segovia, D. Castillo-Barnes, D. Salas-Gonzalez, Functional Brain Imaging Synthesis Based on Image Decomposition and Kernel Modeling: Application to Neurodegenerative Diseases, Frontiers in Neuroinformatics 11 (Nov. 2017). doi:10.3389/fninf.2017.00065.

[8] S. Tiepolt, S. Hesse, M. Patt, J. Luthardt, M. L. Schroeter, K.-T. Hoffmann, D. Weise, H.-J. Gertz, O. Sabri, H. Barthel, Early 18F-florbetaben and 11C-PiB PET images are a surrogate biomarker of neuronal injury in

Alzheimer's disease, European Journal of Nuclear Medicine and Molecular Imaging 43 (9) (2016) 1700–1709. doi:10.1007/s00259-016-3353-1.

[9] S. Daerr, M. Brendel, C. Zach, E. Mille, D. Schilling, M. J. Zacherl, K. Bürger, A. Danek, O. Pogarell, A. Schildan, M. Patt, H. Barthel, O. Sabri, P. Bartenstein, A. Rominger, Evaluation of early-phase [18F]-florbetaben PET acquisition in clinical routine cases, NeuroImage : Clinical 14 (2016) 77–86. doi:10.1016/j.nicl.2016.10.005.

[10] F. Segovia, M. Gómez-Río, R. Sánchez-Vañó, J. M. Górriz, J. Ramírez, E. Triviño-Ibáñez, C. Carnero Pardo, M. D. Lozano Martínez, P. Sopena-Novales, Usefulness of Dual-Point Amyloid PET Scans in Appropriate Use Criteria: A Multicenter Study, Journal of Alzheimer's Disease 65 (3) (2018) 765–779. doi:10.3233/JAD-180232.

[11] J. M. Górriz, J. Ramírez, A. Ortíz, F. J. Martínez-Murcia, F. Segovia, J. Suckling, M. Leming, Y.-D. Zhang, J. R. Álvarez-Sánchez, G. Bologna, P. Bonomini, F. E. Casado, D. Charte, F. Charte, R. Contreras, A. Cuesta-Infante, R. J. Duro, A. Fernández-Caballero, E. Fernández-Jover, P. Gómez-Vilda, M. Graña, F. Herrera, R. Iglesias, A. Lekova, J. Lope, E. López-Rubio, R. M. Tomás, M. A. Molina-Cabello, A. S. Montemayor, P. Novais, D. Palacios-Alonso, J. J. Pantrigo, B. R. Payne, F. la Paz López, M. A. Pinninghoff, M. Rincón, J. Santos, K. Thurnhofer-Hemsi, A. Tsanas, R. Varela, J. M. Ferrández, Artificial intelligence within the interplay between natural and artificial Computation: Advances in data science, trends and applications, Neurocomputing (Jun. 2020). doi:10.1016/j.neucom.2020.05.078.

[12] A. Varrone, S. Asenbaum, T. V. Borght, J. Booij, F. Nobili, K. Nå gren, J. Darcourt, Ö. L. Kapucu, K. Tatsch, P. Bartenstein, K. V. Laere, EANM procedure guidelines for PET brain imaging using 18F-FDG, version 2, European Journal of Nuclear Medicine and Molecular Imaging 36 (12) (2009) 2103. doi:10.1007/s00259-009-1264-0.

[13] J. Ashburner, K. J. Friston, Unified segmentation, NeuroImage 26 (3) (2005) 839–851. doi:10.1016/j.neuroimage.2005.02.018.

[14] V. N. Vapnik, The Nature of Statistical Learning Theory, Springer New York, New York, NY, 2000.

[15] J. Stoeckel, N. Ayache, G. Malandain, P. M. Koulibaly, K. P. Ebmeier, J. Darcourt, Automatic classification of SPECT images of Alzheimer's disease patients and control subjects, in: C. Barillot, D. R. Haynor, P. Hellier (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2004, no. 3217 in Lecture Notes in Computer Science, Springer Berlin Heidelberg, 2004, pp. 654–662.

[16] H. G. Schnack, R. S. Kahn, Detecting Neuroimaging Biomarkers for Psychiatric Disorders: Sample Size Matters, Frontiers in Psychiatry 7 (Mar. 2016). doi:10.3389/fpsyt.2016.00050.

[17] S. Theodoridis, K. Koutroumbas, Pattern Recognition, Fourth Edition, 4th Edition, Academic Press, Amsterdam, 2008.

[18] E. Rodriguez-Vieitez, S. F. Carter, K. Chiotis, L. Saint-Aubert, A. Leuzy, M. Schöll, O. Almkvist, A. Wall, B. Lå ngström, A. Nordberg, Comparison of Early-Phase 11C-Deuterium-l-Deprenyl and 11C-Pittsburgh Compound B PET for Assessing Brain Perfusion in Alzheimer Disease, Journal of Nuclear Medicine 57 (7) (2016) 1071–1077. doi:10.2967/jnumed.115.168732.

[19] K. Varmuza, P. Filzmoser, Introduction to Multivariate Statistical Analysis in Chemometrics, CRC Press, Boca Raton, 2009.

[20] L. Khedher, J. Ramírez, J. M. Górriz, A. Brahim, F. Segovia, Early diagnosis of Alzheimer's disease based on partial least squares, principal component analysis and support vector machine using segmented MRI images, Neurocomputing 151, Part 1 (2015) 139–150. doi:10.1016/j.neucom.2014.09.072.

[21] W. S. Noble, Support vector machine applications in computational biology, in: B. Scholkopf, K. Tsuda, J.-P. Vert (Eds.), Kernel Methods in Computational Biology, MIT Press, 2004, pp. 71–102.

[22] M. Gönen, E. Alpaydı n, Multiple Kernel Learning Algorithms, J. Mach. Learn. Res. 12 (2011) 2211–2268.

[23] F. Segovia, J. Górriz, J. Ramírez, D. Salas-González, I. Álvarez, Early diagnosis of Alzheimer's disease based on Partial Least Squares and Support Vector Machine, Expert Systems with Applications 40 (2) (2013) 677–683. doi:10.1016/j.eswa.2012.07.071.

[24] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, M. Joliot, Automated Anatomical Labeling of Activations in SPM Using a Macroscopic Anatomical Parcellation of the MNI MRI Single-Subject Brain, NeuroImage 15 (1) (2002) 273–289. doi:10.1006/nimg.2001.0978.

[25] A. P. Bradley, The use of the area under the ROC curve in the evaluation of machine learning algorithms, Pattern Recognition 30 (7) (1997) 1145–1159. doi:10.1016/S0031-3203(96)00142-2.

[26] F. Segovia, J. M. Górriz, J. Ramírez, I. Álvarez, J. M. Jiménez-Hoyuela, S. J. Ortega, Improved Parkinsonism diagnosis using a partial least squares based approach, Medical Physics 39 (7) (2012) 4395–4403. doi:10.1118/1.4730289.

[27] D. Lu, K. Popuri, G. W. Ding, R. Balachandar, M. F. Beg, Multimodal and Multiscale Deep Neural Networks for the Early Diagnosis of Alzheimer's Disease using structural MR and FDG-PET images, Scientific Reports 8 (1) (2018) 5697. doi:10.1038/s41598-018-22871-z.

[28] T. Zhou, K.-H. Thung, X. Zhu, D. Shen, Feature Learning and Fusion of Multimodality Neuroimaging and Genetic Data for Multi-status Dementia Diagnosis, Machine learning in medical imaging. MLMI (Workshop) 10541 (2017) 132–140. doi:10.1007/978-3-319-67389-9_16.

[29] A. Selvikvå g Lundervold, A. Lundervold, An overview of deep learning in medical imaging focusing on MRI, Zeitschrift für Medizinische Physik (Dec. 2018). doi:10.1016/j.zemedi.2018.11.002.

[30] L. Fu, L. Liu, J. Zhang, B. Xu, Y. Fan, J. Tian, Comparison of dual-biomarker PIB-PET and dual-tracer PET in AD diagnosis, European Radiology 24 (11) (2014) 2800–2809. doi:10.1007/s00330-014-3311-x.

[31] F. J. Martinez-Murcia, A. Ortiz, J.-M. Gorriz, J. Ramirez, D. Castillo-Barnes, Studying the Manifold Structure of Alzheimer's Disease: A Deep Learning Approach Using Convolutional Autoencoders, IEEE Journal of Biomedical and Health Informatics 24 (1) (2020) 17–26. doi:10.1109/JBHI.2019.2914970.

[32] D. Lu, K. Popuri, G. W. Ding, R. Balachandar, M. F. Beg, Multiscale deep neural network based analysis of FDG-PET images for the early diagnosis of Alzheimer's disease, Medical Image Analysis 46 (2018) 26–34. doi:10.1016/j.media.2018.02.002.

[33] F. Segovia Román, J. M. Gorriz, J. Ramírez Pérez de Inestrosa, F. J. Martinez-Murcia, M. García-Pérez, Using deep neural networks along with dimensionality reduction techniques to assist the diagnosis of neurodegenerative disorders, Logic Journal of the IGPL 26 (6) (2018) 618–628. doi:10.1093/jigpal/jzy026.

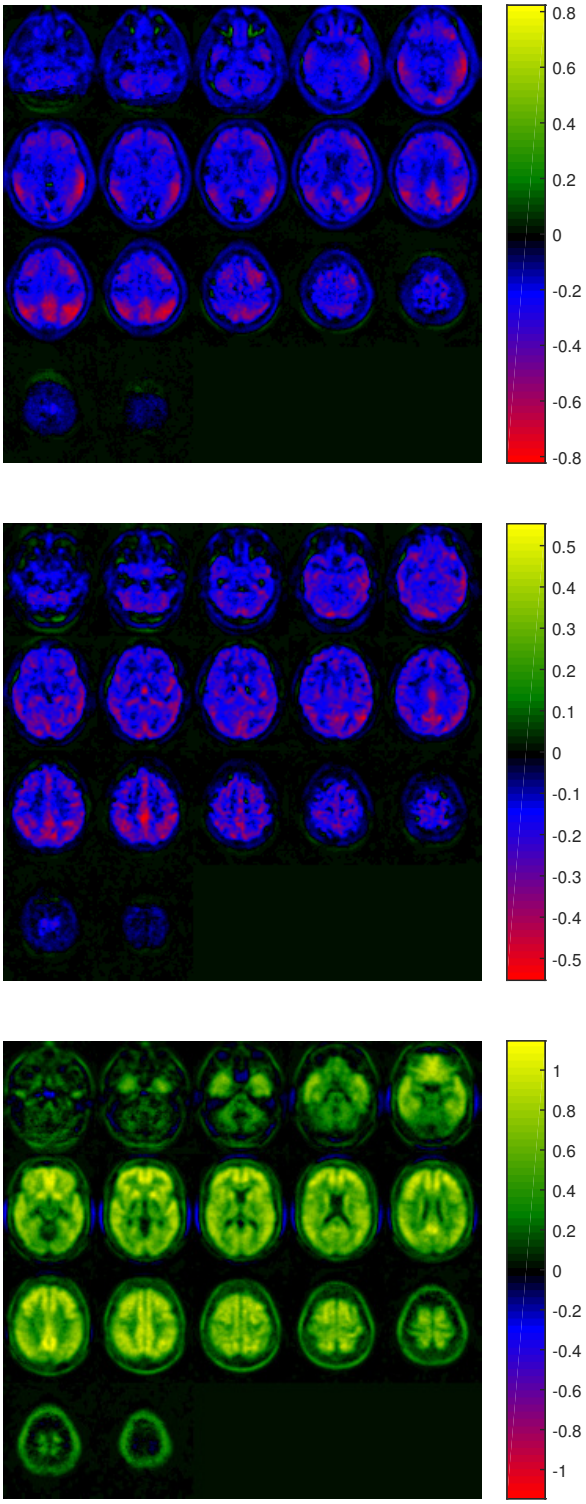

Figure 2: Brain map containing the first column of **X** loading (matrix **P** in equation 2) corresponding to the PLS decomposition of FDG (top), eFBB (middle) and sFBB (bottom) data and their labels.

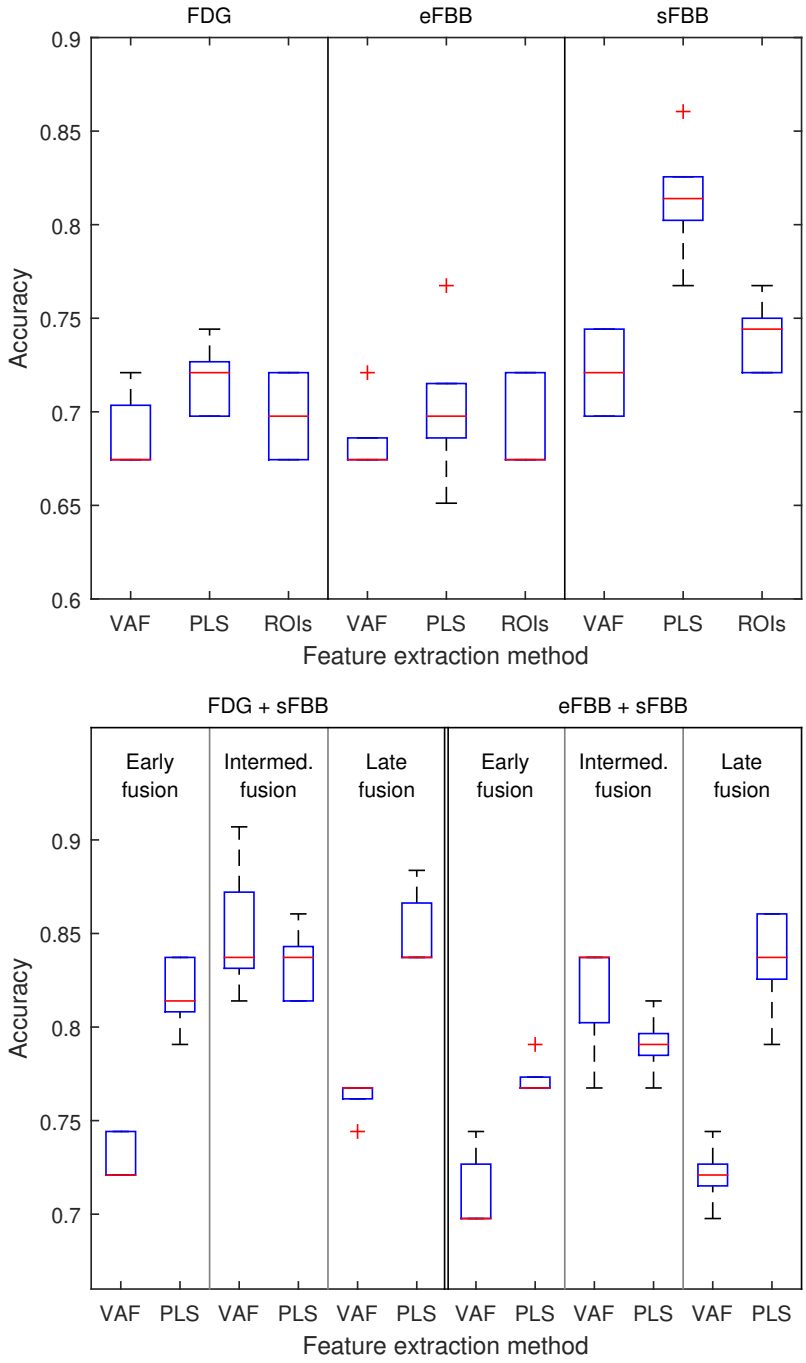

Figure 3: Accuracy obtained in each of the 5 repetitions of the cross validation loop for the systems using: i) neuroimage per patient (top) or ii) two neuroimage per patient (bottom). Blue boxes and circled dots represent accuracies' range and median respectively.

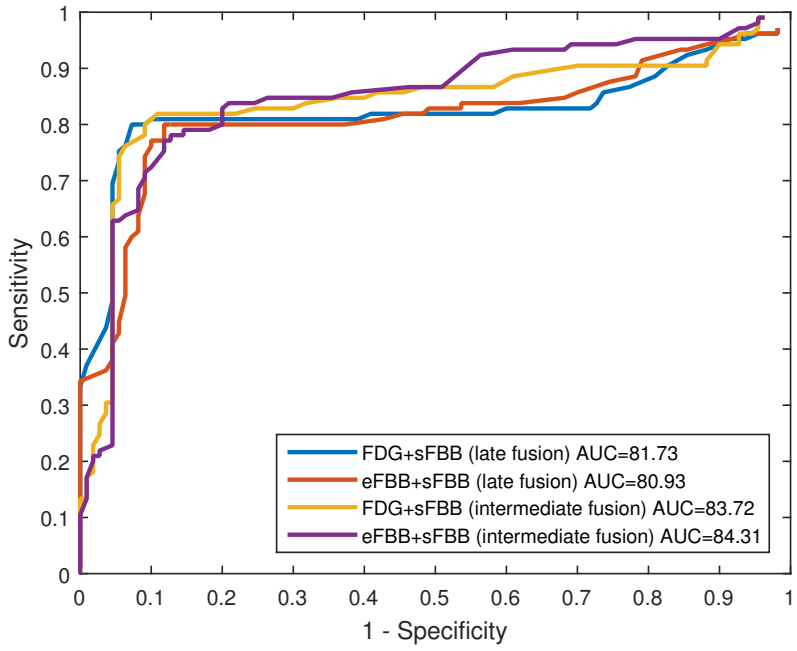

Figure 4: ROC curves for four multimodal systems, two combining FDG and sFBB data and two others combining eFBB and sFBB. The area under the curve (AUC) of each system is shown in the legend.