Automated amyloid PET image-based classification of controls and patients with mild cognitive impairment and Alzheimer's disease

Petra Parviainen [1], Janne Isojärvi[1], Marco Bucci [1,2], Jouni Tuisku[1], Anniina Snellman[1], Matti Sillanpää[1], Laura Ekblad[1], Lauri Nummenmaa[1], Juha Rinne[1,3]

1)Turku PET Centre, Turku University Hospital and University of Turku, Turku, Finland, 2)Department of Neurobiology, Care Sciences and Society, Centre for Alzheimer Research, Division of Clinical Geriatrics, Karolinska Institutet, Stockholm, Sweden, 3) InFLAMES Research Flagship Center, University of Turku, Turku, Finland

Objectives: Longitudinal changes of Alzheimer´s disease (AD) can be identified with positron emission tomography (PET) and beta-amyloid radioligands such as [^{11}C]PIB. **Methods**: The study is based on amyloid images and magnetic resonance images from healthy controls (HC) (N=218, female/male 118/100, age range 40.4; 89.0), individuals with mild cognitive impairment (MCI) (N=64, female/male 27/37, age range 51.8; 88.6) and AD (N=34, female/male 15/19, age range 55.3; 88.4) collected at Turku PET Centre. PET scans were acquired across 5 different scanners and harmonized with preserving the variance of age, gender and study group. Automated Anatomical Atlas 3 and principle component analysis (PCA) were used with *a* machine learning algorithm, Extreme Gradient Boosting (XGBoost) to classify subjects into three diagnostic categories based on regional [^{11}C]PIB load in regions of interest alone and combined with demographical information including body mass index, age, gender and clinical information including apoe status and mini mental test results, and to test whether classification corresponds with clinical diagnosis given by a clinician. **Results:** XGBoost succeeded to classify study subjects into three groups with 0.68 Area Under Curve (AUC) (CI:0.62-0.72) using PET-image data with PCA based approach and increased significantly to 0.79 AUC (CI:0.74-0.83) by adding demographical and clinical information with PET-image data ($p<0.001$). **Conclusion:** Computer-assisted recognition of regional beta-amyloid patterns may support the diagnostic process in patients with cognitive disorders, but clinical evaluation remains essential in the automated classification of diagnostic groups. In addition, a larger and more balanced sample would likely improve the accuracy of the classification model.

1. Introduction

Alzheimer's disease (AD) is the most common type of dementia that causes deficits in memory and other information-processing abilities as it slowly destroys brain tissue[1]. One of the AD pathological hallmarks, beta-amyloid protein deposits, can be detected in the brain decades before the appearance of clinical symptoms and extensive degeneration of nervous tissue[2].

Positron emission tomography (PET) is an imaging technique used to study different stages of AD pathology by using radiolabeled tracers that bind to amyloid plaques allowing evaluation of amyloid status and early diagnosis. In vivo imaging of amyloid plaques can be performed with e.g. the radiotracer N-methyl-[^{11}C]2-(4'-methylaminophenyl)-6-hydroxybenzothiazole ([^{11}C]PIB), which has a high affinity for amyloid plaques[3,4]. Studies using [^{11}C]PIB have shown higher uptake values in several brain regions in AD patients compared with healthy controls, and the analysis is typically based on region-of-interest (ROI) measurements[5]. According to neuropathological studies early amyloid accumulation is typically first detectable in prefrontal cortex, parietal cortex, lateral temporal cortex, anterior cingulate cortex, posterior cingulate cortex and precuneus[6].

The current clinical standard for detecting amyloid deposition is visual assessment of PET images by experienced readers, but this method can be slow and may lead to false-positive or false-negative results, especially in equivocal cases with unsure visual reading. Amyloid PET quantification has recently been approved for clinical use as well as has been shown to provide results comparable to visual assessment[7]. However, diagnosing conditions in the AD spectrum, particularly mild cognitive impairment (MCI) and healthy amyloid positive subjects, remains challenging because subtle cognitive decline and increase of amyloid accumulation can also occur during normal aging. Current measures have limited reliability in distinguishing healthy aging from AD, highlighting the need for improved methods to detect early amyloid pathology using PIB-PET imaging [2,8] .

Recently, different studies have established artificial intelligence methods for AD detection, and these approaches are classified as machine learning techniques (ML)[9]. Application of ML to identify diseases such as AD has increased a lot of interest as a useful tool, identifying automatically different levels of amyloid pathology. In the past decade, more computer-aided pattern recognition algorithms have been developed to evaluate and identify PET patterns that relate to specific disease stages.[10–12] Neuroimaging data includes hundreds to thousands of features. To reduce features, principal component analysis (PCA) approach can effectively reduce the dimensionality of high-resolution brain imaging data by retaining the most informative components and discarding redundant or less significant features.[13]

In the past few years, beta-amyloid antibody treatments that can slow the progression of AD have been approved for clinical use. These therapies are likely to be more effective when starting at an early stage of the disease[14]. In the future, automatic classification methods based on principal component analysis (PCA) could help to identify individuals at early stages of AD and support studies evaluating the effectiveness of anti-amyloid treatments in people at risk of the disease. Therefore, it is important to study the most relevant PCA components that may indicate increased AD risk.

The aim of this study is to automatically classify individuals into three groups: healthy control (HC), MCI and AD by using the local cohort collected at the Turku PET Centre in Finland and estimate a potential clinical applicability of the designed ML approaches. In this study, the PCA approach was used to retain the most informative and predictive ROIs based on PET imaging data together and separately from clinical variables to show the importance of amyloid-PET-imaging data to support clinical diagnosis.

2. Material and Methods

Machine learning methods were used to classify the participants into three different classes (HC, MCI and AD) based on amyloid accumulation detected by [^{11}C]PIB PET scan to test whether the predictions given by the model accord with the clinical diagnosis.

2.1. Data

[^{11}C]PIB PET and MRI scans as well as neuropsychological data consisting of research study projects were collected at the Turku PET Center between 2004-2022. The cohort includes healthy controls, and patients diagnosed with amnestic MCI or AD. The controls had no history of neurological or psychiatric disease and did not have memory complaints. The patients with amnestic MCI (for the sake of simplicity later referred as MCI) had to fulfill the criteria by Petersen et al. (2001)[15], AD patients had to meet the National Institute of Neurological and Communicative Disorders and Stroke–Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) criteria for probable AD[16] as well as DSM-IV criteria for dementia of the Alzheimer type. The participants were diagnosed at Turku University Hospital, Turku City Hospital, or a private medical clinic by a neurologist or geriatrician. The Ethics Committee of the Hospital District of Southwest Finland has approved study protocols. All research participants assigned a written informed consent at enrollment to the studies in accordance with the declaration of Helsinki.

We included 314 individuals, 216 HC, 64 MCI, and 34 AD with [^{11}C]PIB and MRI scans, and MMSE score and the apolipoprotein E (APOE)–status (ε2/ε3, ε2/ε4, ε3/ε3, ε3/ε4, ε4/ε4) from those participants it was available. See the demographic data in Table 1 and from Supplementary material Table 1 for a list of original publications whose data is used in the current study.

Table 1. *Demographics of the study population*

Characteristic	Overall N = 314[1]	HC N = 216[1]	MCI N = 64[1]	AD N = 34[1]	p-value[2]
Age	71.4 (7.1)	70.8 (6.6)	72.1 (7.7)	73.9 (8.2)	0.04
Sex					0.15
Female	160 (51%)	118 (55%)	27 (42%)	15 (44%)	
Male	154 (49%)	98 (45%)	37 (58%)	19 (56%)	
Body Mass Index	26.9 (4.6)	26.8 (4.2)	26.3 (4.0)	28.5 (7.4)	0.29
Not available	6	6	0	0	
MMSE	26.8 (2.9)	27.9 (1.9)	25.3 (3.1)	22.8 (3.2)	<0.01
Not available	9	5	4	0	
APOE genotype					0.02
E2/3	12 (4.7%)	10 (5.2%)	2 (5.3%)	0 (0%)	
E2/4	3 (1.2%)	3 (1.6%)	0 (0%)	0 (0%)	
E3/3	139 (55%)	109 (56%)	19 (50%)	11 (48%)	
E3/4	90 (35%)	69 (36%)	12 (32%)	9 (39%)	
E4/4	10 (3.9%)	2 (1.0%)	5 (13%)	3 (13%)	
Not available	60	23	26	11	

[1] Mean (SD); n (%)

[2] Kruskal-Wallis rank sum test; Pearson's Chi-squared test; Fisher's exact test

The mean age of the whole population is 71.4 years and HC group 70.8 years, MCI 72.1 years and AD 73.9 years. Gender groups are balanced (females 51 % and males 49 %) that helps to ensure fair comparisons and makes the findings more representative of the whole population. Body Mass Index (BMI) is the highest in the AD group (HC 26.8, MCI 26.8 and AD 28.5).

2.3. Neuroimaging data

The synthesis of [^{11}C]PIB has been described earlier[17]. Participants received an injection of approximately 500 MBq and underwent [^{11}C]PIB scan from 40 to 90 minutes after injection (Lindgren et al., 2021). The [^{11}C]PIB images were acquired by five different PET scanners (ECAT EXACT HR + (Siemens Medical Systems, Knoxville, TN, USA), HRRT High-Resolution Research Tomograph (CTI PET Systems, Knoxville, TN, USA), GE Discovery 690 (General Electric Medical Systems, Milwaukee, WI, USA), GE Advance (General Electric Medical Systems, Milwaukee, WI, USA), GE Signa 3T PET/MR (GE Healthcare, Chicago, IL, USA). A brain MRI scan (Philips Ingenuity TF PET/MR, Philips Medical Systems, Cleveland, OH, USA; GE Signa 3T PET/MR GE Healthcare, Chicago, IL, USA), including a 3D T1-weighted sequence, was taken to exclude significant structural changes and to be used in the PET data analysis (PET normalization). Participants were excluded from the analysis if MRI data were unavailable or if PET or MRI image quality did not meet the predefined criteria for reliable analysis. In addition, PET scans were excluded when scanner-related effects could not be sufficiently corrected using harmonization methods, resulting in residual variability that could introduce systematic bias into the results.

Preprocessing and analysis of PET data was carried out using an automated analysis pipeline Magia (https://github.com/tkkarjal/magia)[18]. The Magia toolbox is a standardized and fully automatic analysis pipeline for reproducible pre-processing and kinetic modelling of brain PET data, running on MATLAB (The MathWorks, Inc., Natick, MA, USA). Dynamic [^{11}C]PIB PET images were motion corrected and co-registered with the 3D T1-weighted single subject MRI images and normalized to the Montreal Neurological Institute space.

Automated Anatomical Labelling Atlas 3 (AAL3)[19] was applied to generate cortical gray matter and cerebellar cortex ROIs masks. From the 116 total regions present in the AAL3 atlas, 38 were selected because they are related to AD pathology. Based on literature, amyloid deposition is typically seen in cortical areas such as the precuneus, posterior and anterior cingulate, prefrontal cortex, parietal cortex, and lateral temporal cortex, which are the earliest regions in AD[6]. Medial temporal structures, including the entorhinal cortex, hippocampus, amygdala, and medial temporal cortex, were included because these areas are also linked to the early stages of AD and play an important role in memory decline[20,21]. In addition, subcortical regions such as the caudate, putamen, pallidum, striatum, and thalamus were selected to capture broader brain involvement, since amyloid pathology may also extend beyond cortical regions as the disease progresses[22]. Including both left and right hemisphere regions allowed possible asymmetries in amyloid distribution to be examined. The use of the AAL3 atlas provided a standardized way to define anatomical regions for ROI-based analysis of the [^{11}C]PIB PET images. See ROIs and lateralization in Supplementary Table 2.

Standardized uptake value ratios (SUVRs) were generated over the 60 to 90-minute scan duration by computing the ratio between the different brain target regions (each of the 38 ROIs selected from AAL3 atlas) and the cerebellar cortex which was used as reference region. The 60–90-minute post-injection window was selected because it represents a stable late phase of [^{11}C]PIB retention, minimizing early perfusion effects and providing reliable estimates of amyloid binding[3].

2.4. Data harmonization

To mitigate scanner-related variability in [^{11}C]PIB-PET imaging data while preserving the biological variation of interest, we applied the ComBat harmonization algorithm. The harmonization was implemented using the neuroCombat function from the neuroCombat R package (version 1.0.13). To assess the impact of harmonization, we visualized the Binding Potential (BPnd) estimates for each scanner in six different ROIs (Figure 1).

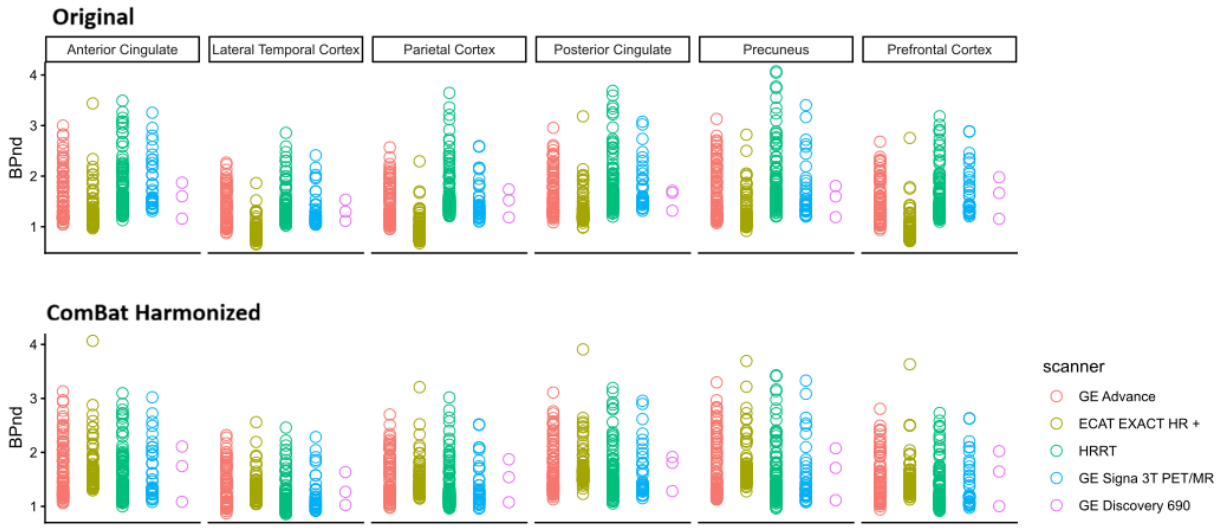

Figure 1*. Each panel shows Binding Potential (BPnd) values for five PET scanners across six ROIs. Circles represent individual observations. Before harmonization (above), systematic differences between scanners are visible across several ROIs. After ComBat harmonization (below), scanner-related variability is reduced while preserving regional differences in BPnd.*

2.5. Machine Learning Models

The data were classified using regional, literature-based and PCA-based PIB data with and without the demographic (age, sex and BMI) and clinical information (MMSE and APOE status). We first developed a baseline classification model using demographic information alone. Next, we constructed imaging-based models utilizing brain-imaging data and compared the classification performance of these models against the baseline model. Finally, we combined imaging-based models with the baseline model and additional clinical information to evaluate whether the inclusion of multimodal information improved classification accuracy (Table 2).

***Table 2**. Summary of the models, predictor variables, and outcome variables used in the analyses*

Model	**Variables**
Baseline model	Demographic information: Age, Sex and Body Mass Index (BMI)
Regional Pittsburgh Compound B (PIB) model	Imaging data, 38 Region of Interests (ROIs) chosen from Automated Anatomical Labelling Atlas 3 (AAL3 Atlas)[6,20,21,236,20,21,23]
Principal Component Analysis (PCA) based PIB model	Three principal components based on 38 selected ROIs and Kaiser's criterion (eigenvalue > 1)
Alzheimer's Disease Signature Based model (ADS)	The image data of 6 ROIs where amyloid accumulation is typically observed first[6]
Regional PIB Model + Baseline model	38 ROIs + Demographic information (Age, Sex and BMI)
PCA based PIB model + Baseline model	Three principal components + Demographic information (Age, Sex and BMI)
ADS + Baseline model	6 ROIs where amyloid accumulation is typically observed first[6] + Demographic information (Age, Sex and BMI)
Regional PIB model + Baseline model + Clinical information	38 ROIs + Demographic information (Age, Sex and BMI) + Clinical information (MMSE and APOE status)
PCA based PIB model + Baseline model + Clinical information	Three principal components + Demographic information (Age, Sex and BMI) + Clinical information (MMSE and APOE status)
ADS + Baseline model + Clinical information	6 ROIs where amyloid accumulation is typically observed first[6] + Demographic information (Age, Sex and BMI) + Clinical information (MMSE and APOE status)

2.6. Study pipeline and machine learning

The study pipeline and machine learning workflow are illustrated in Figure 2. The imaging data were first pre□processed, after which white matter (WM) regions were excluded prior to harmonization because PIB has non-specific uptake in WM[24]. Scanner-related variability was harmonized using the ComBat algorithm. After pre-processing, machine learning models were used to classify subjects into three diagnostic categories (HC, MCI, and AD) based on regional [^{11}C]PIB binding patterns. XGBoost was selected as the optimal classifier due to its robust regularization and proven efficacy in neuroimaging applications (Fernández-Blázquez et al., 2025). Model selection and final evaluation were performed using a 10-fold nested cross validation. Hyperparameters (inner loop) were tuned using Bayesian optimization with Optuna. For each model, 50 optimization trials were performed. During tuning, a 5-fold stratified cross validation procedure was used. The outer loop evaluated the tuned model on held out folds. For each

outer fold, predicted class probabilities and predicted labels were stored. After all folds were completed, predictions were concatenated to form a single out of sample prediction set for each model.

Model performance was evaluated using two primary metrics: the Average Weighted AUC score with Confidence Interval (95%) and the Weighted Cohen's Kappa derived from the confusion matrix, averaging across the cross-validation folds (https://pubmed.ncbi.nlm.nih.gov/19673146[3] /). All analyses were performed in Python(version 3.14) using NumPy, pandas, scikit learn, XGBoost, Optuna, and Seaborn packages.

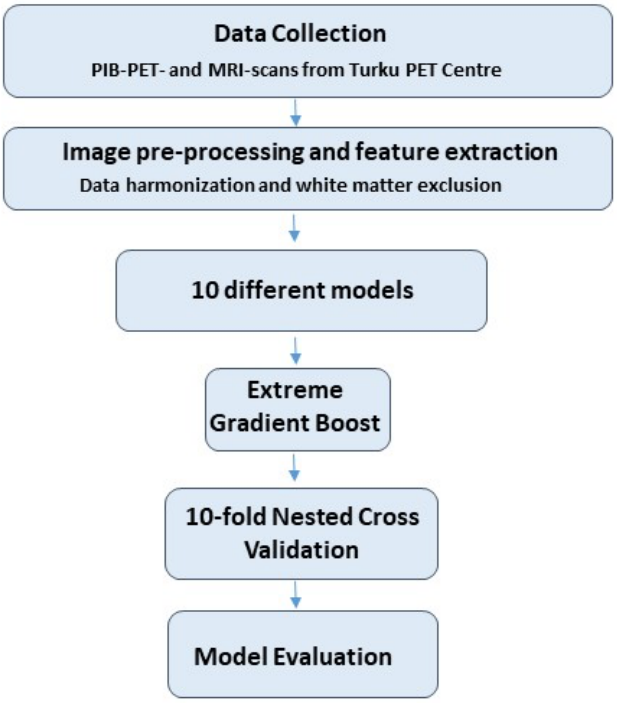

Figure 2. *Flow chart of the study pipeline. 1) (11)C-labeled Pittsburgh Compound-B (PIB-PET) and Magnetic Resonance Imaging (MRI) scans data collected in the Turku PET Centre (N=314) were 2) pre-processed, 3) analyzed parallel based on 10 different approaches to 4) select statistically classifier that predicts the best diagnoses given by clinicians. 5) Extreme Gradient Boost was chosen as the best classifier and 10-fold Nested cross-validation was performed to it by splitting whole data 10 times as 90 % / 10 % train/test split to ob*tain an unbiased estimate of model performance, and finally 6) *the best classifier was used in a model evaluation.*

3. Results

3.1. Descriptive statistics

The ε4 allele of the *APOE* gene significantly increases AD risk and APOE ε4 drives amyloid formation[25]. The percentage of APOE ε4 carriers was 38.6% in HC, 45% in MCI and 52% in AD group. In the HC group 1,6% of the participants were carriers of ε2 allele (that is known to be a protective against AD[25]), but no ε2 carries were seen in MCI and AD groups. See Table 1 for details.

Figure 3 shows the group mean SUVR images and the distribution of [^{11}C]PIB in the study participant groups. Across the study groups, [^{11}C]PIB uptake seems to be most salient in prefrontal, parietal and temporal cortex, and visually uptake clearly increases through study groups from HC to AD.

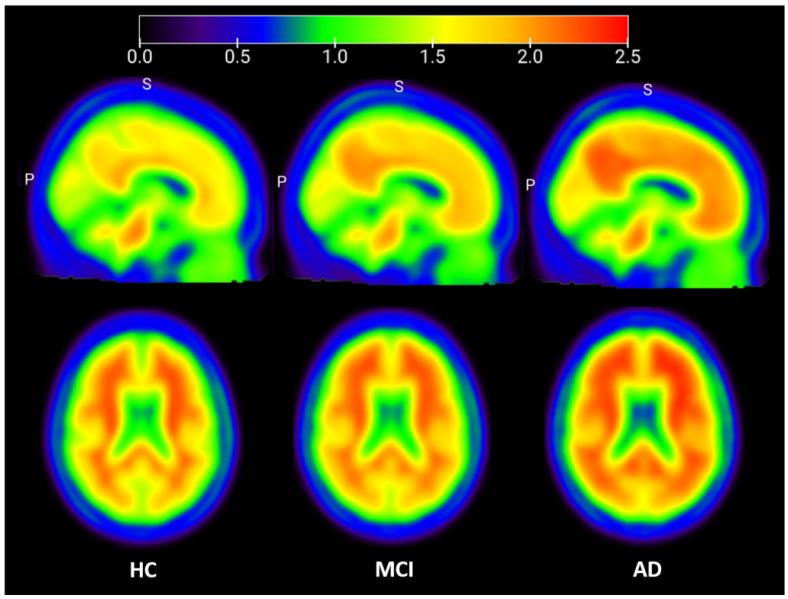

Figure 3. *The mean [^{11}C]PIB PET sagittal and axial images (scale from 0-2.5, MNI coordinates x=-6, y=-18 and z=-20) of the harmonized whole data set (N= 314) as its study groups healthy control (HC) (N=216), Mild Cognitive Impairment (MCI) (N=64) and Alzheimer's disease (AD) (N=34). The letters S and P indicate anatomical orientation: S = superior (up), P = posterior (back).*

Group-specific regional standardized uptake value ratios (SUVRs) are shown in Figure 4. In several regions, the binding levels demonstrated a clear graded pattern across the diagnostic spectrum, where the SUVRs increased progressively from the HC group to MCI and were highest in the AD group. Pairwise Mann-Whitney U tests confirmed statistically significant differences between the cohorts' groups in multiple regions; the most robust effects were typically observed when comparing the HC and AD groups while comparisons between MCI and AD yielded modest significance levels in the parietal cortex.

.

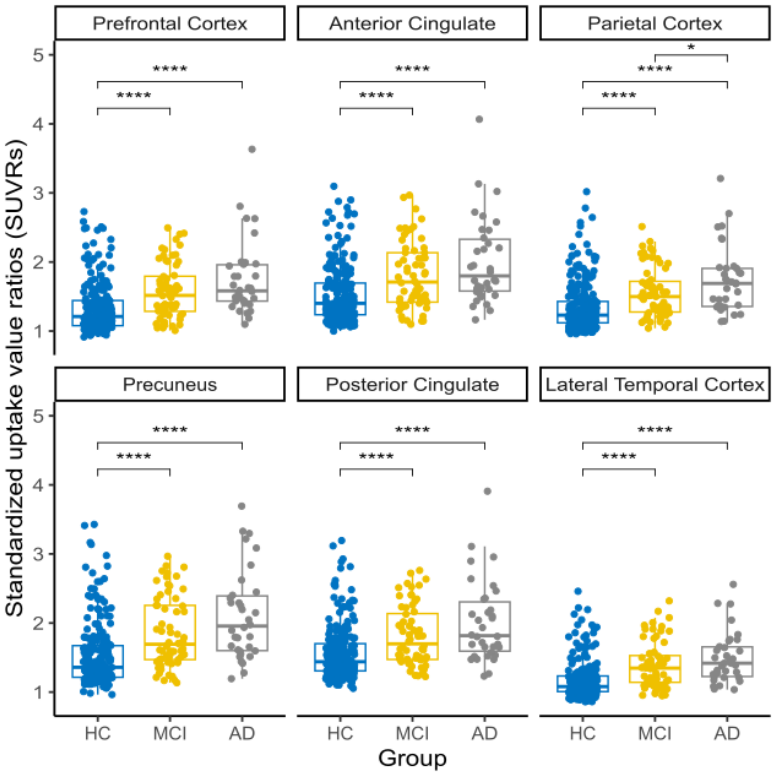

Figure 4. *Regional standardized uptake value ratios (SUVRs) across healthy control (HC), mild cognitive impairment (MCI), and Alzheimer's disease (AD) groups. Asterisks denote statistically significant pairwise differences identified via Mann-Whitney U tests: **** $p < 0.0001$, * $p < 0.05$.*

3.2. Comparison of AUC Across Different Approaches

We evaluated the predictive performance XGBoost model across three different groups (HC, MCI, and AD) by using 38 ROIs. Average area under curve (AUC) values with corresponding confidence intervals were reported for each model. This comparison allowed us to determine the relative effectiveness of each model in supporting accurate classification.

The baseline model demonstrated limited ability to classify groups, achieving an AUC of 0.59 (95% CI: 0.54–0.65). The regional PIB model alone achieved an AUC of 0.71 (95% CI: 0.65–0.71), while the PCA-based model yielded an AUC of 0.68 (95% CI: 0.62–0.72). The ADS model alone achieved an AUC of 0.71 (95% CI: 0.65–0.76). When combined with baseline model, the regional PIB, PCA, and ADS models achieved AUCs of 0.70 (95% CI: 0.64–0.74), 0.69 (95% CI: 0.63–0.73), and 0.69 (95% CI: 0.64–0.74), respectively. The highest predictive performance was observed when PIB imaging features were combined with baseline model and clinical information, with AUCs of 0.80 (95% CI: 0.76–0.85) for the regional PIB model, 0.79 (95% CI: 0.74–0.83) for the PCA model, and 0.80 (95% CI: 0.76–0.85) for the ADS model. (Figure 5.)

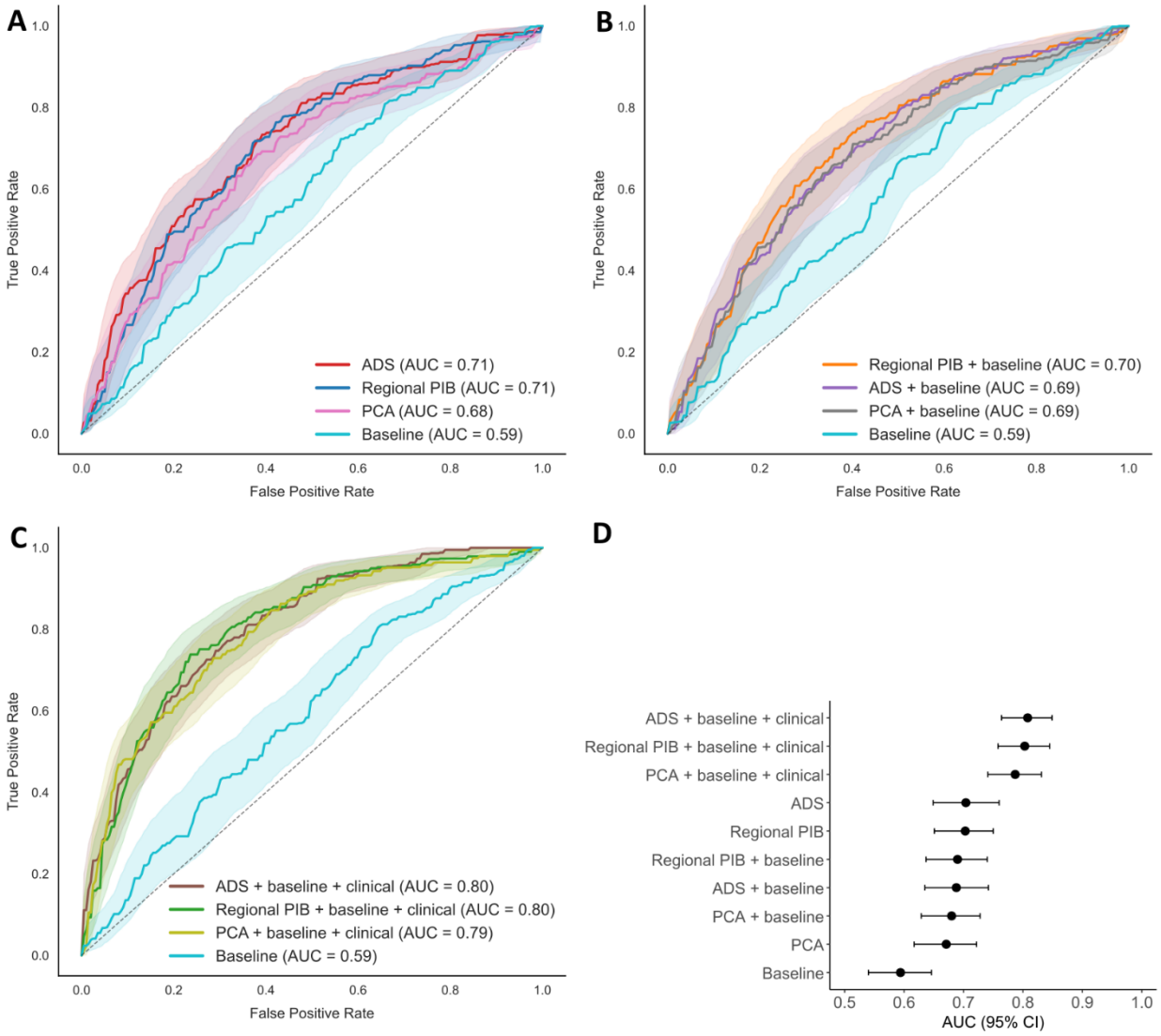

Figure 5*. Comparison of Area Under Curves (AUC) with 95% confidence intervals across Alzheimer's Disease Signature (ADS), Regional Pittsburgh Compound B (PIB) and Principal Component Analysis (PCA) models shown in A,B and C. Forest plot of model AUCs with 95% confidence intervals, arranged from best to worst D.*

Model performance was assessed using Accuracy, Precision, Recall, F1-score, Cohen's Kappa, and Macro-AUC. Macro-AUC represents the unweighted average of class-specific AUC values, providing a more balanced view in the presence of class imbalance (Table 3). Figure 5(D) shows the model-specific AUC values along with their 95% confidence intervals.

***Table 3**: Comparison of performance metrics for different models ranked by macro-AUC.*

Model	Accuracy	Precision	Recall	F1	Kappa	Macro-AUC
Alzheimer's Disease Signature (ADS) model with baseline model and clinical information	0.69	0.53	0.54	0.53	0.35	0.80 (CI: 0.76-0.85)
Regional Pittsburgh Compound B (PIB) model with baseline model and clinical information	0.71	0.53	0.51	0.52	0.37	0.80 (CI: 0.76-0.85)
Principal Component Analysis (PCA) model with baseline model and clinical information	0.70	0.58	0.63	0.60	0.42	0.79 (CI: 0.74-0.83)
ADS model	0.61	0.49	0.53	0.50	0.28	0.71 (CI: 0.65-0.76)
Regional PIB model	0.64	0.42	0.42	0.42	0.20	0.71 (CI: 0.65-0.75)
Regional PIB model with baseline model	0.61	0.41	0.42	0.41	0.20	0.70 (CI: 0.64-0.74)
ADS model with baseline model	0.60	0.45	0.47	0.45	0.22	0.69 (CI: 0.64-0.74)
PCA model with baseline model	0.57	0.43	0.45	0.43	0.22	0.69 (CI: 0.63-0.73)
PCA model	0.55	0.43	0.45	0.42	0.20	0.68 (CI: 0.62-0.72)
Baseline model	0.52	0.39	0.40	0.38	0.12	0.59 (CI: 0.54-0.65)

3.3 Statistical Evaluation of Differences in AUC

Bootstrap pairwise comparisons based on 10 000 resamples of the AUC revealed a clear hierarchical pattern of model performances. All models that combined clinical, baseline, and imaging-derived features significantly outperformed all other configuration ($p < 0.0001$), aligning with AUC-curves (Figure 4). The models that combine demographical information with PCA, Atlas, or ROI-based representations also showed significantly higher AUCs than all brain-ROI based models ($p < 0.0001$). In contrast, Regional PIB-, PCA-, and ADS -models did not differ significantly from one another, forming a lower-performing cluster. Baseline alone yielded the poorest performance. The full 10×10 bootstrap p-value matrix is shown in Figure 7.

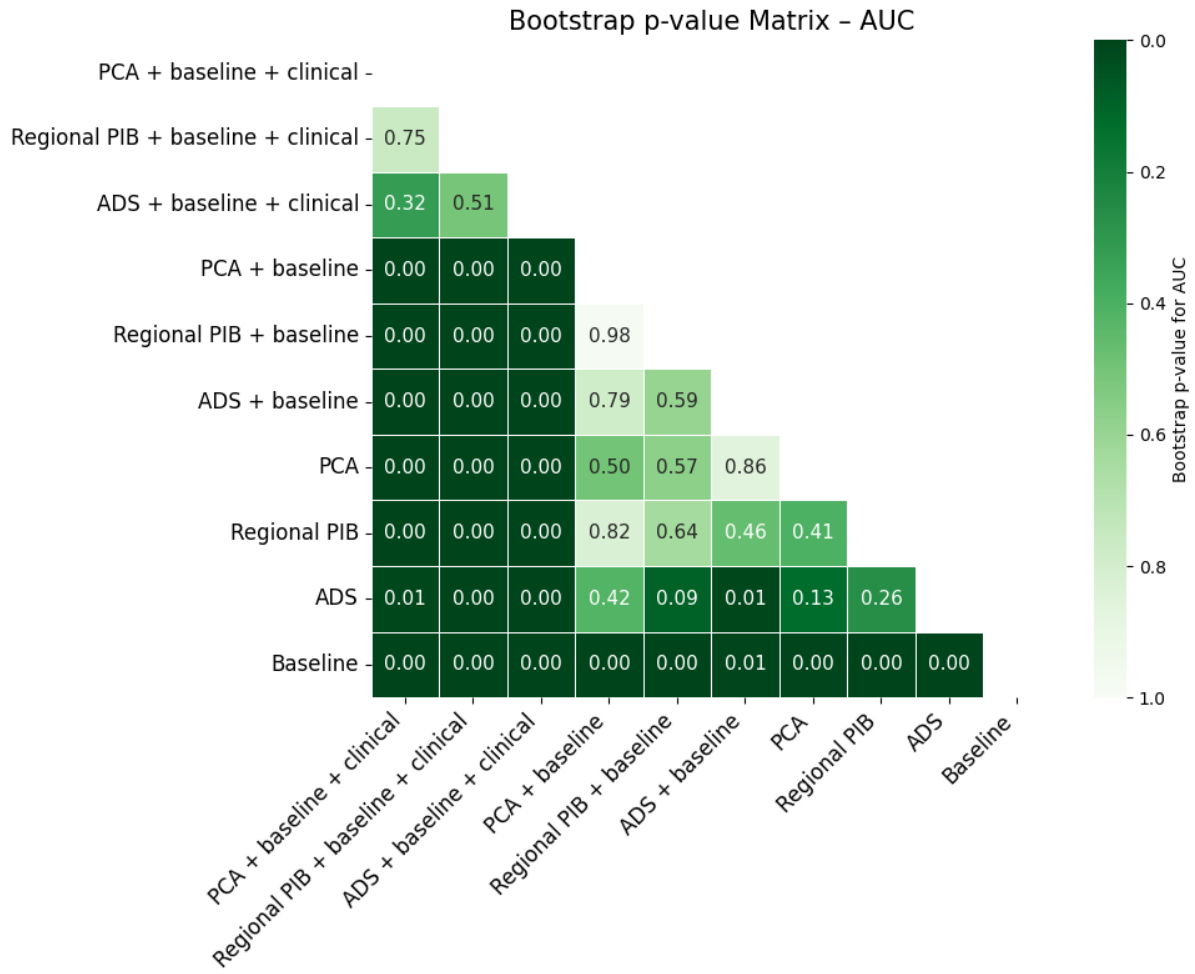

Figure 7. *Bootstrap-derived pairwise p-values for Area Under Curve (AUC) differences. Two-sided p-values were obtained from 10 000 bootstrap resamples of the fold-level AUC differences between all model pairs. The matrix is visualized as a lower-triangle heatmap with the diagonal removed. The color scale is inverted (dark-to-light green) to emphasize lower p-values. Abbreviations: Alzheimer's Disease Signature (ADS), Regional Pittsburgh Compound B (PIB) and Principal Component Analysis (PCA)*

4. Discussion

In our study, we demonstrate that machine learning applied to amyloid PET imaging can successfully distinguish between healthy controls (HC), mild cognitive impairment (MCI), and Alzheimer's disease (AD) within the local PIB PET cohort (including research studies). Importantly, our findings highlight that combining an imaging biomarker with readily available clinical information yields the most accurate classification performance. This work contributes to the growing effort to translate imaging-based biomarkers and machine learning approaches into clinically applicable diagnostic tools, particularly in heterogeneous datasets that better reflect routine clinical practice.

When comparing different models, PCA did not perform better than other models, adding baseline characteristics did not perform better than using the PIB data only. Numerically highest AUC values were obtained when PIB data was combined with baseline and clinical information leading to the most accurate classification of the three diagnoses in the Alzheimer's disease continuum. In particular, amyloid pathology derived from PIB-PET imaging, representing biological information, plays a central role in improving classification accuracy when combined with basic demographic variables such as age, sex, and BMI. These findings are in line with previous literature, which emphasizes the importance of integrating biological and clinical information for more accurate characterization and prediction of AD[26]. In addition, reducing the imaging data into three components using PCA did not improve the classification accuracy.

Dimensionality reduction using PCA did not improve classification accuracy compared to models using the full set of 38 regional PIB variables. However, the comparable model performance suggests that most of the relevant information was captured within a small number of underlying dimensions. While PCA-based approach did not enhance predictive accuracy, it offers practical advantages such as reduced data complexity, and improved computational efficiency and potential gains in model generalizability. Therefore, PCA remains a useful approach, especially in clinical contexts where simpler and more interpretable models are desirable.

Our findings are broadly aligned with previous machine learning studies utilizing amyloid PET imaging, though differences in methodology and datasets influence the reported performance to compare with. For example, Jiang et al., 2015[27] showed predictive power, with 88% accuracy in HC, MCI and AD classification, based on [^{11}C]PIB PET imaging technique by using PCA-reduced features for AD early-stage detection by using Alzheimer's disease neuroimaging initiative (ADNI) dataset, multiclass and voxel based analysis. In that study, PCA and Support Vector Machine (SVM) were combined to develop computer-aided diagnostic tools. The results highlighted increased amyloid accumulation in the posterior cingulate gyrus and precuneus, as well as the lateral-temporal region, mesial-temporal cortex, frontal temporal, occipital regions that occurred also in our results. In addition, previous literature supports amyloid accumulation typically appears in parietal cortex, anterior cingulum, posterior cingulum, precuneus and lateral temporal cortex[28–31] which were also shown in our results.

However, an important distinction lies in dataset characteristics. Jiang et al., 2015[27] used the ADNI dataset in their study instead we used a local cohort. Compared to a local cohort, ADNI data may present challenges for real-world generalization due to potential spectrum bias and variability across multi-site imaging protocols. In addition, its widespread reuse requires careful data handling to avoid unintended data leakage. However, ADNI also provides a well-characterized and standardized dataset that supports reproducibility and comparability across studies.[2,12,32]

Differences in classification tasks also need to be considered when comparing results across the studies. Studies focusing on binary classification (e.g. amyloid-positive vs. amyloid-negative, or AD vs. HC) typically achieve higher accuracy than multiclass models. For instance, Ladefoged et al., 2023[32] performed the study with a local memory clinic cohort that included cognitively impaired patients (n=1309) using [^{11}C]PIB PET/CT and PET/MRI scans. They achieved 98% accuracy when classifying amyloid status by employing a voxel-based approach for analyzing PET images. They developed convolutional neural networks (CNNs)

that directly processed entire PET image volumes to predict SUVRs and classify amyloid status in a binarized category. However, their cohort did not include healthy controls, nor did they use their model to classify patients into diagnostic groups (HC, MCI, and AD) as we did. Instead, their approach focused on binary amyloid status (amyloid-positive vs. amyloid-negative), which is inherently a less complex task. Similarly, Yan et al., 2018[33] used Support Vector Machine models with PET-derived features to classify AD and predict MCI conversion, achieving relatively high accuracies 93.6% for AD vs. HC and 88.9% for progressive vs. stable MCI. In contrast, our all-vs-all multiclass framework (HC vs MCI vs AD) translates better to a heterogeneous clinical population which it could be applied in the future. Comparing the two studies, Yan et al., 2018[33] vs present study, the main differences were: 1) multi-site vs single site; 2) multi-tracer vs single tracer; 3) scanner differences apparently not harmonized (although in ADNI these are partially accounted on, residual scanner effects remain in the results) vs harmonization with ComBat. An important methodological consideration when comparing our results to studies such as Yan et al., 2018[33] is the handling of inter-scanner (and inter-tracer, but not necessary in our study) variability. If such differences are not explicitly harmonized, machine learning models may inadvertently exploit scanner- or tracer-specific features rather than disease-related biological patterns. This can lead to artificially inflated classification performance, particularly if diagnostic groups are unevenly distributed across acquisition protocols. In contrast, our study applied harmonization procedures to mitigate scanner effects, making the classification task more challenging but also improving the validity and generalizability of the results.

The added value of multimodal integration is further supported by studies combining neuroimaging and clinical variables. Choi et al., 2025[34] developed a ML based models to predict amyloid-beta (Aβ) positivity using clinical and neuroimaging-related variables. They identified the most influential predictive features contributing to Aβ classification performance across diagnostic group (controls, MCI, dementia). Combining early-phase (0-10 min after injection) ^{18}F-Florbetaben PET and clinical data improved prediction of amyloid-beta positivity. Early-phase PET alone showed moderate performance for Aβ prediction with 80.6% accuracy by using Random Forest In the study, the best-performing model combined clinical and PET imaging data by using Gradient Boosting achieved 88.9% accuracy. The findings of the study support the clinical value of PET imaging and demonstrate the synergistic potential of combining neuroimaging with clinical data in AD assessment as our results did by increasing the accuracy when combining imaging data together with demographic and clinical information. Of note is that the early frames of ^{18}F-Florbetaben (and other amyloid PET tracers) reflect blood flow and correlate with ^{18}F-luorodeoxyglucose uptake (see Choi et al., 2025[34]) whereas later time point uptake reflects better beta-amyloid accumulation[35,36]. Their study population was relatively small compared to our cohort (176 vs. 314 participants), which may limit the generalizability and robustness of the machine learning models. A larger dataset provides greater variability across cognitive groups, reduces the risk of overfitting, and improves the model's ability to identify clinically relevant features more reliably. On the other hand, the diagnostic groups in their study cohort (HC 38 vs 214, MCI 94 vs 64, and AD 44 vs 34) were more balanced than in our dataset, which may have reduced class imbalance bias and supported more stable model training across diagnostic categories. In our cohort, the uneven distribution between groups may have influenced classifier performance and sensitivity toward the majority class. In another study similar to ours, Yi et al., 2023[37] demonstrated that ML models using amyloid-related biomarkers and clinical variables can accurately classify HC, MCI, and AD, and that integrating multimodal information improves

classifiers' performance. However, while their work focused on constructing an interpretable framework using Shapley additive explanation (SHAP) values and algorithmic strategies to handle class imbalance, our study evaluated both ROI-based and PCA-reduced PIB-PET features within a local, clinically representative cohort. In contrast to their large multi-site ADNI dataset, our cohort required scanner harmonization and better reflects typical clinical variability.

From a methodological perspective, ROI-based approaches, as used in our study, offer several advantages compared to voxel methods. While voxel-based machine learning provides high spatial resolution and enables deeper data-driven feature selection[38,33], ROI-based machine learning methods offer advantages in predicting AD, especially using the PCA method that summarizes information within each ROI into a small number of components, enabling noise reduction while retaining the most informative features, and thus reduces overfitting, thereby making it a robust alternative to voxel-based analyses[39,34].

One major strength is the integration of multimodal information, demonstrating that combining amyloid pathology from PIB-PET with clinical information yields the most accurate classifications. In addition, reducing the imaging data into three components using PCA did not statistically degrade the classification accuracy. Another strength is that our MCI population includes amnestic MCI and not all-type MCI. However, amnestic MCI patients are more often amyloid positive than non-amnestic ones[40], which on the other hand might have reduced possibility to separate our MCI and AD groups. Furthermore, in accordance with this the conversion rate to AD is higher in the amnestic vs. non-amnestic MCI patients[41].

This study had certain limitations, for example PIB-PET scans were not performed with the same PET scanner which is why the SUV-values needed to be harmonized before using machine learning which might have an impact on the results, possibly risking partial loss of biological variance. On the other hand, the difference between PET scanners represents a real-world situation because hospitals usually have PET scanners from different vendors. In addition, our cohort wasn't balanced in terms of groups, especially the HC group was larger than MCI and AD group. Our HC group included amyloid positive study subjects that maybe affected bias on the classifier results by mixing some of amyloid positive HC into either MCI or AD group. Using amyloid PET-imaging alone to totally separate HC, MCI and AD encompass inherent challenge, since a considerable proportion on cognitively unimpaired elderly are amyloid positive (evaluated by cerebrospinal fluid or amyloid PET) and this proportion increases with age and the presence of APOE ε4 positivity[42]. Furthermore, incorporating clinical information was important for achieving higher classification accuracy.

Despite the limitations, our study showed that it is possible to classify study subjects into three different diagnostic groups by using in-house imaging dataset together and separately from demographic and clinical information. The approach developed in our study could characterize an individual's amyloid burden for diagnostic purposes, for example in drug development clinical trials to see the efficacy of the drug candidate. SUVR-based models could also increase confidence in visual reading in the clinic when evaluating brain amyloid load[7] and coupled to ML techniques like hereby presented. In future studies, voxel-wise analysis could provide more accurate results than ROI-based analyzes. In addition, balancing study groups by adding more data would be considered. Furthermore, a potential direction for future research would be to evaluate the proposed approach in a more diverse and clinically representative

population, such as patients from a memory disorder center. In the present study, the cohort may not fully capture the variability seen in real-world diagnostic settings. Therefore, applying this method to a diagnostic patient cohort could provide further insight into its generalizability and clinical utility. Testing the approach in populations with varying stages and types of cognitive impairment would help assess its robustness and potential for practical implementation.

Conclusion

Alzheimer's Disease Signature, Regional Pittsburgh Compound B and Principal Component Analysis approaches separately and combined with demographic information and clinical information succeeded in classifying study subjects into three different diagnosis classes. Thus, our pipeline in amyloid-PET image classification could be a useful tool to predict MCI and AD, and increase confidence in visual reading when evaluating amyloid positivity and clinicians' confidence in diagnosing MCI and AD. Furthermore, the approach could be a valuable tool in drug development in longitudinal studies when monitoring the effectiveness of anti-amyloid drugs, and to identify individuals with amyloid pathology in the brain with no or minimal cognitive symptoms.

References

1. Deture MA, Dickson DW. The neuropathological diagnosis of Alzheimer's disease. *Mol Neurodegener*. *BioMed Central Ltd.* 2019;14(1). doi:10.1186/s13024-019-0333-5
2. Jack CR, Bennett DA, Blennow K, et al. NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. *Alzheimer's and Dementia*. *Elsevier Inc.* 2018;14(4):535-562. doi:10.1016/j.jalz.2018.02.018
3. Lopresti BJ, Klunk WE, Mathis CA, et al. Simplified quantification of Pittsburgh Compound B amyloid imaging PET studies: a comparative analysis. *J Nucl Med*. 2005;46(12):1959-1972.
4. Mountz JM, Laymon CM, Cohen AD, et al. Comparison of qualitative and quantitative imaging characteristics of [11C]PiB and [18F]flutemetamol in normal control and Alzheimer's subjects. *Neuroimage Clin*. 2015;9:592-598. doi:10.1016/j.nicl.2015.10.007
5. Klunk WE, Engler H, Nordberg A, et al. Imaging brain amyloid in Alzheimer's disease with Pittsburgh Compound-B. *Ann Neurol*. 2004;55(3):306-319. doi:10.1002/ana.20009
6. Braak H, Braak E. Frequency of Stages of Alzheimer-Related Lesions in Different Age Categories. *Neurobiol Aging*. 1997;18(4):351-357. doi:10.1016/S0197-4580(97)00056-0
7. Bucci M, Chiotis K, Nordberg A. Alzheimer's disease profiled by fluid and imaging markers: tau PET best predicts cognitive decline. *Mol Psychiatry*. 2021;26(10):5888-5898. doi:10.1038/s41380-021-01263-2
8. Diogo VS, Ferreira HA, Prata D. Early diagnosis of Alzheimer's disease using machine learning: a multi-diagnostic, generalizable approach. *Alzheimers Res Ther*. 2022;14(1). doi:10.1186/s13195-022-01047-y
9. Mohsen S. Alzheimer's disease detection using deep learning and machine learning: a review. *Artif Intell Rev*. 2025;58(9):262. doi:10.1007/s10462-025-11258-y
10. de Vries BM, Golla SS V., Ebenau J, et al. Classification of negative and positive 18F-florbetapir brain PET studies in subjective cognitive decline patients using a convolutional neural network. *Eur J Nucl Med Mol Imaging*. 2021;48(3):721-728. doi:10.1007/s00259-020-05006-3
11. Kavitha C, Mani V, Srividhya SR, Khalaf OI, Tavera Romero CA. Early-Stage Alzheimer's Disease Prediction Using Machine Learning Models. *Front Public Health*. 2022;10. doi:10.3389/fpubh.2022.853294
12. Samper-González J, Burgos N, Bottani S, et al. Reproducible evaluation of classification methods in Alzheimer's disease: Framework and application to MRI and PET data. *Neuroimage*. 2018;183:504-521. doi:10.1016/j.neuroimage.2018.08.042
13. Campbell MC, Markham J, Flores H, et al. Principal component analysis of PiB distribution in Parkinson and Alzheimer diseases. *Neurology*. 2013;81(6):520-527. doi:10.1212/WNL.0b013e31829e6f94
14. Cummings J. Anti-Amyloid Monoclonal Antibodies are Transformative Treatments that Redefine Alzheimer's Disease Therapeutics. *Drugs*. 2023;83(7):569-576. doi:10.1007/s40265-023-01858-9
15. Petersen RC, Doody R, Kurz A, et al. Current Concepts in Mild Cognitive Impairment. *Arch Neurol*. 2001;58(12):1985. doi:10.1001/archneur.58.12.1985
16. McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease. *Neurology*. 1984;34(7):939-939. doi:10.1212/WNL.34.7.939

17. Kemppainen NM, Aalto S, Wilson IA, et al. Voxel-based analysis of PET amyloid ligand [^{11} C]PIB uptake in Alzheimer disease. *Neurology*. 2006;67(9):1575-1580. doi:10.1212/01.wnl.0000240117.55680.0a
18. Karjalainen T, Tuisku J, Santavirta S, et al. Magia: Robust Automated Image Processing and Kinetic Modeling Toolbox for PET Neuroinformatics. *Front Neuroinform*. 2020;14. doi:10.3389/fninf.2020.00003
19. Rolls ET, Huang CC, Lin CP, Feng J, Joliot M. Automated anatomical labelling atlas 3. *Neuroimage*. 2020;206:116189. doi:10.1016/j.neuroimage.2019.116189
20. Braak H, Braak E. Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathol*. 1991;82(4):239-259. doi:10.1007/BF00308809
21. Stouffer KM, Grande X, Düzel E, et al. Amidst an amygdala renaissance in Alzheimer's disease. *Brain*. 2024;147(3):816-829. doi:10.1093/brain/awad411
22. Cho SH, Shin JH, Jang H, et al. Amyloid involvement in subcortical regions predicts cognitive decline. *Eur J Nucl Med Mol Imaging*. 2018;45(13):2368-2376. doi:10.1007/s00259-018-4081-5
23. Cho SH, Shin JH, Jang H, et al. Amyloid involvement in subcortical regions predicts cognitive decline. *Eur J Nucl Med Mol Imaging*. 2018;45(13):2368-2376. doi:10.1007/s00259-018-4081-5
24. Zeydan B, Schwarz CG, Lowe VJ, et al. Investigation of white matter PiB uptake as a marker of white matter integrity. *Ann Clin Transl Neurol*. 2019;6(4):678-688. doi:10.1002/acn3.741
25. Raulin AC, Doss S V., Trottier ZA, Ikezu TC, Bu G, Liu CC. ApoE in Alzheimer's disease: pathophysiology and therapeutic strategies. *Mol Neurodegener*. *BioMed Central Ltd*. 2022;17(1). doi:10.1186/s13024-022-00574-4
26. Jack CR, Andrews JS, Beach TG, et al. Revised criteria for diagnosis and staging of Alzheimer's disease: Alzheimer's Association Workgroup. *Alzheimer's & Dementia*. 2024;20(8):5143-5169. doi:10.1002/alz.13859
27. Jiang J, Shu X, Liu X, Huang Z. A Computed Aided Diagnosis tool for Alzheimer's disease based on 11C-PiB PET imaging technique. In: *2015 IEEE International Conference on Information and Automation*. IEEE; 2015:1963-1968. doi:10.1109/ICInfA.2015.7279610
28. Ekblad LL, Johansson J, Helin S, et al. Midlife insulin resistance, *APOE* genotype, and late-life brain amyloid accumulation. *Neurology*. 2018;90(13). doi:10.1212/WNL.0000000000005214
29. Toppala S, Ekblad LL, Tuisku J, et al. Association of Early β-Amyloid Accumulation and Neuroinflammation Measured With [^{11} C]PBR28 in Elderly Individuals Without Dementia. *Neurology*. 2021;96(12). doi:10.1212/WNL.0000000000011612
30. Pietilä E, Snellman A, Tuisku J, et al. Midlife insulin resistance, APOE genotype, and change in late-life brain beta-amyloid accumulation – A 5-year follow-up [11C]PIB-PET study. *Neurobiol Dis*. 2024;190. doi:10.1016/j.nbd.2023.106385
31. Grothe MJ, Barthel H, Sepulcre J, et al. In vivo staging of regional amyloid deposition. *Neurology*. 2017;89(20):2031-2038. doi:10.1212/WNL.0000000000004643
32. Ladefoged CN, Anderberg L, Madsen K, et al. Estimation of brain amyloid accumulation using deep learning in clinical [11C]PiB PET imaging. *EJNMMI Phys*. 2023;10(1). doi:10.1186/s40658-023-00562-7

33. Yan Y, Somer E, Grau V. Classification of amyloid PET images using novel features for early diagnosis of Alzheimer's disease and mild cognitive impairment conversion. *Nucl Med Commun*. 2019;40(3):242-248. doi:10.1097/MNM.0000000000000953
34. Choi DH, Ahn SH, Chung Y, Kim JS, Jeong JH, Yoon HJ. Machine learning model for predicting Amyloid-β positivity and cognitive status using early-phase 18F-Florbetaben PET and clinical features. *Sci Rep*. 2025;15(1):21987. doi:10.1038/s41598-025-00743-7
35. Becker GA, Ichise M, Barthel H, et al. PET Quantification of 18 F-Florbetaben Binding to β-Amyloid Deposits in Human Brains. *Journal of Nuclear Medicine*. 2013;54(5):723-731. doi:10.2967/jnumed.112.107185
36. Yoon HJ, Kim BS, Jeong JH, et al. Dual-phase 18F-florbetaben PET provides cerebral perfusion proxy along with beta-amyloid burden in Alzheimer's disease. *Neuroimage Clin*. 2021;31:102773. doi:10.1016/j.nicl.2021.102773
37. Yi F, Yang H, Chen D, et al. XGBoost-SHAP-based interpretable diagnostic framework for alzheimer's disease. *BMC Med Inform Decis Mak*. 2023;23(1):137. doi:10.1186/s12911-023-02238-9
38. Klyuzhin IS, Fu JF, Hong A, et al. Data-driven, voxel-based analysis of brain PET images: Application of PCA and LASSO methods to visualize and quantify patterns of neurodegeneration. *PLoS One*. 2018;13(11). doi:10.1371/journal.pone.0206607
39. Mwangi B, Tian TS, Soares JC. A review of feature reduction techniques in Neuroimaging. *Neuroinformatics*. *Humana Press Inc.* 2014;12(2):229-244. doi:10.1007/s12021-013-9204-3
40. Wolk DA, Price JC, Saxton JA, et al. Amyloid imaging in mild cognitive impairment subtypes. *Ann Neurol*. 2009;65(5):557-568. doi:10.1002/ana.21598
41. Fischer P, Jungwirth S, Zehetmayer S, et al. Conversion from subtypes of mild cognitive impairment to Alzheimer dementia. *Neurology*. 2007;68(4):288-291. doi:10.1212/01.wnl.0000252358.03285.9d
42. Jansen WJ, Ossenkoppele R, Knol DL, et al. Prevalence of Cerebral Amyloid Pathology in Persons Without Dementia. *JAMA*. 2015;313(19):1924. doi:10.1001/jama.2015.4668

Supplementary

Table 1. *The list of original publications whose data is used in the current study*

Authors	**Year**	**Journal**
Kemppainen N, Aalto S, Wilson I, Någren K, Helin S, Brück A, Oikonen V, Kailajärvi M, Scheinin M, Viitanen M, Parkkola R, Rinne JO.	2006	Neurology
Scheinin NM, Aalto S, Koikkalainen J, Lötjönen J, Karrasch M, Kemppainen N, Viitanen M, Någren K, Helin S, Scheinin M, Rinne JO.	2009	Neurology
Scheinin NM, Aalto S, Kaprio J, Koskenvuo M, Räihä I, Rokka J, Hinkka-Yli-Salomäki S, Rinne JO.	2011	Neurology
Koivunen J, Scheinin N, Virta JR, Aalto S, Vahlberg T, Någren K, Helin S, Parkkola R, Viitanen M, Rinne JO.	2011	Neurology
Joutsa J, Rinne JO, Hermann B, Karrasch M, Anttinen A, Shinnar S, Sillanpää M	2017	JAMA Neurology
Leinonen V, Rauramaa T, Johansson J, Bottelbergs A, Tesseur I, van der Ark P, Pemberton D, Koivisto AM, Jääskeläinen JE, Hiltunen M, Herukka SK, Blennow K, Zetterberg H, Jokinen P, Rokka J, Helin S, Haaparanta-Solin M, Solin O, Okamura N, Kolb HC, Rinne JO.	2018	Journal of Alzheimer´s disease
Ekblad L, Johansson J, Helin S, Viitanen M, Laine H, Pauli P, Jula A, Rinne JO.	2018	Neurology
Lindgren N, Kaprio J, Karjalainen T, Ekblad L, Helin S, Karrasch M, Teuho J, Rinne JO , Vuoksimaa E	2021	Neurobiology of Aging
Snellman A, Ekblad LL, Tuisku J, Koivumäki M, Ashton NJ, Lantero-Rodriguez J, Karikari TK, Helin S, Bucci M, Löyttyniemi E, Parkkola R, Karrasch M, Schöll M, Zetterberg H, Blennow K, Rinne JO	2023	Alzheimer´s Research and Therapies
Koivumäki M, Ekblad L, Lantero-Rodriguez J, Ashton NJ, Karikari TK, Helin S, Parkkola R, Lötjönen J, Zetterberg H, Blennow K, Rinne JO, Snellman A	2024	Alzheimer´s Research and Therapies

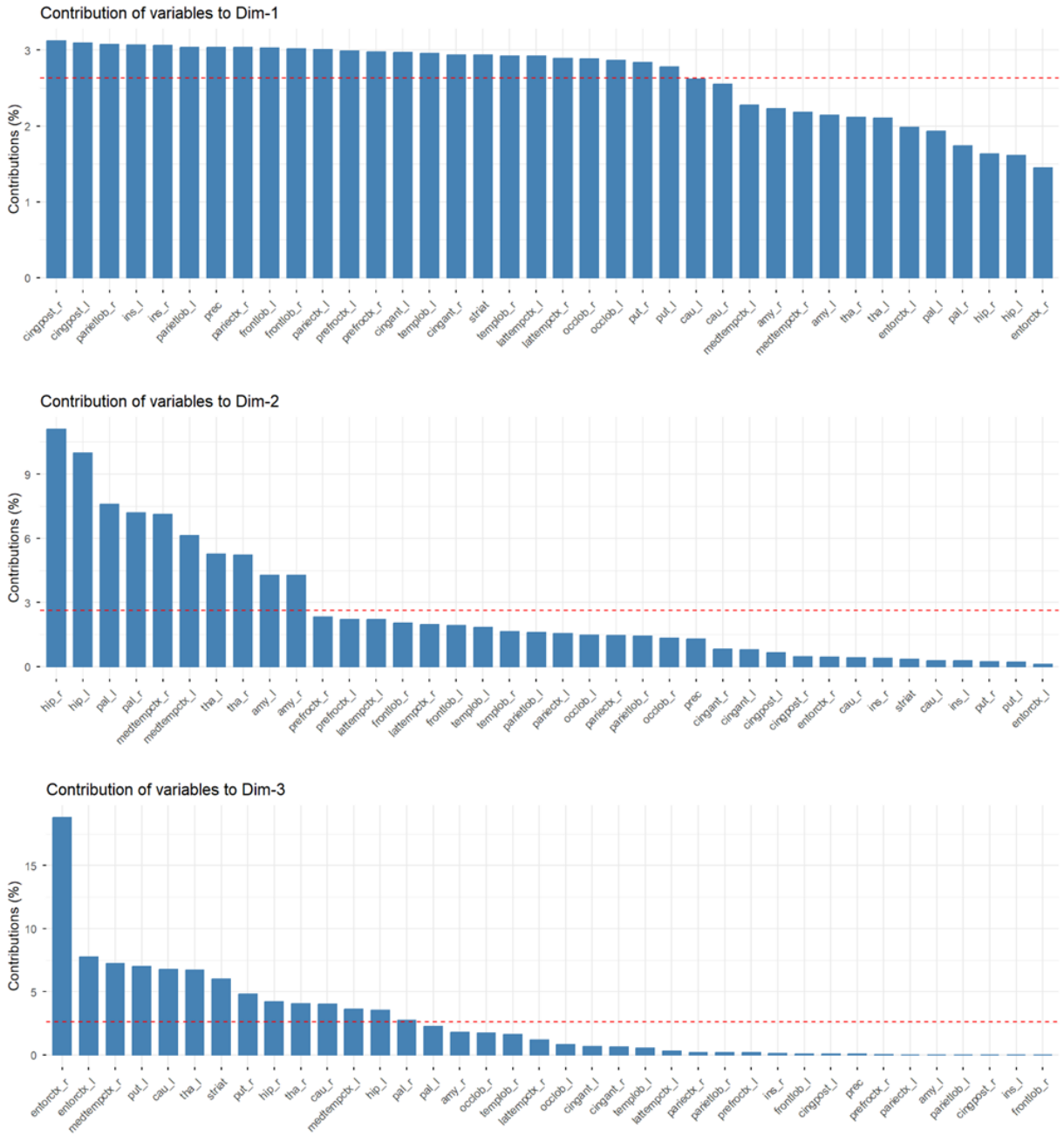

Figure 2. *Contribution of individual variables to the first three principal components (Dim-1, Dim-2, and Dim-3) of 38 ROIs. Bars represent the percentage contribution of each variable, and the red dashed line indicates the threshold for variables contributing above the average expected level. Different sets of brain regions dominate each dimension, highlighting distinct underlying patterns in the model. Abbreviations in Supplementary Table 2.*

Table 2. *Regions of interest (ROIs) used in analyzes.*

ROIs	Abbrevation
Amygdala left	amy_l
Amygdala right	amy_r
Anterior cingulate left	cingant_l
Anterior cingulate right	cingant_r
Caudate left	cau_l
Caudate right	cau_r
Entorhinal cortex left	entorctx_l
Entorhinal cortex right	entorctx_r
Frontal lobe left	frontlob_l
Frontal lobe right	frontlob_r
Hippocampus left	hip_l
Hippocampus right	hip_r
Insula left	ins_l
Insula right	insl_r
Lateral temporal cortex left	lattempctx_l
Lateral temporal cortex right	lattempctx_r
Medial temporal cortex left	meditempctx_l
Medial temporal cortex right	meditempctx_r
Occipital lobe left	occilob_l
Occipital lobe right	occilob_r
Pallidum left	pal_l
Pallidum right	pal_r
Parietal cortex left	parietctx_l
Parietal cortex right	parietctx_r
Parietal lobe left	parietlob_l
Parietal lobe right	parietlob_r
Posterior cingulate left	cingpost_l
Posterior cingulate right	cingpost_r
Prefrontal cortex left	prefroctx_l
Prefrontal cortex right	prefroctx_r
Precuneus	prec
Putamen left	put_l
Putamen right	put_r
Striatum	striat
Temporal lobe left	templob_l
Temporal lobe right	templob_r
Thalamus left	tha_l
Thalamus right	tha_r