university of groningen

faculty of science and engineering

Feature Extraction Methods for 3D FDG-PET Brain Scans for the Diagnosis of Neurodegenerative Disorders

August 25, 2025

Author: Luuk Westerhoek
Student Number: S4151674
First supervisor: Michael Biehl
Second supervisor: Kerstin Bunte
Daily supervisor: Sofie Lövdal

Abstract

This thesis presents a comparative analysis of feature extraction methods for neurodegenerative disease diagnosis using 3D fluorodeoxyglucose positron emission tomography (FDG-PET) brain imaging. We compare Principal Component Analysis (PCA) and Region-of-Interest (ROI) aggregation, using 10-fold stratified cross-validation for consistent comparison between the methods. Additionally, we investigate the reconstruction accuracy of both methods and between different atlases. For the ROI method, we also examine the disease-specific regions to determine optimal atlas selection for specific diseases. Results show PCA's great performance in information retention, through reconstruction, and classification performance. For ROI aggregation, despite higher reconstruction errors, shows comparable or even slightly better classification results than PCA. This result indicates that anatomical dimensionality reduction can very well capture metabolic patterns even with significant information loss.

Contents

1 Introduction

1.1 Background and Motivation

Cases of neurodegenerative diseases such as Alzheimer's and Parkinson's disease are increasing globally, affecting millions of people and putting a heavy load on our healthcare systems [1]. Alzheimer's disease alone affects over 50 million people and Parkinson's disease impacts another 10 million, worldwide [2]. These conditions should be diagnosed precisely and on time to evaluate the right treatment plan and patient care.

[^{18}F]Fluorodeoxyglucose positron emission tomography (FDG-PET) directly tracks the uptake of glucose in the brain [3, 4], which is directly effected by neurodegenerative disease [5, 6]. Therefore, it can be noted as one of the most effective methods for metabolic activity mapping in the brain. Since metabolic changes often precede observable anatomical changes [7], we can detect diseases like Alzheimer's and Parkinson's much earlier using FDG-PET [8]. However, there is a drawback; the high dimensionality of these scans makes automated analysis and classification way to computationally expensive and therefore, we need to reduce this high dimensionality.

In this thesis, two types of feature extraction methods will be discussed to reduce the high dimensional FDG-PET data. A data-driven technique using Principal Component Analysis (PCA) [9] and an anatomy based technique using Region-of-Interest (ROI) aggregation [10]. Each of these approaches is beneficial in different ways to reduce the computational complexity and to capture important features that are relevant to the diagnosis of disease. Such a pipeline may only perform as well as the features given as input. This work evaluates what types of features are the most informative to distinguish between different neurodegenerative diseases.

Such an assessment is relevant as it responds to the important issue of developing effective and reproducible diagnostic systems that can be applied in clinical practice [11].

1.2 Problem Statement

Each FDG-PET scan creates a high-dimensional 3D image of the brain containing spatial information about its metabolic activity. Yet, this high dimensionality, referring to the large number of voxels ($\approx 10^5$) in the image , causes issues when trying to implement and execute automated data analysis and classification, especially with limited scans available. In addition, a part of the full data set (reference set) needs to be held out only for PCA to prevent overfitting.

One way to tackle this problem of high dimensionality is by using dimensionality reduction, where the original data is transformed from a high dimensional space into a low dimensional space and the low-dimensional representation preserves most of the critical properties of the original data. Principal Component Analysis, or PCA, is such a technique used in neuroimaging [12]. It extracts the principal components by identifying patterns in the data of

highest variance, making the size of the data more manageable. Even though it is a straightforward approach and computationally efficient, it assumes these patterns can be captured by linear relationships between variables and fails to capture more complex, non-linear, relationships that may exist in the FDG-PET scans, for distinguishing between healthy and diseased brains. [13, 14]

Besides PCA, the use of predefined regions of interest (ROI) is an adequate strategy. In this approach, an anatomical atlas is used to divide the brain into segments, based on anatomy, function or probability [15, 16]. Each segment is summarized, by taking the average uptake of the segment, giving a low dimensional space that is easier to interpret. This allows for analysis of a specific area of the brain more likely to be affected by disease. That being said, this ROI-based approach depends heavily on the level of detail of the atlas and may overlook subtleties within specific regions, hindering early diagnoses [17]. On the contrary, an atlas may contain too many ROIs, such that the dimensionality is still too high for a robust model.

1.3 Research Objectives

For this research, we propose a comparative analysis of two feature extraction methods for neurodegenerative disorders, using FDG-PET brain scans that reveal patterns of glucose hypo- & hypermetabolism. However, direct analysis of these scans is way too computationally intensive. That is because each scan is composed of approximately 10^5 voxels, called the "high dimensionality".

Therefore, our main research question states: "Which feature extraction method is best for brain FDG-PET in the context of neurodegenerative diseases?" To be able to answer this, we will compare two feature extraction methods, Principal Component Analysis (PCA) and Region-of-Interest (ROI) aggregation, to see which one gives better results for analyzing FDG-PET data. Here, "best" will be judged in terms of information preservation and classification performance, i.e. how well the reduced data set retains the original information and how effectively the extracted features can help distinguish between several diseases and healthy controls.

We will evaluate PCA and ROI-based feature extraction, testing multiple atlases of varying region counts, on the brain scan data. This will show the trade-offs between dimensionality reduction and information loss, by their impact on the accuracy of the diagnosis. The comparison of these methods allows us to see which method is most effective and under what conditions and what type of atlas, for diagnosing neurodegenerative diseases.

2 Theoretical Background

2.1 Previous Work

Recent research by van Veen et al. (2022) serves as the foundational on which this research builds [18]. Their study showed the effectiveness of combining principal component analysis (PCA) and Generalized Matrix Learning Vector Quantization (GMLVQ) for classification of neurodegenerative diseases

They showed that GMLVQ could effectively learn disease-specific patterns in FDG-PET data and that the relevance matrix learned by GMLVQ could be projected back to the original voxel space. This enabled them to identify brain regions most discriminative for disease classification. By using 10-fold cross-validation, they managed to achieve classification accuracies ranging from 74% to 91%.

Apart from the van Veen study, early research by Minoshima et al. (1997) identified characteristic patterns of glucose hypometabolism in AD patients, affecting posterior cingulate and parietotemporal regions [19]. Other studies have shown other disease-specific metabolic patterns [7, 20, 21].

2.2 Dimensionality Reduction

Dimensionality reduction methods allow for efficient analysis of FDG-PET brain scans by transforming them from very high-dimensional space into much lower-dimensional representations and at the same time preserving the most important information of the original data.

2.2.1 Principal Component Analysis

Principal Component Analysis tries to maximize the variance of the retained data using eigenvalue decomposition (though implemented via singular value decomposition in most computational frameworks). To fit the high-dimensional brain scan data, we rotate our coordinate system to align with the directions of maximum variance. This way, the most important dimensions are preserved and the least important ones are discarded [9].

2.2.1.1 Mathematical Foundation

For a matrix $\mathbf{X} \in \mathbb{R}^{n \times p}$ with n samples and p voxels, where each row represents a brain scan and each column a voxel intensity, the decomposition is done using singular value decomposition (SVD):

$$\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T \tag{1}$$

where $\mathbf{U} \in \mathbb{R}^{n \times n}$ and $\mathbf{V} \in \mathbb{R}^{p \times p}$ are orthogonal matrices and $\mathbf{\Sigma} \in \mathbb{R}^{n \times p}$ is a diagonal matrix containing the singular values $\sigma_1 \geq \sigma_2 \geq \ldots \geq 0$.

The columns of $\mathbf{V}$ correspond to the principal directions and the squared singular values correspond to the eigenvalues of the covariance matrix $\mathbf{C}$, the variances:

$$\lambda_i = \frac{\sigma_i^2}{n-1} \tag{2}$$

The sample covariance matrix shows the relationships between all pairs of voxels:

$$\mathbf{C} = \frac{1}{n-1}\mathbf{X}^T\mathbf{X} \in \mathbb{R}^{p\times p} \tag{3}$$

The principal components are the solutions to:

$$\mathbf{C}\mathbf{v}_i = \lambda_i \mathbf{v}_i \tag{4}$$

where $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p \geq 0$ are the eigenvalues, representing the explained variances and $\mathbf{v}_i$ are the corresponding eigenvectors, showing the principal directions. So, the eigenvalue-eigenvector relation of the covariance matrix $\mathbf{C}$ is determined from the singular value decomposition.

2.2.1.2 Dimensionality Reduction and Reconstruction Quality

For dimensionality reduction to $k < p$ components, the transformation projects data onto the subspace spanned by the first k principal components:

$$\mathbf{Y} = \mathbf{X}\mathbf{V}_k \in \mathbb{R}^{n\times k} \tag{5}$$

where $\mathbf{V}_k = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k]$ contains the most significant principal components.

The reconstruction calculation reverses the projection:

$$\hat{\mathbf{X}} = \mathbf{Y}\mathbf{V}_k^T = \mathbf{X}\mathbf{V}_k\mathbf{V}_k^T \tag{6}$$

Then the reconstruction error, given by the matrix norm, equals the sum of discarded eigenvalues:

$$\|\mathbf{X} - \hat{\mathbf{X}}\|^2 = \sum_{i=k+1}^{p} \lambda_i \tag{7}$$

Note that the reconstruction error can be significantly larger for out of sample data points when $d >> n$, or when the dimensionality is much larger than the number of available data points.

2.2.1.3 Explained Variance and Component Selection

The total variance explained by the first k components is:

$$R_k = \frac{\sum_{i=1}^{k}\lambda_i}{\sum_{i=1}^{p}\lambda_i} = \frac{\sum_{i=1}^{k}\sigma_i^2}{\sum_{i=1}^{p}\sigma_i^2} \tag{8}$$

This holds under the singular value decomposition without computing the covariance matrix, which is beneficial when $p \gg n$.

2.2.1.4 Limitations and Assumptions

PCA has a few limitations however. It cannot capture non-linear relationships that may characterize disease specific metabolic patterns. Additionally, it focuses on variance maximization and may not capture features represented by low-variance signals, depending on how many components are included.

When the number of features exceeds the number of samples ($p > n$), PCA can only identify at most $n-1$ components. This may suggest that PCA primarily captures patterns that are present in the training data, potentially missing important variations in new, never before seen, data.

2.2.2 Region-of-Interest Aggregation

Region-of-Interest (ROI) methods take a different approach by incorporating the anatomical and functional structure of the brain to subdivide it in specific regions. ROI feature extraction leverages the fact that neurodegenerative diseases do not affect the brain uniformly. ROI methods try to capture these disease-specific patterns, by focusing on anatomically meaningful regions [10, 15].

2.2.2.1 Mathematical Foundation

The ROI approach transforms the three-dimensional brain volume into a feature vector, starting with the brain parcellation, where atlases divide the entire brain volume into discrete regions based on anatomical, functional or probabilistic patterns.

Let's define the brain volume $\mathbf{V} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ and a corresponding atlas $\mathbf{A} \in \mathbb{Z}^{d_1 \times d_2 \times d_3}$ containing R distinct regions. Each voxel in the atlas volume carries an integer label indicating its regional assignment:

$$\mathcal{R}_r = \{(i, j, k) : \mathbf{A}(i, j, k) = r\}, \quad r \in \{1, 2, \ldots, R\} \tag{9}$$

This defines the set of voxel coordinates belonging to region r.

The next step is transforming the metabolic information within each region into scalar features. The choice of aggregation function significantly impacts the resulting feature quality. The regional feature extraction follows:

$$\mathbf{f}_r = g(\{\mathbf{V}(i, j, k) : (i, j, k) \in \mathcal{R}_r\}) \tag{10}$$

where $g(\cdot)$ represents the aggregation function that transforms the collection of voxel intensities within region r into a single representative value. The aggregation function can be the mean, median or the mode, but mean is most intuitive to use.

Mean aggregation is done by computing the average metabolic activity across all voxels within each region:

$$f_r^{(\text{mean})} = \frac{1}{|\mathcal{R}_r|} \sum_{(i,j,k) \in \mathcal{R}_r} \mathbf{V}(i, j, k) \tag{11}$$

This approach assumes that the average metabolic activity is representative of each region.

The standard deviation signifies the heterogeneity within regions and is given by:

$$f_r^{(\text{std})} = \sqrt{\frac{1}{|\mathcal{R}_r| - 1} \sum_{(i,j,k) \in \mathcal{R}_r} (\mathbf{V}(i,j,k) - f_r^{(\text{mean})})^2} \tag{12}$$

2.2.2.2 Atlas Types and Parcellation

The choice of brain atlas shapes the ROI analysis, as different atlases embody distinct perspectives on brain organization. Each atlas type reflects different theoretical frameworks for understanding brain structure and function, leading to complementary insights into neurodegenerative processes.

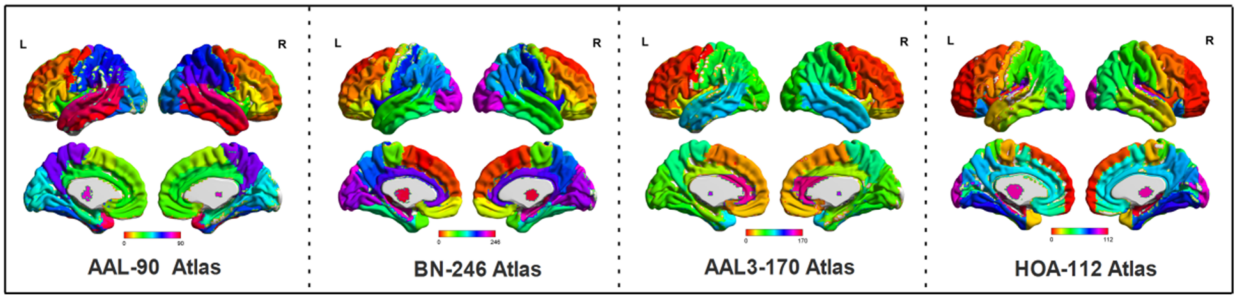

Figure 1: Visualization of brain atlases used in this study: AAL-90 Atlas (90 regions), BN-246 Atlas (246 regions), AAL3-170 Atlas (170 regions), and HOA-112 Atlas (112 regions), shown in both lateral view (top row) and sagittal (bottom row) view. Figure reprinted from [22].

Anatomical atlases define regions based on the macroscopic structure visible in anatomical MRI scans. These atlases partition the brain according to sulcal and gyral patterns that remain relatively consistent across individuals. The Automated Anatomical Labeling (AAL) atlas, developed by Tzourio-Mazoyer et al. (2002), has 116 regions, both cortical and subcortical, based on sulcal and gyral patterns [10]. The Destrieux atlas has 148 cortical anatomical regions [23], with finer granularity [16, 24].

Functional atlases derive brain parcellations from patterns of neural activity and connectivity. The Schaefer atlas, derived from resting-state functional MRI data, identifies regions based on similar temporal patterns of activity [15].

Probabilistic atlases derive brain parcellations by determining the likelihood that each voxel belongs to a specific region. The Harvard-Oxford atlases are probabilistic maps with 48 cortical and 21 subcortical regions derived from parcellation of healthy references [25].

For probabilistic atlases, a threshold of 25% probability for voxel assignment was used, following FSL software defaults [26]. Using this threshold makes sure that only voxels with

sufficient confidence of being in a specific region are used. Additionally, making sure there is still enough brain coverage for the analysis. Voxels with a probability below 25% for all regions are discarded. This is done to make sure we don't introduce noise into the atlas and therefore the analysis.

2.2.2.3 Dimensionality Reduction and Information retention

ROI methods can achieve a compression ratio of approximately $1000 : 1$ with dimensionality reduction, transforming brain volumes containing $p \approx 10^5$ voxels into feature vectors with $R \approx 10^2$ values, making it computationally efficient.

However, this will inevitably lead to information loss. The aggregation discards all information within the regions of the atlas. The information loss can be expressed as follows:

$$\mathcal{L}_{\text{ROI}} = \sum_{r=1}^{R} \sum_{(i,j,k)\in\mathcal{R}_r} (\mathbf{V}(i,j,k) - f_r)^2 \tag{13}$$

representing the total squared difference between original voxel values and their regional representatives.

2.2.2.4 Atlas Granularity

The granularity of an atlas affects its ability to retain information. Coarse atlases with fewer, larger regions have a higher change of increasing bias, as it mixes different tissue types by averaging unhealthy and healthy regions. On the other hand, fine-grained atlases with many small regions have a higher chance of high variance due to insufficient number of patients to compensate for individual variance in the data. This is more precise anatomically, but can lead to noisy and unreliable analysis.

2.2.2.5 Advantages and Limitations

The ROI approach offers some advantages over PCA. The extracted features maintain anatomical information, making it easier to see which brain regions contribute to what disease.

However, the predetermined parcellations may miss important information depending on what type of atlas is used, the coverage of said atlas and the granularity of that atlas. So, different atlases perform differently for different application, making it harder to find the optimal atlas that works for all applications. Additionally, the aggregation removes patterns within regions that can contain important information about the tissue.

2.3 Classification

We now have established the feature extraction methods that transform high-dimensional brain scans into manageable representations, but we still need classification algorithms to identify the subtle patterns within these features that distinguish between healthy and diseased states and even between different neurodegenerative diseases.

2.3.1 Generalized Matrix Learning Vector Quantization

Generalized Matrix Learning Vector Quantization (GMLVQ) learns both representative prototypes for each diagnostic category and an adaptive distance metric that emphasizes the most discriminative feature combinations [27].

2.3.1.1 Prototype-Based Learning

The idea of GMLVQ is that classification problems can be solved by identifying representative examples, or prototypes, for each class and classifying new observations based on how similar they are to these prototypes.

GMLVQ maintains a set of prototype vectors $\{\mathbf{w}_j\}_{j=1}^{M}$, where each prototype $\mathbf{w}_j \in \mathbb{R}^p$ represents a characteristic pattern for a specific diagnostic class $c(\mathbf{w}_j)$. These prototypes serve as bases in the feature space, capturing the characteristics that define each category. The learning process also adapts the distance metric to emphasize features that give maximal discriminative power.

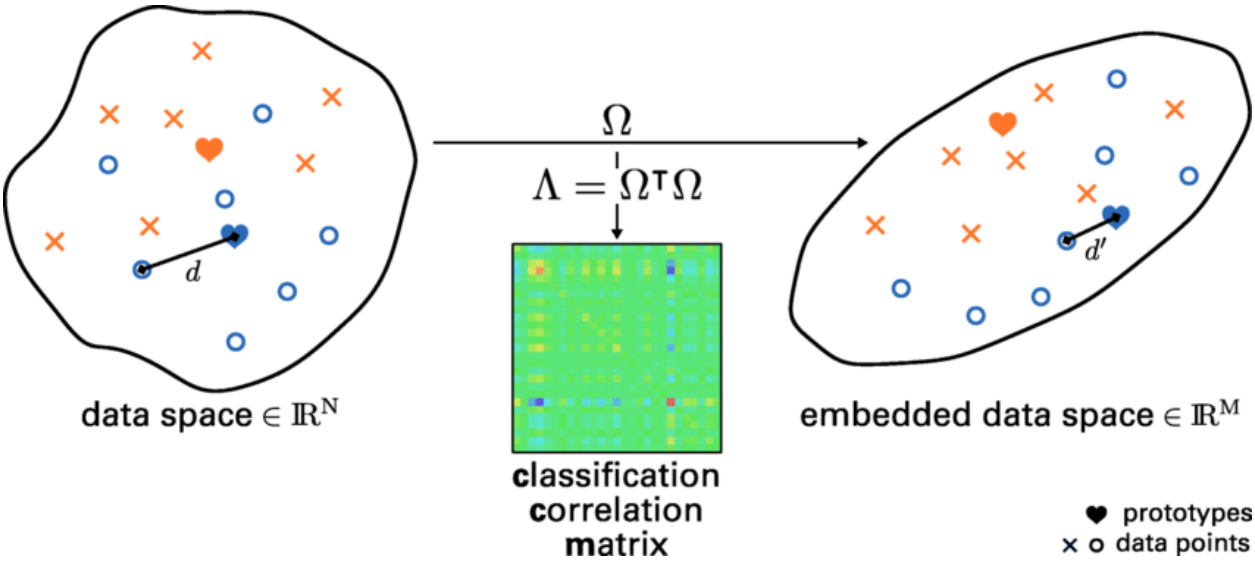

Figure 2: Illustration of dimensionality reduction from high-dimensional data space ($\mathbb{R}^N$) to lower-dimensional embedded space ($\mathbb{R}^M$). The transformation is performed through matrix Ω, where $\Lambda = \Omega^T\Omega$ is the classification correlation matrix. Data points (circles and crosses) and prototypes (hearts) are preserved in the embedding, as indicated by the distance measures d and d'. Figure reprinted from [28].

2.3.1.2 Distance Metric Learning

GMLVQ learns a relevance matrix $\mathbf{\Lambda} \in \mathbb{R}^{p \times p}$ that weights and matches features based on their discriminative importance [29].

The adaptive distance between input $\mathbf{x}$ and prototype $\mathbf{w}_j$ is computed as:

$$d_{\mathbf{\Omega}}(\mathbf{x}, \mathbf{w}_j) = (\mathbf{x} - \mathbf{w}_j)^T \mathbf{\Lambda} (\mathbf{x} - \mathbf{w}_j) \tag{14}$$

where $\mathbf{\Lambda} = \mathbf{\Omega}^T \mathbf{\Omega}$ represents the learned relevance matrix. The diagonal elements of $\mathbf{\Lambda}$ determine the importance of individual features and the off-diagonal elements capture relationships between features. The irrelevant features receive near-zero weights by the gradient-based optimization.

The relevance matrix effectively transforms the feature vector space: stretching dimensions along discriminative directions and compressing irrelevant variations.

2.3.1.3 Classification Decision Making

For any input sample $\mathbf{x}$, the algorithm identifies the closest prototypes from both the correct class and the incorrect classes:

$$d^+ = \min_{j:c(\mathbf{w}_j)=c(\mathbf{x})} d_{\mathbf{\Omega}}(\mathbf{x}, \mathbf{w}_j) \quad \text{(correct class distance)} \tag{15}$$

$$d^- = \min_{j:c(\mathbf{w}_j)\neq c(\mathbf{x})} d_{\mathbf{\Omega}}(\mathbf{x}, \mathbf{w}_j) \quad \text{(incorrect class distance)} \tag{16}$$

The relative distance can be used to determine the confidence of the algorithm between correct and incorrect class assignments:

$$\mu(\mathbf{x}) = \frac{d^+ - d^-}{d^+ + d^-} \tag{17}$$

where $d^+, d^- \geq 0$. The magnitude of μ reflects the decision confidence, with values near zero, with $d^+ = d^-$, indicating low confidence and values near ± 1 indicating high confidence, either $d^+ = 0$ and the sample lies exactly on the correct prototype ($\mu = -1$) or $d^- = 0$ and the sample lies exactly on the wrong prototype ($\mu = +1$).

2.3.1.4 Activation Type

The learning process optimizes the prototypes and relevance matrix with derivative-based minimization of a differentiable cost function, called the activation type. The activation type used in this study is the sigmoid activation function:

$$E = \sum_{i=1}^{n} \Phi(\mu(\mathbf{x}_i)) = \sum_{i=1}^{n} \frac{1}{1 + \exp(-\mu(\mathbf{x}_i))} \tag{18}$$

2.3.2 Cross-Validation

Neurodegenerative disease datasets most often contain limited samples with imbalanced class distributions. For example, Alzheimer's disease samples often outnumber rarer conditions like Dementia with Lewy Bodies.

2.3.2.1 Stratified K-Fold Cross-Validation

For dataset $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{n}$ with class distribution $\mathbf{p} = [p_1, p_2, \ldots, p_K]$ where $p_j = |\{i : y_i = j\}|/n$, each fold $\mathcal{F}_k$ maintains approximate class proportions:

$$\left| \frac{|\{i \in \mathcal{F}_k : y_i = j\}|}{|\mathcal{F}_k|} - p_j \right| \leq \epsilon \tag{19}$$

for small tolerance $\epsilon > 0$. This constraint ensures that each validation fold contains representative samples from categories, preventing favoring larger classes.

The 10-fold cross-validation performance estimate averages across all folds:

$$\hat{R}_{\text{CV}} = \frac{1}{10} \sum_{k=1}^{10} R_k(\mathcal{F}_k) \tag{20}$$

where $R_k(\mathcal{F}_k)$ is performance on fold k, using a model trained on $\mathcal{D} \setminus \mathcal{F}_k$. Ten-fold cross-validation was chosen to maximize training data per fold, minimize bias and maintain acceptable variance, given our limited dataset size and class imbalance [30].

3 Materials and Methods

3.1 Dataset Characteristics and Processing

3.1.1 Dataset Composition

The analysis makes use of the same FDG-PET neuroimaging dataset used by van Veen et al. (2022), comprising 236 brain scans from patients with neurodegenerative conditions and healthy controls. The dataset includes: Parkinson's Disease (PD, $n = 41$), Alzheimer's Disease (AD, $n = 56$), Dementia with Lewy Bodies (DLB, $n = 23$), idiopathic REM sleep behavior disorder (RBD, $n = \text{RBD1} + \text{RBD2} = 47$), and Healthy Controls (HC, $n = 69$) [18].

The images are stored in NIfTI format (`.nii`) with corresponding participant metadata in tab-separated value (`.tsv`) files. Each image has dimensions of $91 \times 109 \times 91$ voxels, with each voxel having dimensions of $2 \times 2 \times 2$ millimeters.

3.1.2 Diagnostic Category

To address the imbalance within the dataset between diseases, the decision was made to group related disease subtypes: AD variants (AD, AD-P, AD-ATYP, AD-MCI) combined into AD, PD kept as a distinct category, RBD variants (RBD1, RBD2) combined into RBD as a separate category, DLB kept as a distinct category and HC kept as healthy controls.

RBD remains as a separate category due to its uncertain progression. iRBD can develop into PD, DLB or multiple system atrophy (MSA), with patterns varying heavily across patients [31].

3.1.2.1 Class Distribution

The final class distribution vector follows $\mathbf{n} = [n_{\text{PD}}, n_{\text{AD}}, n_{\text{DLB}}, n_{\text{RBD}}, n_{\text{HC}}]$ with five distinct diagnostic categories.

3.1.3 Data Splitting Strategy

To make sure the class imbalance doesn't cause any problems during the analysis, stratified random sampling is used to create training and reference datasets, with a split ratio of 90/10, such that there was proportional class representation across splits. Each split configuration is verified to have minimum class representation of at least 5 samples.

3.2 Image Preprocessing and Optimization

3.2.1 Intensity Normalization

To make the brain scans comparable between scans, voxel intensity normalization is applied to all scans beforehand, by dividing each voxel value by the mean intensity of all in-brain voxels.

For brain volume $\mathbf{V} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ and brain mask $\mathbf{M} \in \{0,1\}^{d_1 \times d_2 \times d_3}$, the normalization follows:

$$\tilde{\mathbf{V}}(i,j,k) = \frac{\mathbf{V}(i,j,k)}{\bar{\mathbf{V}}_{\text{brain}}} \tag{21}$$

where $\bar{\mathbf{V}}_{\text{brain}}$ is the mean in-brain intensity and $\mathbf{M}(i,j,k) = \mathbf{1}_{brain}(i,j,k)$ with $\mathbf{1}_{brain}(i,j,k)$, the indicator function that equals 1 if the voxel at coordinates (i,j,k) is an in-brain voxel.

3.3 Feature Extraction Implementation

3.3.0.1 PCA Feature Extraction

PCA models are fitted exclusively on reference data (10% of total samples) and then the training data is projected onto the principal components for classification analysis. For reference dataset $\mathcal{D}_{\text{ref}}$ and training dataset $\mathcal{D}_{\text{train}}$, the transformation follows:

$$\mathbf{T}_{\text{PCA}}(\mathbf{x}) = (\mathbf{x} - \boldsymbol{\mu}_{\text{ref}})\mathbf{V}_{\text{ref},k} \tag{22}$$

where $\mathbf{T}_{\text{PCA}}(\mathbf{x})$ is the PCA-transformed version of input sample $\mathbf{x}$, $\boldsymbol{\mu}_{\text{ref}}$ is the mean vector computed from the reference dataset and $\mathbf{V}_{\text{ref},k}$ is the principal components matrix with the k most important eigenvectors learned from reference data.

3.3.1 ROI Feature Extraction

For the ROI aggregation method, 6 brain parcellation atlases are used: AAL (116 regions), Destrieux (148 regions, 47.8% brain coverage), Harvard-Oxford Cortical (48 regions, 65.1% brain coverage), Harvard-Oxford Subcortical (21 regions, 83.9% brain coverage), Schaefer 2018 (100 regions, 65.3% brain coverage; 200 regions), representing two atlases for each type: anatomical, probabilistic and functional, respectively. These atlases have varying brain coverages and cover varying regions in the brain. This will help us determine which regions are most important for which disease types.

Then, for each atlas, the mean FDG uptake values within each region is calculated. For region r with voxel set $\mathcal{R}_r$, the feature extraction follows:

$$f_r = \frac{1}{|\mathcal{R}_r|} \sum_{(i,j,k) \in \mathcal{R}_r} \mathbf{V}(i,j,k) \tag{23}$$

In the aggregation process, we do voxel-wise multiplication with brain mask M defined previously to only include in-brain voxels and then the mean uptake calculation across the valid voxels within each region.

3.4 Experimental Design

For the PCA methods, transformation parameters are learned from reference data and applied to training data for cross-validation. The ROI-based methods use standard stratified cross-validation only on the training data, using the exact same data splits as the PCA experiments, since the atlas-based feature extraction requires no training phase.

10-fold stratified cross-validation was used with scikit-learn's `StratifiedKFold` with `shuffle=True` and `random_state=42`. We determine the weighted F1-score to account for class imbalance. Additionally, the overall accuracy, class-specific precision and recall, and confusion matrices. The weighted F1-score follows:

$$\mathrm{F1}_{\text{weighted}} = \sum_{i=1}^{K} \frac{n_i}{n} \cdot \mathrm{F1}_i \tag{24}$$

where n_i is the true count for class i and $\mathrm{F1}_i$ is the class-specific F1-score.

3.5 Classification and Reconstruction Analysis

3.5.1 Classifier Implementation

Scikit-learn's GMLVQ is used from the sklvq library [32] with the following hyperparameters: The prototype learning rate of 0.1 avoids overshooting and adapts quickly to class-specific patterns, which is important for our limited sample sizes (23-69 per class). The relevance matrix rate of 0.01 prevents oscillation in the adaptive distance metric, important for convergence in high-dimensional data [27].

3.5.2 Reconstruction Analysis

The reconstruction quality is determined by mean squared error (MSE) between original and reconstructed voxel values. The PCA reconstruction was performed using the standard PCA inverse transformation:

$$\mathbf{X}_{\text{reconstructed}} = \mathbf{X}_{\text{projected}} \mathbf{V}_k^T \tag{25}$$

The ROI reconstruction was performed by assigning regional mean values to all voxels within each region against original normalized data.

3.5.3 Region Importance Analysis

Following GMLVQ training, the learned relevance matrix $\mathbf{\Lambda} = \mathbf{\Omega}^T\mathbf{\Omega}$ contains feature importance weights. For ROI-based features, the diagonal elements Λ_{rr} show the discriminative importance of region r. These values are the learned weights that the algorithm applies to each feature during sample-prototype distance calculations. The higher these diagonal elements are, the more significantly their corresponding regions contribute to class separation.

Each diagonal element is then normalized by dividing it by the maximum diagonal element across all regions, giving us relative importance scores between 0 and 1. A score of 1 corresponds to the most discriminative region and a score of 0 corresponds to the least discriminative region for distinguishing between disease categories.

For the disease-specific analysis, the differences between disease group means and healthy control means are determined:

$$z_{r,d} = \frac{\bar{f}_{r,d} - \bar{f}_{r,HC}}{\sigma_{r,HC}} \tag{26}$$

where $\bar{f}_{r,d}$ is mean FDG uptake in region r for disease d, $\bar{f}_{r,HC}$ is the healthy control mean and $\sigma_{r,HC}$ is the healthy control standard deviation.

3.5.4 Experimental Pipeline

The experimental pipeline follows these steps:

For PCA:

1. Apply stratified split to create reference data (10%) and training data (90%)
2. Learn PCA transformation parameters ($\boldsymbol{\mu}_{\text{ref}}, \mathbf{V}_{\text{ref},k}$) only from reference data
3. Apply learned transformation to training data: $\mathbf{Y}_{\text{train}} = (\mathbf{X}_{\text{train}} - \boldsymbol{\mu}_{\text{ref}})\mathbf{V}_{\text{ref},k}$
4. Perform 10-fold stratified cross-validation on transformed training data using GMLVQ
5. Evaluate reconstruction quality using reference data transformation

For ROI:

1. Apply atlas-based feature extraction to training data (no reference data needed)
2. Perform 10-fold stratified cross-validation on training data using GMLVQ
3. Evaluate reconstruction quality by averaging all voxels within each region and comparing with the original data.

The maximum number of meaningful principal components is limited by the reference dataset size. For reference datasets, PCA can extract at most $n_{\text{ref}} - 1$ meaningful components.

4 Results and Discussion

4.1 Reconstruction Performance Analysis

4.1.1 PCA Reconstruction Results

Overall, principal component analysis showed substantially superior reconstruction performance across all metrics compared to all ROI methods. As expected, PCA-23, with maximum available components, achieved a reconstruction MSE score of 0, PCA-20 produced an MSE of $4.15 \times 10^{-4} \pm 5.73 \times 10^{-4}$ with normalized MSE of 7.54×10^{-3}. PCA-15 achieved an MSE of $1.33 \times 10^{-3} \pm 9.00 \times 10^{-4}$ with normalized MSE of 2.42×10^{-2}, and PCA-10 produced an MSE of $2.59 \times 10^{-3} \pm 1.29 \times 10^{-3}$ with normalized MSE of 4.70×10^{-2}.

The variance explained shows a clear negatively correlated trend in Figure 3: PCA-23 achieved the best information retention at 100.0% variance explained, followed by PCA-20 at 96.2%, PCA-15 at 87.8%, and PCA-10 at 76.2%. This shows that PCA has very impressive information retention performance, even with very little components extracted.

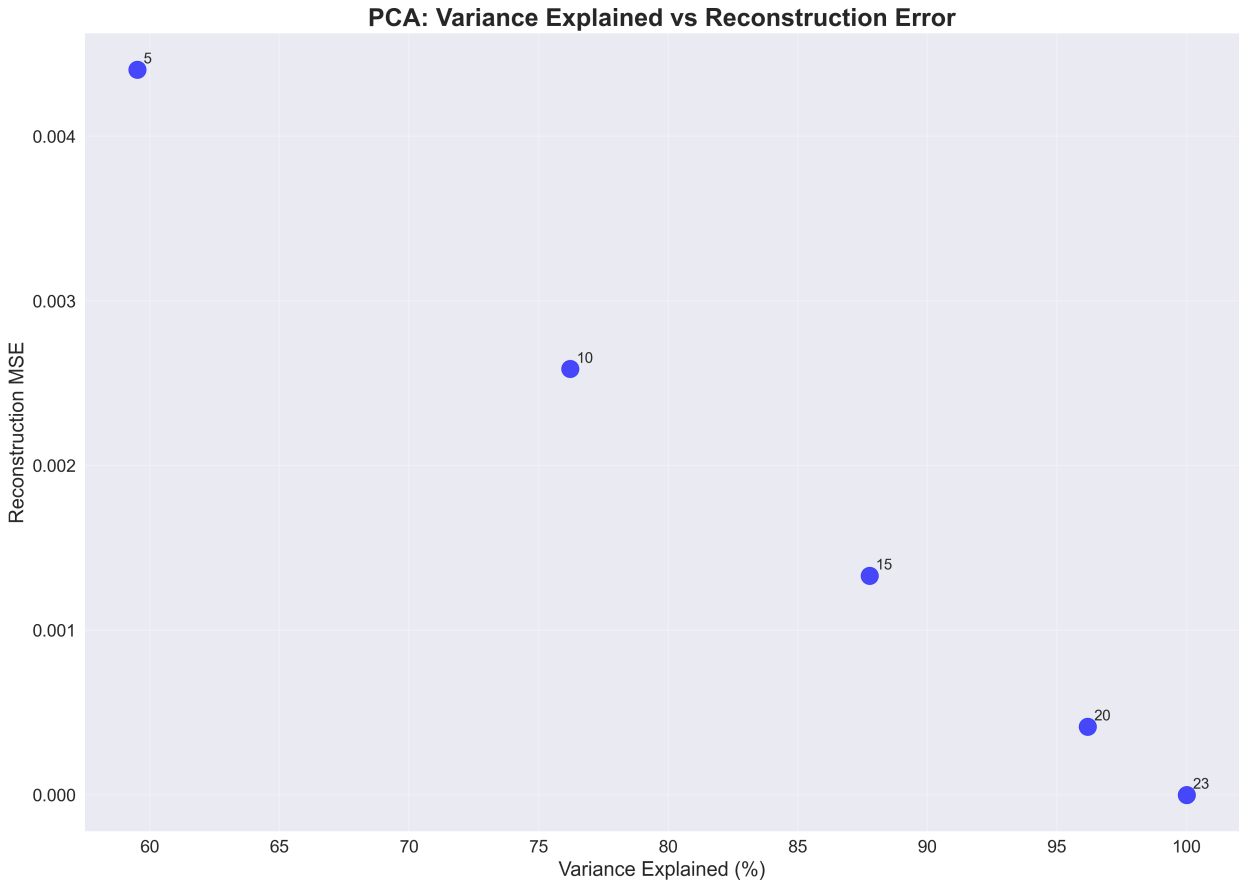

Figure 3: Variance explained by principal components

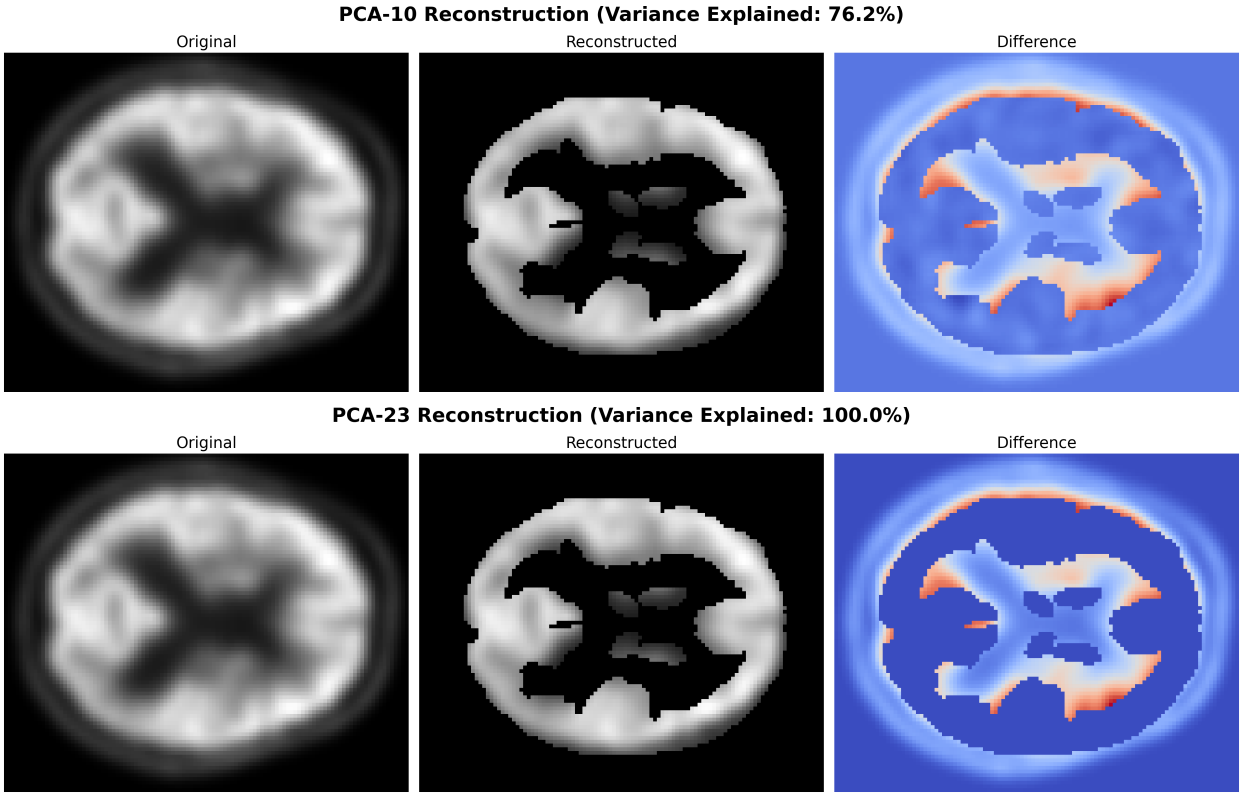

Figure 4: PCA reconstruction quality comparison: (Top) PCA-10 showing higher reconstruction error with visible loss of fine details, (Bottom) PCA-23 demonstrating perfect reconstruction.

4.1.2 ROI Reconstruction Results

The Region-of-Interest methods perform substantially worse, compared to PCA, with much higher reconstruction errors. The AAL atlas achieved the lowest ROI reconstruction MSE of $8.88 \times 10^{-2} \pm 3.35 \times 10^{-3}$, followed by Harvard-Oxford-Subcortical ($1.65 \times 10^{-1} \pm 1.45 \times 10^{-2}$) and Harvard-Oxford-Cortical ($3.12 \times 10^{-1} \pm 1.60 \times 10^{-2}$). The functional atlases showed much higher reconstruction errors, compared to anatomical and probabilistic atlases: Schaefer-200 ($3.12 \times 10^{-1} \pm 1.61 \times 10^{-2}$), Schaefer-100 ($3.13 \times 10^{-1} \pm 1.61 \times 10^{-2}$). The Destrieux ($4.94 \times 10^{-1} \pm 9.54 \times 10^{-3}$) atlas produced the highest reconstruction errors, due to its low brain coverage of 47.8%.

These reconstruction error values are approximately 2 to 3 orders of magnitude higher compared to PCA, see Figure 7. This clearly shows the advantage PCA has over ROI-aggregation. PCA, even with very little components used, retains information way better. Important to note is that the performance of the anatomical atlas AAL compared to any functional atlases suggests that an anatomical parcellation with medium fine granularity (116 regions) performs better in preserving metabolic structure than both very fine-grained anatomical or any functional atlases.

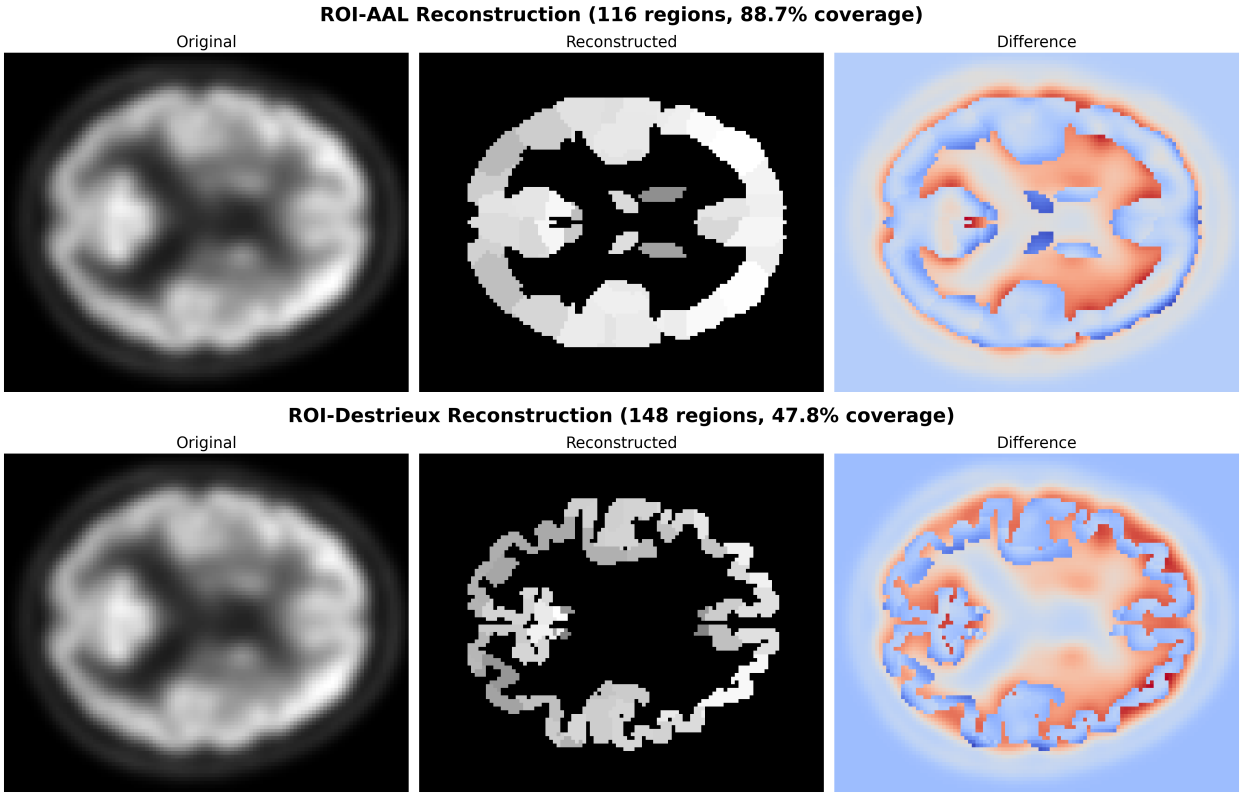

Figure 5: ROI reconstruction comparison: (Top) AAL atlas showing the best ROI reconstruction performance, (Bottom) Destrieux atlas showing the highest reconstruction error among ROI methods due to its limited brain coverage (47.8%).

In Figure 8, we see that when we only include the regions contained in an atlas, the reconstruction performance improves significantly. Which is not surprising, but shows us that with the right implementation, ROI-aggregation can be competitive with PCA.

Figure 6 shows that this approach also results in approximately equal performance across atlas types, where before there wasn't a trend between reconstruction performance and atlas type.

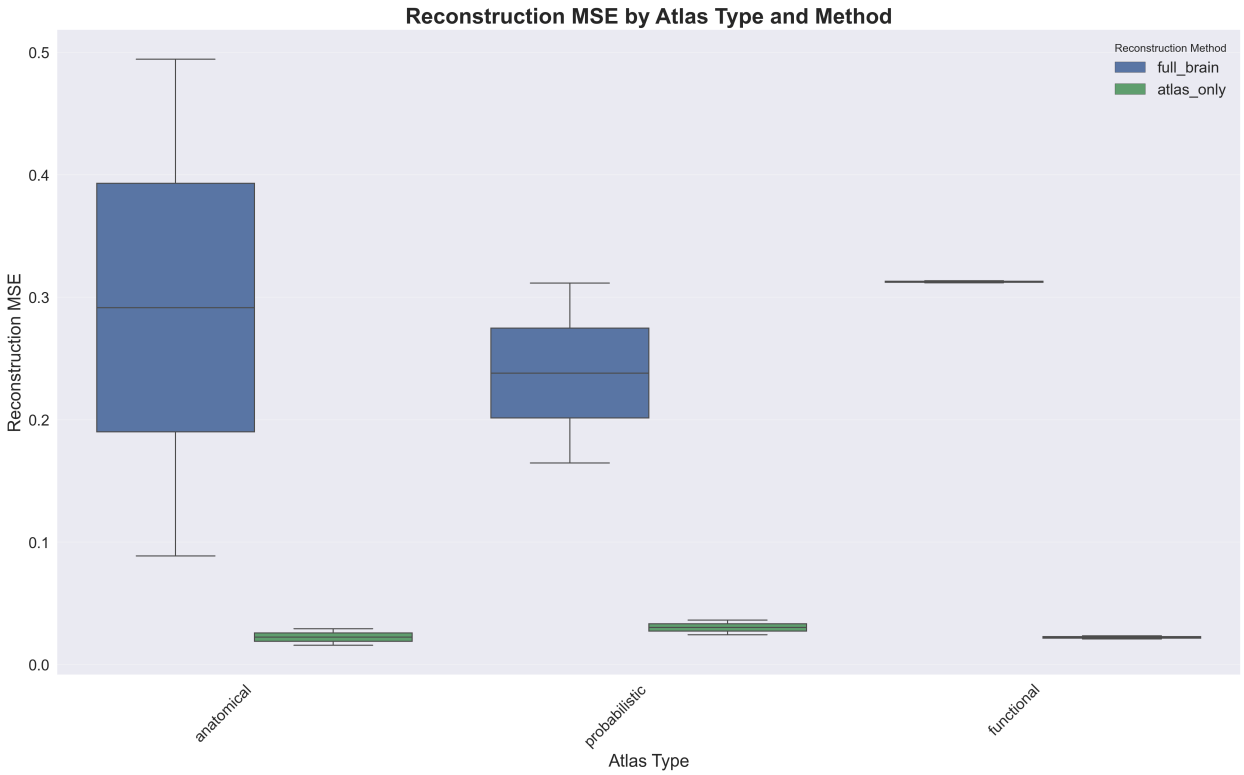

Figure 6: Reconstruction performance comparison between PCA and ROI, both for reconstruction of the full brain and reconstruction of only the regions each atlas contains. PCA methods still outperform ROI methods, but only by one order of magnitude.

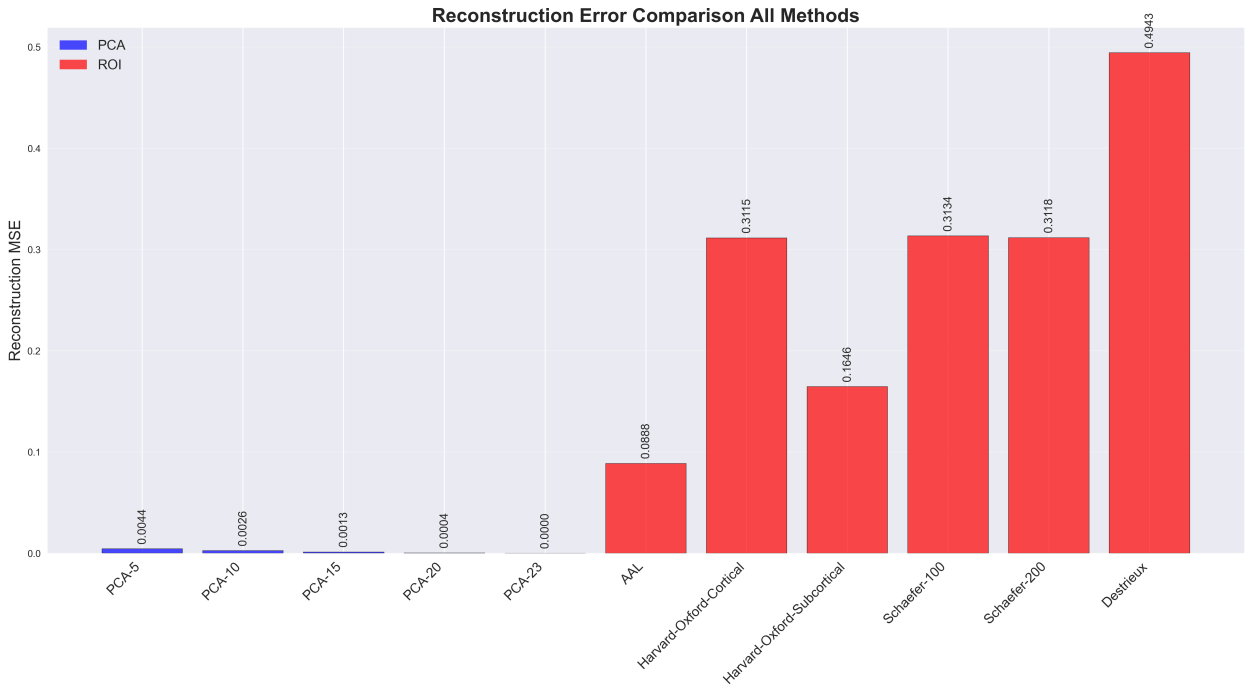

Figure 7: Reconstruction performance comparison between PCA and ROI. PCA methods achieve 2-3 orders of magnitude lower reconstruction errors than ROI methods.

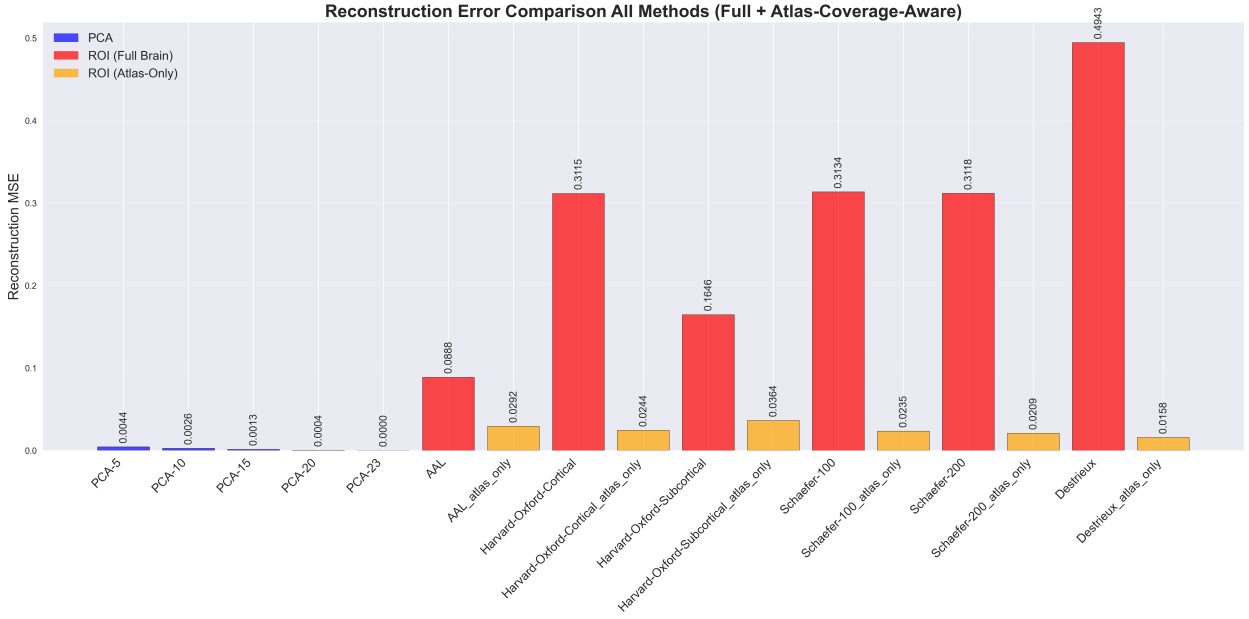

Figure 8: Reconstruction performance comparison between PCA and ROI, both for reconstruction of the full brain and reconstruction of only the regions each atlas contains. PCA methods still outperform ROI methods, but only by one order of magnitude.

4.2 Classification Results

4.2.1 PCA Classification Results

In the case of classification, again PCA scored very well. PCA-10 showed great classification performance with a cross-validation F1-score of 0.606 ± 0.098. However, PCA-20 achieved 0.704 ± 0.089, which is the best result of all methods.

The confusion matrix for PCA-10 shows that it performs decent in DLB and RBD classification, getting 66.7% true positive rate for these categories (Figure 9). Overall, PCA-10 scored well across all disease categories, with healthy controls getting 90.3% classification accuracy and Parkinson's disease reaching 51.8% accuracy.

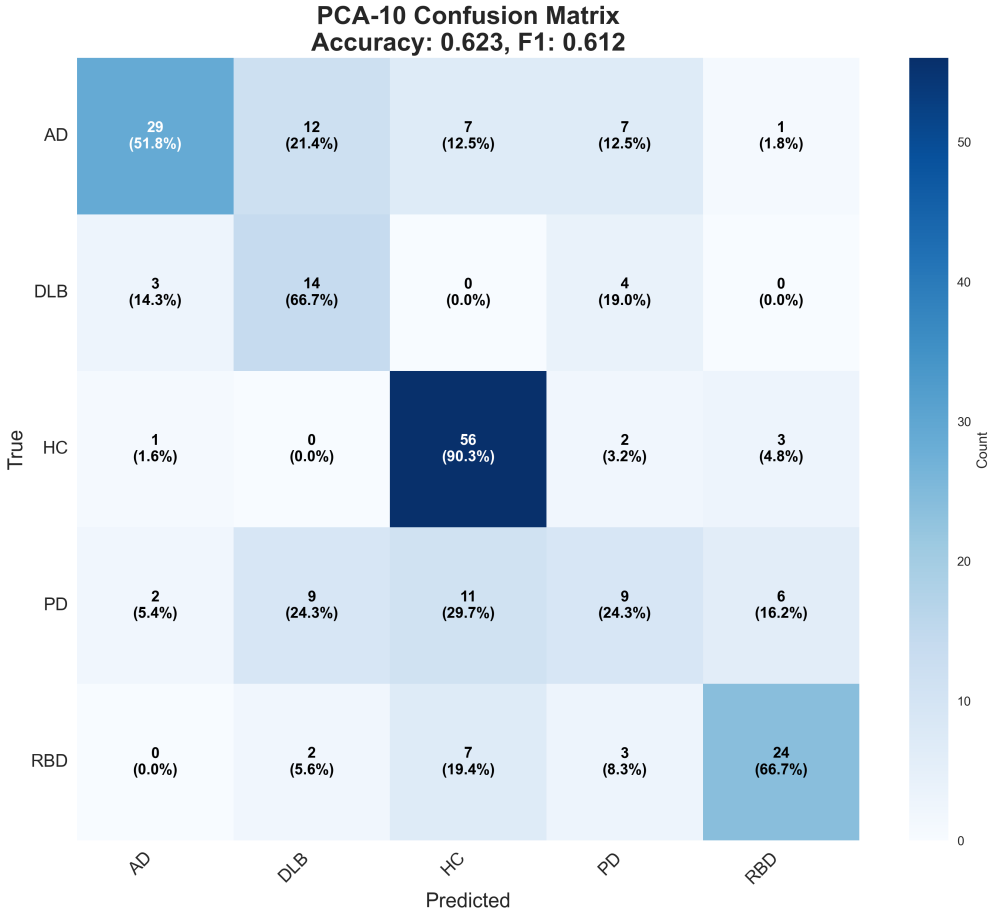

Figure 9: Confusion matrix for PCA-10 classification showing true versus predicted diagnostic categories. It shows strong performance for healthy controls (90.3% accuracy) and DLB classification (66.7% true positive rate), with average performance for other disease categories.

4.2.2 ROI Classification Results

The ROI classification shows varying performance depending on parcellation type and brain coverage percentage. Destrieux achieved an F1-score of 0.641 ± 0.098, making it the best performing ROI method and second-best overall. This suggests that anatomical connectivity-based parcellations with fine granularity are very well suited for classification of neurodegenerative diseases. Schaefer-200 showed moderate performance with 0.632 ± 0.067 and AAL showed weaker performance at 0.554 ± 0.124.

The Harvard-Oxford cortical atlas shows moderate performance (0.436 ± 0.099), while the subcortical variant showed comparable results (0.466 ± 0.048). Schaefer-100 performed worse (0.598 ± 0.151) than Schaefer-200. Which suggests that higher granularity in functional atlases can improve performance, similar to anatomical atlases where fine parcellation captures important discriminative information.

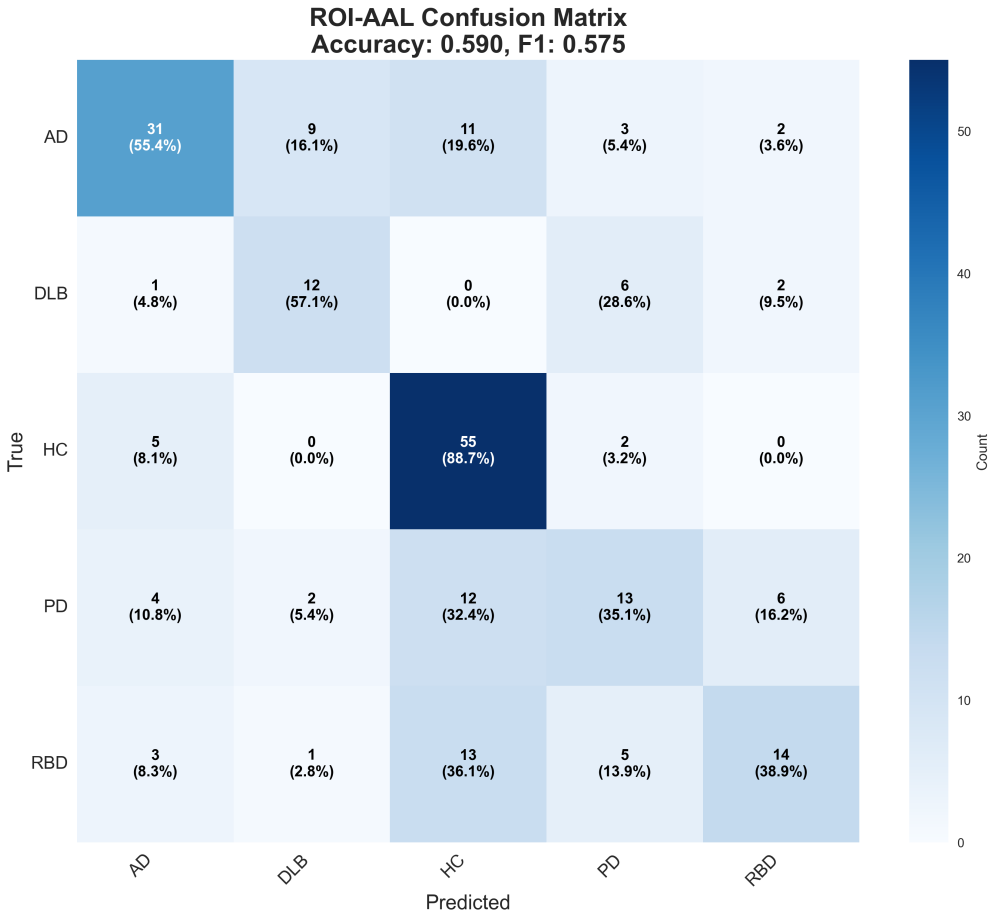

Figure 10: Confusion matrix for AAL classification showing true versus predicted diagnostic categories. Healthy controls reached 88.7% accuracy. The matrix reveals challenges in distinguishing PD from other neurodegenerative diseases, consistent with known overlapping metabolic patterns.

The confusion matrix for the AAL atlas (Figure 10) shows good performance in healthy control classification (88.7% accuracy). However, it struggled in distinguishing PD from other neurodegenerative diseases.

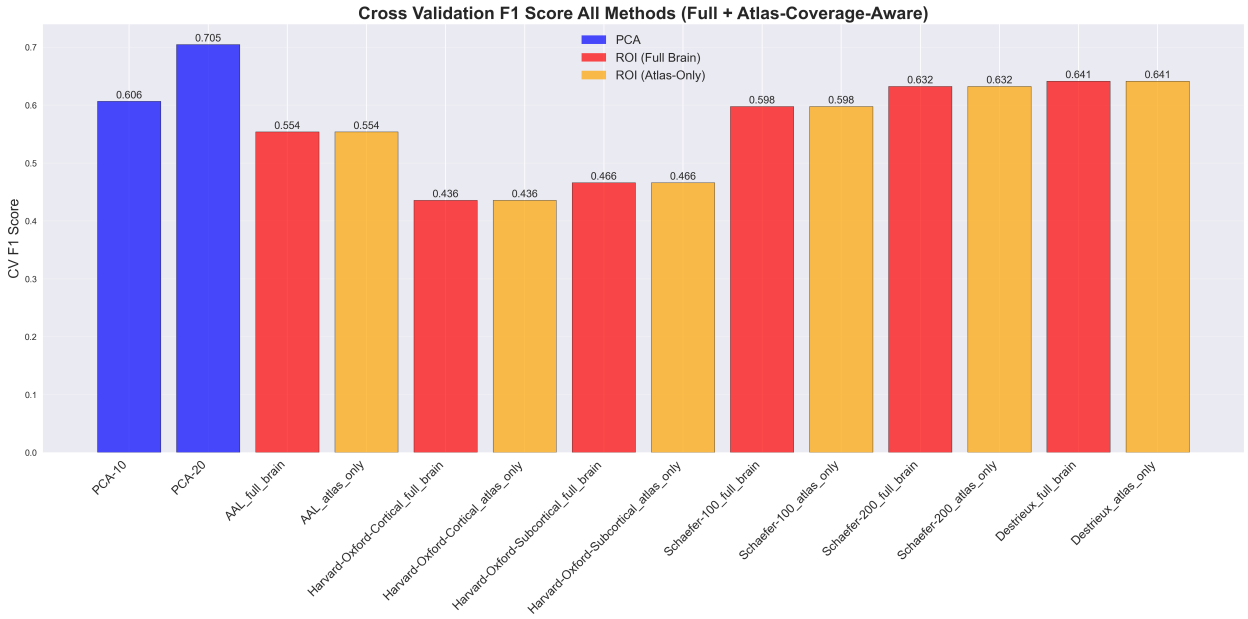

Figure 11: Cross-validation F1-score comparison across all feature extraction methods. PCA methods (blue bars) show competitive performance with PCA-10 getting 0.606 and PCA-20 reaching 0.705. ROI methods (red bars) show varied performance with Schaefer-200 getting 0.632, while Harvard-Oxford subcortical atlas showed the worst performance at 0.436. The yellow bars are redundant.

4.3 Method Comparison

4.3.1 Performance vs Dimensionality Analysis

PCA methods, with only 10 or 20 components, performed very well in both classification and reconstruction. ROI methods, with 21 to 200 regions, showed great performance at intermediate granularity, with Schaefer-200 outperforming both lower-dimensional Harvard-Oxford subcortical (21 regions) and lower-granularity Schaefer-100 atlases.

4.3.2 Information Retention vs Classification Performance

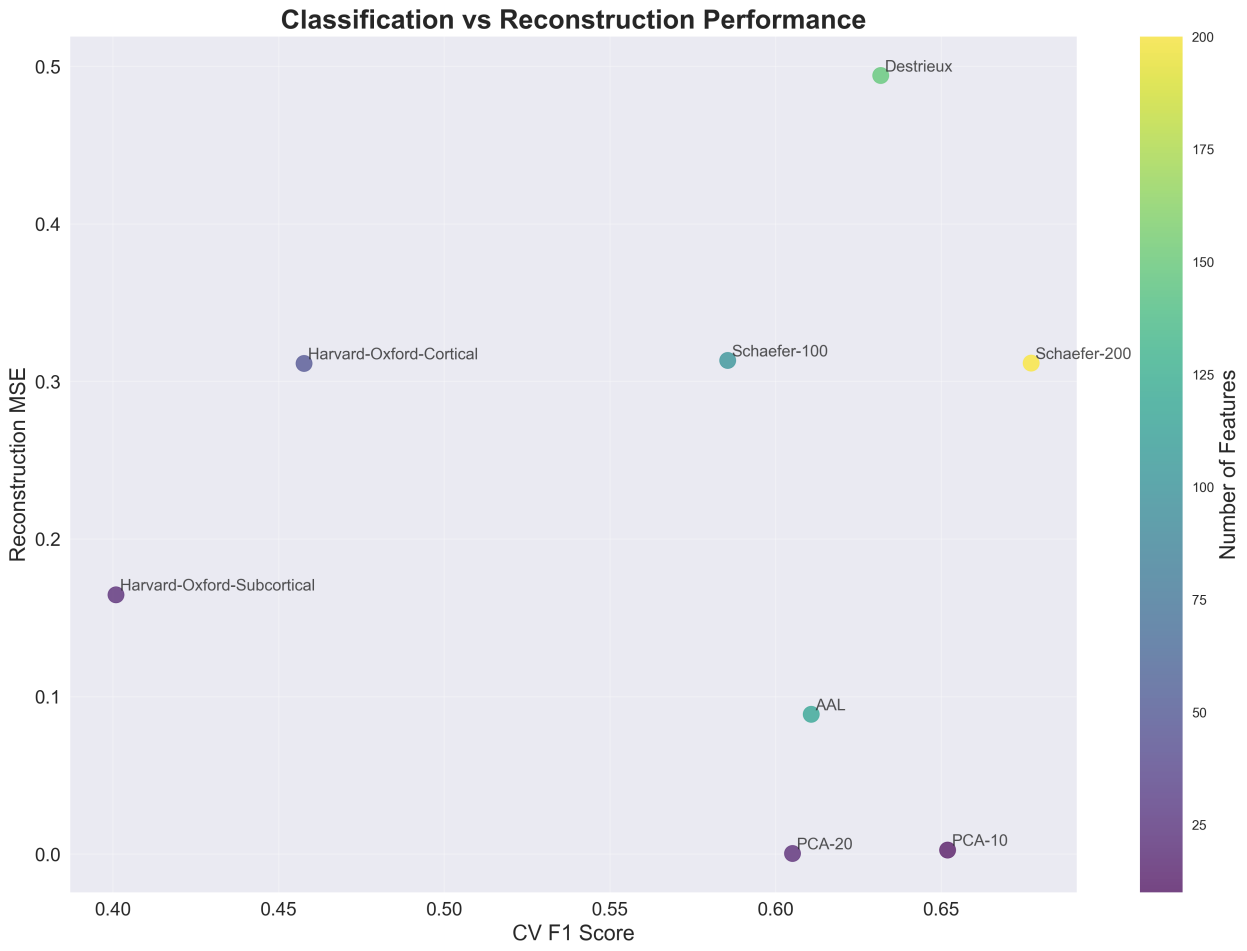

Figure 12: Classification F1-score versus reconstruction MSE. PCA methods (bottom right quadrant) perform the best overall, both in reconstruction quality (MSE < 0.003) and classification performance (CV F1-score > 0.606). ROI methods show way higher reconstruction errors, but have comparable or even better classification performance in some cases.

4.4 Disease-Specific Brain Region Analysis

4.4.1 Region Importance and Disease Patterns

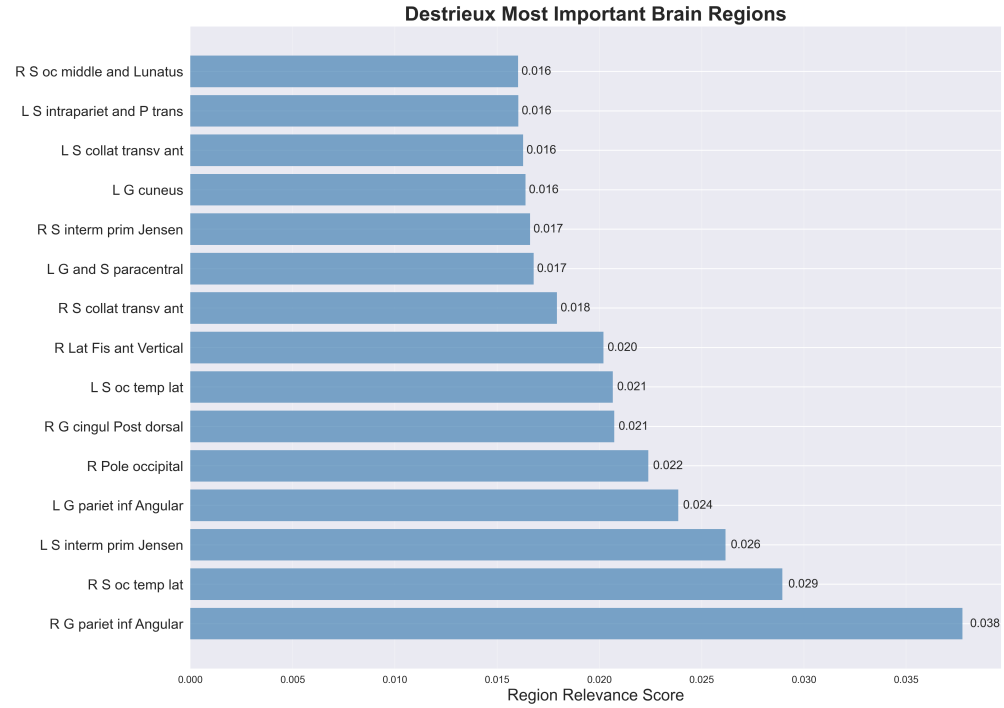

Figure 13: Most important brain regions for the Destrieux atlas ranked by GMLVQ relevance scores. These scores are the dimensionless relevance weights learned by GMLVQ, representing the relative importance of each brain region for classification.

From Figure 13 we can see that the Destrieux atlas, the most important regions included the R G parietal inf Angular (0.038), R S oc temp lat (0.029), and L S interm prim Jensen (0.026). In Figure 14, a visualization can be seen of spatial position of these most important regions. In figure 15, we can see the metabolic deviations from healthy controls.

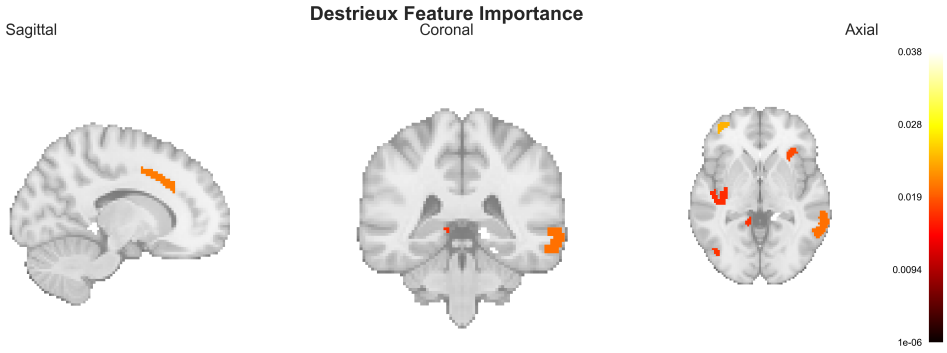

Figure 14: Visualization of the most important regions of the Destrieux atlas

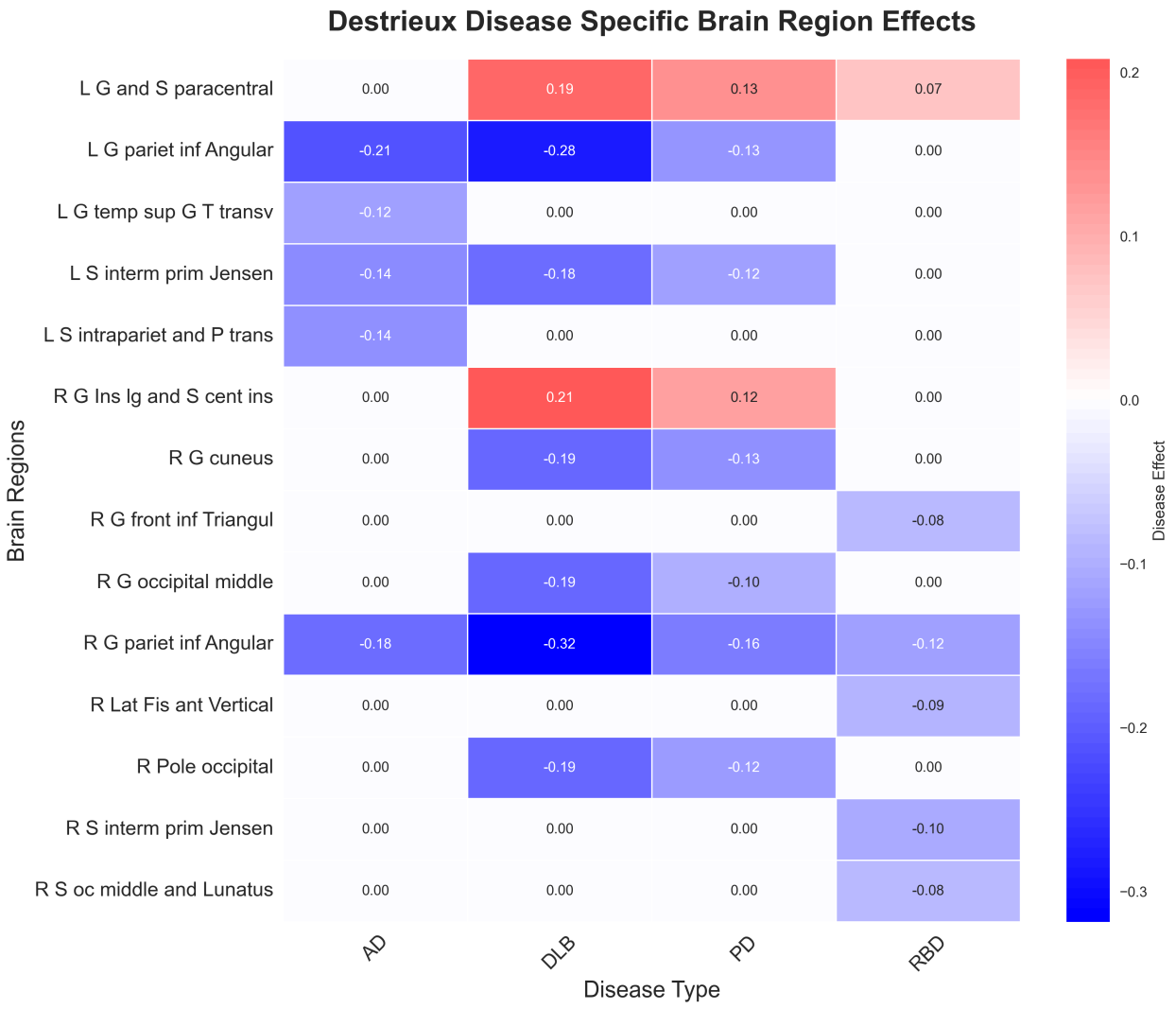

Figure 15: Disease-specific brain region metabolic deviations for the Destrieux atlas. The heatmap displays metabolic deviations from healthy controls, with red indicating hypermetabolism and blue indicating hypometabolism compared to healthy controls. Disease effect values determined by Equation 26

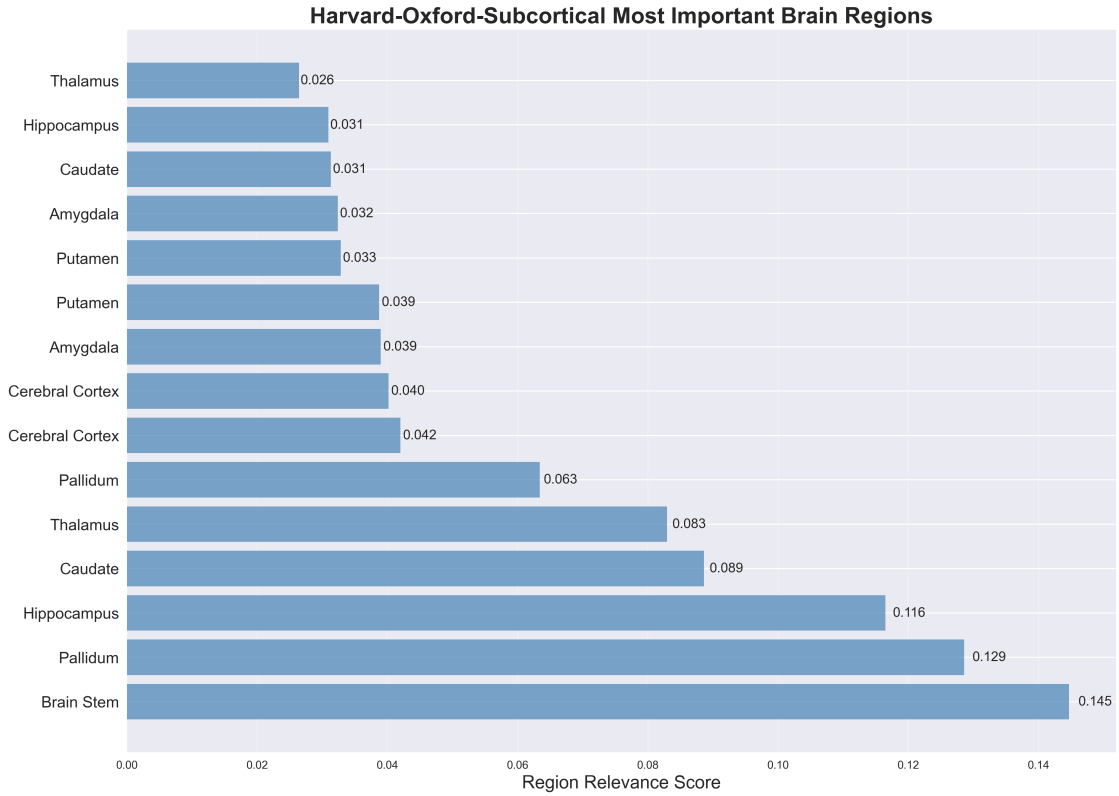

Figure 16: Most important brain regions for the Harvard-Oxford-Subcortical atlas ranked by GMLVQ relevance scores. These scores are the dimensionless relevance weights learned by GMLVQ, representing the relative importance of each brain region for classification.

From Figure 16 we can see that the Harvard-Oxford-Subcortical atlas, the most discriminative brain regions included the Brain Stem (relevance score: 0.145), Pallidum (0.129), and Hippocampus (0.116). In Figure 17, a visualization can be seen of spatial position of these most important regions. In figure 18 we can see the metabolic deviations from healthy controls.

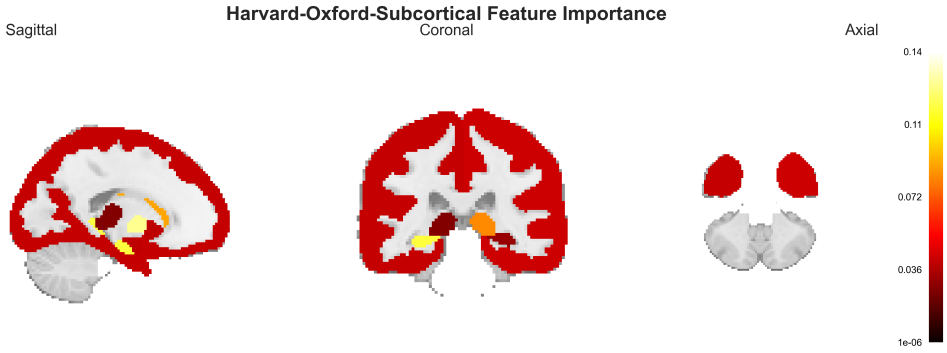

Figure 17: Visualization of the most important regions of the Harvard-Oxford-Subcortical atlas

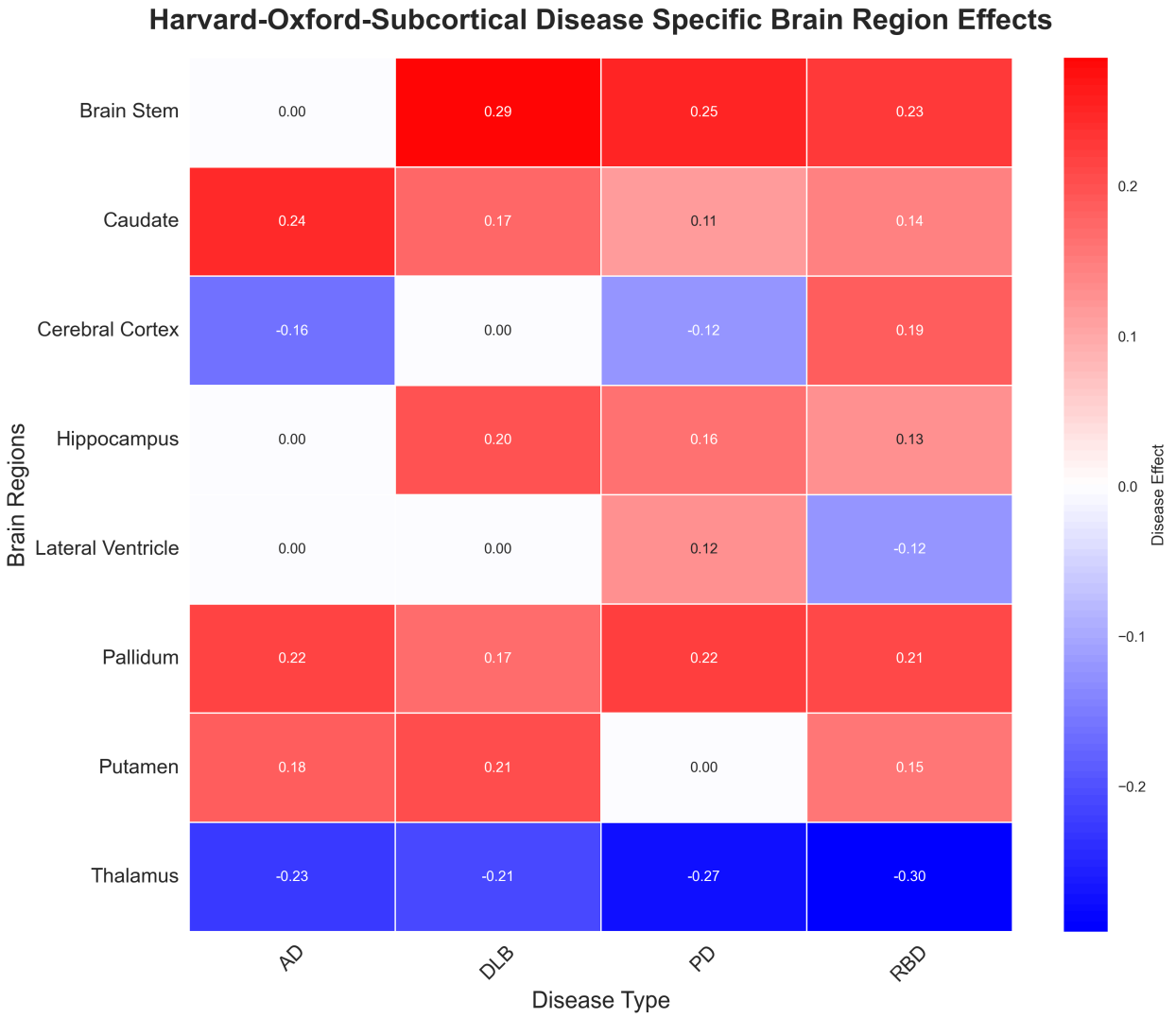

Figure 18: Disease-specific brain region metabolic deviations for the Harvard-Oxford-Subcortical atlas. The heatmap displays metabolic deviations from healthy controls, with red indicating hypermetabolism and blue indicating hypometabolism compared to healthy controls. Disease effect values determined by Equation 26

4.4.2 Atlas Types Performance

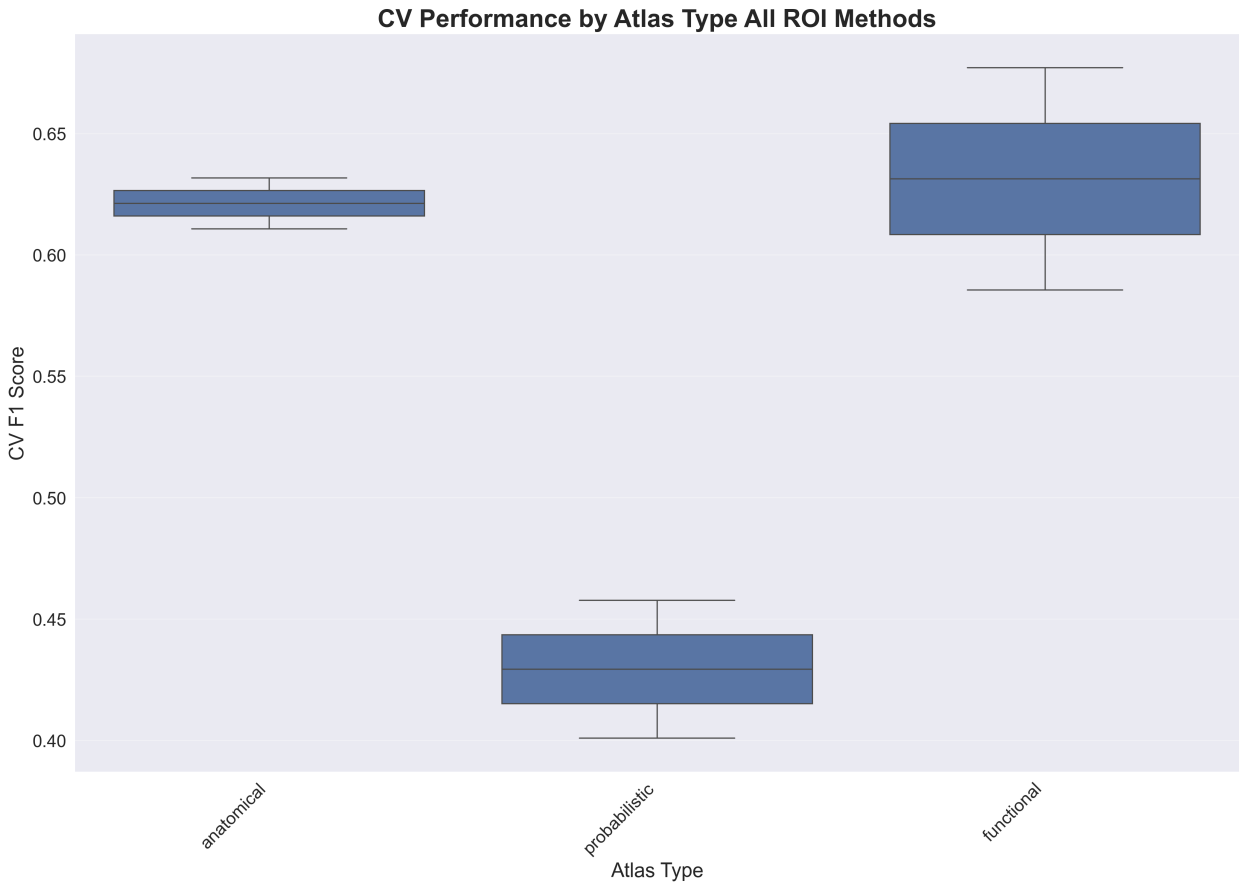

Figure 19: Classification performance comparison by atlas type. Functional atlases achieve median F1-scores of ≈ 0.632, anatomical atlases achieve ≈ 0.622, but probabilistic atlases achieve ≈ 0.430.

From Figure 19, we can see that the atlas parcellation type significantly impacts the classification performance, with functional and anatomical atlas types outperforming probabilistic ones. This suggests that connectivity-based atlases capture metabolic patterns more effectively than purely anatomical ones.

5 Study Limitations

5.1 Sample Size and Dataset

Our study had a small number of patients (189 total), making it harder to find small differences between diseases and means our results might not translate well to other datasets. Additionally, keeping RBD patients in their own separate group made our individual groups even smaller.

5.2 Technical Problems

For the PCA method, we could only use a limited number of components because our reference dataset was small. We also had to set aside some data just to set up PCA, leaving us with less data for training our models.

For the ROI approach, we took the average of all the values within each region. This approach throws away information from within each region that might actually be useful. We could have used other aggregation methods and compared the performance of those in reconstruction and classification.

5.3 Real-World Use

Our study only looked at how accurate the information retention and classifications were. What we didn't consider is how well these results translate to the real world and how confident we can be in the results, when applied in practice.

6 Conclusion

6.1 Method Performance

Principal Component Analysis showed to have both great reconstruction and classification performance. This is likely due to probably is the case because of the way it captures variance, as it purely focuses on trying to maximize the variance. Region-of-Interest methods showed seem to perform much worse when it comes to reconstruction quality, but achieved comparable and even slightly better classification performance. This indicates that dimensionality reduction using an anatomic method like ROI-aggregation is still able to capture relevant patterns despite heavy information loss. We also see that the performance of connectivity-based functional atlases is comparable to, and sometimes even better than, anatomical parcellations, which supports the hypothesis that neurodegenerative diseases primarily alter functional brain networks.

6.2 Atlas Selection

Atlas choice makes a huge difference in the classification performance, with the best results achieved at higher functional parcellation granularity. The performance of Schaefer-200 functional atlas compared to both lower-dimensional and lower-granularity functional shows that for functional atlases, higher granularity gives better performance.

So, we can conclude that PCA is a better choice in most cases over ROI-aggregation. However, for ROI, functional connectivity-based with fine parcellations and anatomical atlases are the best choice for neurodegenerative disease classification. The 65.3% brain and 47.8% coverage of the Schaefer-200 and Destrieux atlases, respectively, proved to be sufficient for this analysis. This suggests that complete brain coverage may be unnecessary and counterproductive due to potentially including irrelevant regions and, depending on the granularity, increasing computation times.

References

1. Feigin, V. L. *et al.* Global, regional, and national burden of neurological disorders during 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. *The Lancet Neurology* **16,** 877–897 (2017).
2. Nichols, E. *et al.* Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: an analysis for the Global Burden of Disease Study 2019. *The Lancet Public Health* **7,** e105–e125 (2022).
3. Phelps, M. E. Positron emission tomography provides molecular imaging of biological processes. *Proceedings of the National Academy of Sciences* **97,** 9226–9233 (2000).
4. Sokoloff, L. *et al.* The [14C] deoxyglucose method for the measurement of local cerebral glucose utilization: theory, procedure, and normal values in the conscious and anesthetized albino rat. *Journal of Neurochemistry* **28,** 897–916 (1977).
5. Frisoni, G. B., Fox, N. C., Jack Jr, C. R., Scheltens, P. & Thompson, P. M. The clinical use of structural MRI in Alzheimer disease. *Nature Reviews Neurology* **6,** 67–77 (2010).
6. Thompson, P. M. *et al.* Dynamics of gray matter loss in Alzheimer's disease. *Journal of Neuroscience* **23,** 994–1005 (2003).
7. Tang, C. C., Poston, K. L., Dhawan, V. & Eidelberg, D. Abnormalities in metabolic network activity precede the onset of motor symptoms in Parkinson's disease. *Journal of Neuroscience* **30,** 1049–1056 (2010).
8. Mosconi, L. Brain glucose metabolism in the early and specific diagnosis of Alzheimer's disease. *European Journal of Nuclear Medicine and Molecular Imaging* **32,** 486–510 (2005).
9. Jolliffe, I. T. *Principal component analysis* 2nd (Springer, New York, 2002).
10. Tzourio-Mazoyer, N. *et al.* Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage* **15,** 273–289 (2002).
11. Teipel, S. J. *et al.* Multivariate network analysis of fiber tract integrity in Alzheimer's disease. *NeuroImage* **34,** 985–995 (2007).
12. Eidelberg, D. Metabolic brain networks in neurodegenerative disorders: a functional imaging approach. *Trends in Neurosciences* **32,** 548–557 (2009).
13. Schölkopf, B., Smola, A. & Müller, K.-R. Kernel principal component analysis. *Artificial Neural Networks—ICANN'97,* 583–588 (1997).
14. Lee, J. A. & Verleysen, M. Nonlinear dimensionality reduction. *Information Sciences* **138,** 1–37 (2007).
15. Schaefer, A. *et al.* Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. *Cerebral Cortex* **28,** 3095–3114 (2018).
16. Rolls, E. T., Huang, C.-C., Lin, C.-P., Feng, J. & Joliot, M. Automated anatomical labelling atlas 3. *NeuroImage* **206,** 116189 (2020).

17. Klein, A. *et al.* Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *NeuroImage* **46,** 786–802 (2009).

18. Van Veen, R. *et al.* FDG-PET combined with learning vector quantization allows classification of neurodegenerative diseases and reveals the trajectory of idiopathic REM sleep behavior disorder. *Computer Methods and Programs in Biomedicine* **225,** 107042 (2022).

19. Minoshima, S. *et al.* Metabolic reduction in the posterior cingulate cortex in very early Alzheimer's disease. *Annals of Neurology* **42,** 85–94 (1997).

20. Diehl-Schmid, J. *et al.* Frontotemporal dementia: patient characteristics, cognition, and behaviour. *International Journal of Geriatric Psychiatry* **19,** 894–901 (2004).

21. Lobotesis, K. *et al.* Occipital hypoperfusion on SPECT in dementia with Lewy bodies but not AD. *Neurology* **56,** 643–649 (2001).

22. Long, Z. *et al.* A Multi-Modal and Multi-Atlas Integrated Framework for Identification of Mild Cognitive Impairment. *Brain Sciences* **12.** ISSN: 2076-3425. `https://www.mdpi.com/2076-3425/12/6/751` (2022).

23. Destrieux, C., Fischl, B., Dale, A. & Halgren, E. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *NeuroImage* **53,** 1–15 (2010).

24. Rolls, E. T., Joliot, M. & Tzourio-Mazoyer, N. Implementation of a new parcellation of the orbitofrontal cortex in the automated anatomical labeling atlas. *NeuroImage* **122,** 1–5 (2015).

25. Frazier, J. A. *et al.* Structural brain magnetic resonance imaging of limbic and thalamic volumes in pediatric bipolar disorder. *American Journal of Psychiatry* **162,** 1256–1265 (2005).

26. Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W. & Smith, S. M. FSL. *NeuroImage* **62,** 782–790 (2012).

27. Hammer, B. & Villmann, T. Generalized relevance learning vector quantization. *Neural Networks* **15,** 1059–1068 (2002).

28. Bittrich, S. *et al.* Application of an interpretable classification model on Early Folding Residues during protein folding. *BioData Mining* **12** (Jan. 2019).

29. Schneider, P., Hammer, B. & Biehl, M. Adaptive relevance matrices in learning vector quantization. *Neural Computation* **21,** 3532–3561 (2009).

30. Kohavi, R. *et al.* A study of cross-validation and bootstrap for accuracy estimation and model selection. *International Joint Conference on Artificial Intelligence* **14,** 1137–1145 (1995).

31. Postuma, R. B. *et al.* Risk and predictors of dementia and parkinsonism in idiopathic REM sleep behaviour disorder: a multicentre study. *Brain* **142,** 744–759 (2019).

32. Van Veen, R., Biehl, M. & de Vries, G.-J. sklvq: Scikit learning vector quantization. *Journal of Machine Learning Research* **22,** 1–6 (2021).