ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

Autoencoders versus PCA for feature extraction in FDG PET scans in neurodegenerative diseases

Roland J. Veen[1,2], S. Sofie Lövdal[1,3a], Kaitlin X. Vos[1], Ciro Setolino[1,4], Sanne K. Meles[3b] and Michael Biehl[1] *

1- Univ. of Groningen, Bernoulli Institute for Mathematics, Computer Science and Artificial Intelligence, Groningen, The Netherlands

2- Medical Research Council Laboratory of Medical Sciences (MRC LMS) London, United Kingdom

3- University Medical Center Groningen, Groningen, The Netherlands
a. Dept of Nuclear Medicine and Molecular Imaging — b. Dept of Neurology

4- University of Salerno - Department of Information Engineering, Electrical Engineering and Applied Mathematics (DIEM) Fisciano - Italy

Abstract. Positron Emission Tomography (PET) neuroimaging is a valuable tool for studying neurodegenerative disorders. Using $N = 236$ FDG PET scans from healthy individuals and three patient classes, we compare linear and non-linear feature extraction using Principal Component Analysis (PCA), a convolutional autoencoder (CAE) and a variational autoencoder (VAE). We investigate whether non-linear dimensionality reduction improves disease classification performance when used with Generalised Matrix Learning Vector Quantisation (GMLVQ) classifiers trained in the latent space. Although PCA had a smaller reconstruction error between the original and reconstructed images, the features from both AEs achieved higher classification performance, with the VAE showing a slight advantage. The interpretability of the AE-GMLVQ combination was retained by visualising the GMLVQ classification space and the decoded prototypes in voxel space. Even with limited training data, using AE for feature extraction improved classification performance by a significant margin while maintaining interpretability.

1 Introduction

Two of the main aims of machine learning in neuroimaging are automating diagnosis and interpreting models to better understand the underlying disease processes. Deep learning models focus on maximising performance, with limited insight into the workings of the parameters, while explainable models offer interpretability at the cost of lower model complexity. Neuroimaging with Positron Emission Tomography (PET) enables the study of various biological processes in the living human brain. This makes it an invaluable tool for the study of neurodegenerative disorders: Incurable, progressive conditions whose prevalence is increasing worldwide due to longer life expectancies [1]. Differential diagnosis in

*R.J. Veen and S.S. Lövdal share first authorship. We thank the Center for IT of the UG for their support and for providing access to the Hábrók high-performance computing cluster.

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

an early stage of the neurodegenerative disease process is crucial but challenging due to overlapping symptoms, which include cognitive decline, motor issues, and psychiatric changes. PET scanning with [^{18}F]FDG can support the diagnostic procedure by reflecting brain metabolic activity, with different diseases showing patterns of regionally increased or decreased glucose uptake [2]. When considering FDG PET for the multi-class classification of neurodegenerative diseases, datasets are typically small. At the same time, interpretability is critical, both when used as a diagnostic tool and to further understand the underlying disease mechanisms, while it is not desirable to trade off high classification performance for the explainability of a (potentially) lower-performing model. Therefore, it is currently unclear whether shallow or deep learning provides the overall best benefit in this context, and methods that combine both properties/aims are needed.

Using FDG PET scans from Alzheimer's disease (AD), Parkinson's disease (PD), dementia with Lewy bodies (DLB) and healthy controls (HC), we compare linear to non-linear feature extraction using PCA, a convolutional, and a variational autoencoder (AE). Quality is evaluated using the corresponding reconstruction error, as well as classification performance using Generalised Matrix Learning Vector Quantisation (GMLVQ) [3, 4] trained in the latent space. Furthermore, the resulting prototypes are decoded with the corresponding decoder and visualised in voxel space. This approach was previously proposed in [5], where it was applied to the two-dimensional MNIST dataset as a proof-of-concept. However, it is not clear how the approach performs in a much higher-dimensional setting with limited training data, and how the interpretability compares to the linear baseline. Van Veen et al. [6] previously laid the foundation for a fully interpretable neurodegenerative disease classification scenario by combining PCA with GMLVQ, using the same dataset and classes as in this work.

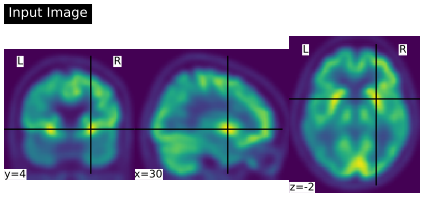
Previous work combining AE with FDG PET has evaluated how the reconstruction error varies depending on which classes were included in training [7], and has used AE as a feature selection method to improve classification performance [8].

This paper has two novel contributions. First, we aim to clarify the performance of deep learning-based vs linear (PCA) feature extraction in molecular imaging for the classification of neurodegenerative disorders. Second, we further expand on the methodology proposed in [5], combining AE with GMLVQ to improve interpretability in the deep learning domain, and evaluate its interpretability relative to its PCA-based counterpart.

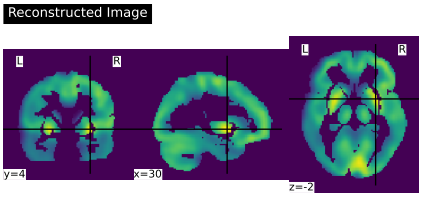
2 Methods

2.1 Dataset: This work involved FDG PET scans of 63 AD patients, 41 PD patients, 23 DLB patients, and 69 HC, all diagnosed according to clinical criteria. Additionally, two scans each from 20 subjects with REM-sleep behaviour disorder (RBD, a prodromal stage of PD and DLB) were included only in the feature extraction step (PCA or AE training), for a total of 236 images. Specific

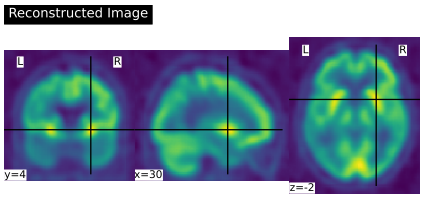
ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.



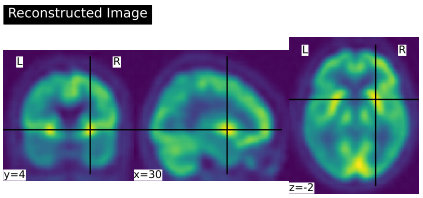
(a) Original



(b) PCA



(c) CAE



(d) VAE

Fig. 1: Example of an input with its reconstruction by each method

imaging protocols and additional patient information are detailed in [6], where this work included an additional 9 AD patients. Here, the scans were preprocessed according to the pipeline in [9]. For the PCA application, the images were masked to include only voxels inside the brain [9], while the unmasked 3D volumes were used for AE training.

2.2 Models: We considered three feature extraction methods: Principal Component Analysis (PCA), a Convolutional Autoencoder (CAE), and a Convolutional Variational Autoencoder (VAE). All have the advantage of being reversible, thereby preserving most of the interpretability from the latent space. For PCA, we preserved 59 principal components in each application, matching the dimensionality used in [6]. For both AEs, we also realised latent dimensions of 59. The CAE was built according to an architecture with two layers (output sizes 8 and 16), both followed by leaky ReLU activation with negative slope 0.05. Kernel size and stride were set to 3 and 2, respectively. A linear layer was added to produce the 59-dim latent space. The VAE layers were similar to the CAE, but with linear layers to produce the mean and log-variance of the latent space representation. The input to the AEs consisted of full FDG-PET brain volumes ($91 \times 109 \times 91$), and the loss function consisted of the mean squared error, computed only within the masked brain volume, in combination with the KL divergence term for the VAE. Both corresponding decoders represented a reverse architecture compared to the encoders. GMLVQ was run with 50 steps of waypoint gradient descent, soft+ activation with $\beta = 1$, and initial step sizes of 0.1 and 0.01 for the prototypes and relevance matrix, according to [6]. The AE architectures were chosen as representative examples, and optimising the model architecture is referred to as future work.

2.3 Validation pipeline: The dataset was stratified and divided into five folds.

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

Subsequently, the AEs were trained (and the PCA was fitted) on four of the five folds, and the remaining fold was used for validation, with the process repeated five times per method. The reconstruction error was measured by evaluating the difference between the original and reconstructed image after projecting the PCA feature vectors back into voxel space, or using the decoder pair of the corresponding AE to reconstruct the image from its latent space representation. The GMLVQ classifier was then trained and validated in matching 80/20 splits in the latent space of each model (transformed with PCA or AE). However, the RBD patients were excluded from the classification step, and from each validation set when computing the reconstruction error. This is because, while they are beneficial for fitting or learning, they are less suitable for classification due to the inclusion of two measurements per patient, and since the eventual disease label is uncertain (the majority of RBD patients progress to either PD or DLB). Hence, the GMLVQ model was trained with four classes (HC-PD-DLB-AD). The dataset was projected onto the two leading eigenvectors (most important discriminative directions) of the trained GMLVQ model and visualised in 2D for each of the three methods. We measured the reconstruction error in the validation sets using mean absolute error (MAE) and structural similarity index measure (SSIM) of the voxels within the brain mask, and compute accuracy, balanced accuracy and area under the ROC curve as performance measures for the classification.

3 Results and Discussion

Figure 1 shows an example of input-reconstruction pairs for an arbitrarily selected AD patient. Visually, all methods produced slightly imperfect reconstructions. The AE reconstructions resembled slightly blurrier versions of the original, while the PCA reconstruction showed a less detailed, more general representation of the input image.

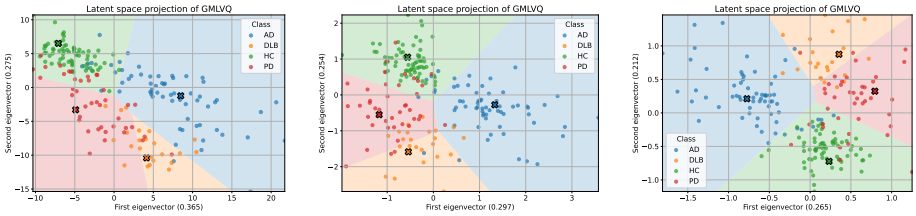
The mean reconstruction error and classification performances of the 5-fold cross-validation procedure for each of the three approaches are shown in Table 1. PCA had the lowest reconstruction error across both MAE and SSIM, and in particular for SSIM, a significant difference was observed. However, this advantage was not reflected in the classification performances. The VAE showed the highest performance (BAC 0.72 and AUC 0.92), closely followed by the CAE (BAC 0.69 and AUC 0.90), followed by PCA (BAC 0.61 and AUC 0.86). This indicates that while the features extracted by the AE overall preserved less of the original information, they were still more specific to disease-related changes. This implies that the images contain some non-linear information that, when extracted by the AE, helps discriminate between the included diseases. The balanced accuracy obtained for PCA is slightly lower than the result reported by van Veen et al. [6] (BAC 0.66) using the same dataset and diseases, while the AUC obtained is the same. However, a designated set of selected patients was used there for the PCA application only, which may have improved the generalisability of the extracted features, and the pipelines are not fully comparable.

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

	MAE	SSIM	ACC	BAC	AUROC
PCA	5.2 ± 0.1	99.0 ± 0.0	0.67 ± 0.05	0.61 ± 0.07	0.86 ± 0.03
CAE	5.5 ± 2.3	84.6 ± 4.7	0.75 ± 0.06	0.69 ± 0.09	0.90 ± 0.02
VAE	5.6 ± 2.4	84.3 ± 4.1	0.77 ± 0.08	0.72 ± 0.09	0.92 ± 0.03

Table 1: Reconstruction error (RE) and mean five-fold cross-validation scores with standard deviation when classifying HC-PD-DLB-AD with GMLVQ, shown for three different feature extraction methods: PCA, a convolutional AE, and a variational AE. RE shown as the mean value multiplied by 100.

To visualise the model space, we projected the features onto the first two eigenvectors of the GMLVQ relevance matrix (Figure 2). While the GMLVQ space for PCA shows a spectrum with partially overlapping distributions and misclassified outliers, both the CAE and the VAE showed more clustered classes. Interestingly, the portion of misclassified PD patients falling into the HC space ended up in the middle of the VAE cluster, rather than closer to the boundary, compared to PCA. Reconstructed prototypes for the VAE are shown in Figure 3. Here, the prototypes were decoded and voxel values were z-scored to the mean and standard deviation of HC for display. The brain slices show regions of relative hyper- (red) and hypometabolism (blue), and were highly similar to previously found typical metabolic patterns in PD, DLB and AD (compare e.g. [6]).



(a) PCA + GMLVQ (b) CAE + GMLVQ (c) VAE + GMLVQ

Fig. 2: Projection of the data and prototypes on the first two eigenvectors of the relevance matrix of GMLVQ when combined with the different feature extractors.

4 Conclusion and further work

We have seen that, even with limited training data, using AE for feature extraction significantly improved classification performance for discriminating between neurodegenerative disorders, compared to PCA (BAC 0.61 vs 0.72, and AUC 0.84 vs 0.92). Since the reconstruction error was worse for the AE, this indicates that the regions in the FDG PET scans exhibit non-linear dependencies relevant to the disease process. Future work should clarify the nature and causes of these relations, how the performance of the methods scales with larger amounts of training data, and evaluate additional methods and models.

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964.
Available from http://www.i6doc.com/en/.

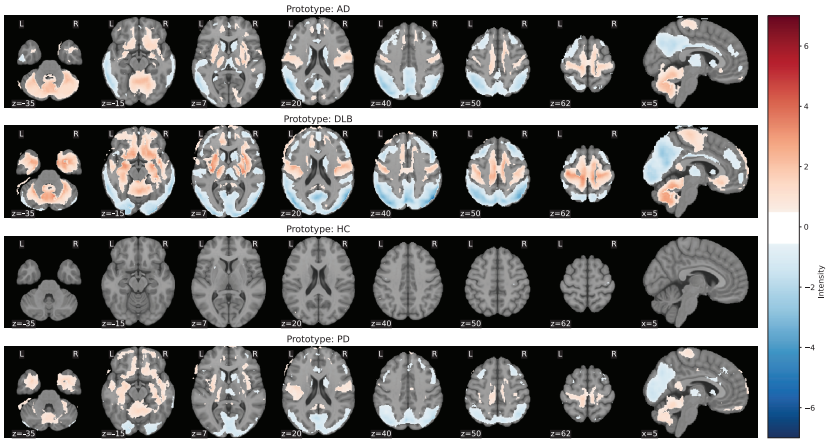


Fig. 3: Decoded prototypes for the VAE. Negative values (blue) represent relative hypometabolism, while positive values (red) represent relatively higher glucose uptake. The voxel values are shown as a z-score with respect to HC.

References

[1] E Nichols, J D Steinmetz, S E Vollset, K Fukutaki, 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(2):e105–e125, 2022.

[2] F Nobili, C Festari, D Altomare, et al. Automated assessment of FDG-PET for differential diagnosis in patients with neurodegenerative disorders. *European journal of nuclear medicine and molecular imaging*, 45(9):1557–1566, 2018.

[3] Atsushi Sato and Keiji Yamada. Generalized learning vector quantization. *Advances in neural information processing systems*, 8, 1995.

[4] Petra Schneider, Michael Biehl, and Barbara Hammer. Adaptive relevance matrices in learning vector quantization. *Neural Computation*, 21(12):3532–3561, 2009.

[5] Roland J. Veen, Christodoulos Hadjichristodoulou, and Michael Biehl. Interpreting hybrid AI through autodecoded latent space entities. In *Proc. Europ. Symp. Artificial Neural Networks (ESANN) 2024*, pages 267–272, 01 2024.

[6] Rick van Veen, Sanne K. Meles, Remco J. Renken, Fransje E. Reesink, Wolfgang H. Oertel, Annette Janzen, Gert-Jan de Vries, Klaus L. Leenders, and Michael Biehl. 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.

[7] R John, J Penning, H Chandler, P Fielding, C Marshall, and R Smith. Quantitative evaluation of synthesized brain PET using a variational autoencoder. In *2021 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, pages 1–4, 2021.

[8] Pham Minh Tuan, Trong-Le Phan, Mouloud Adel, Eric Guedj, and Nguyen Linh Trung. Autoencoder-based feature ranking for Alzheimer Disease classification using PET image. *Machine Learning with Applications*, 6:100184, 2021.

[9] S S Lövdal, R van Veen, G Carli, R J Renken, T Shiner, N Bregman, R Orad, D Arnaldi, B Orso, S Morbelli, P Mattioli, K L Leenders, R Dierckx, S K Meles, M Biehl, and for the Alzheimer's Disease Neuroimaging Initiative. IRMA: Machine learning-based harmonization of ^{18}F-FDG PET brain scans in multi-center studies. *European Journal of Nuclear Medicine and Molecular Imaging*, 52(8):2941–2958, July 2025.