

University of Groningen

FDG-PET combined with learning vector quantization allows classification of neurodegenerative diseases and reveals the trajectory of idiopathic REM sleep behavior disorder

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

Published in:
Computer Methods and Programs in Biomedicine

DOI:
[10.1016/j.cmpb.2022.107042](https://doi.org/10.1016/j.cmpb.2022.107042)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2022

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Veen, R. V., Meles, S. K., Renken, R. J., Reesink, F. E., Oertel, W. H., Janzen, A., de Vries, G.-J., Leenders, K. L., & Biehl, M. (2022). 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]. <https://doi.org/10.1016/j.cmpb.2022.107042>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

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

FDG-PET combined with learning vector quantization allows classification of neurodegenerative diseases and reveals the trajectory of idiopathic REM sleep behavior disorder

Rick van Veen^{a,b,*}, Sanne K. Meles^c, Remco J. Renken^d, Fransje E. Reesink^c, Wolfgang H. Oertel^{e,f}, Annette Janzen^e, Gert-Jan de Vries^h, Klaus L. Leenders^g, Michael Biehl^{a,i}

^a Bernoulli Institute for Mathematics, Computer Science and Artificial Intelligence, University of Groningen, Groningen, the Netherlands

^b Data Science Department, Software Competence Center Hagenberg, Hagenberg, Austria

^c Department of Neurology, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands

^d Department of Biomedical Sciences of Cells & Systems, University of Groningen, University Medical Center Groningen, Cognitive Neuroscience Center, Groningen, the Netherlands

^e Department of Neurology, Philipps-Universität Marburg, Marburg, Germany

^f Institute for Neurogenomics, Helmholtz Center for Health and Environment, Munich, Germany

^g Department of Nuclear Medicine and Molecular Imaging, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands

^h Philips Research, Eindhoven, the Netherlands

ⁱ SMQB, Institute of Metabolism and Systems Research, College of Medical and Dental Sciences, Birmingham, United Kingdom

ARTICLE INFO

Article history:

Received 1 April 2022

Revised 11 July 2022

Accepted 27 July 2022

Keywords:

Neurodegenerative diseases
Idiopathic REM sleep behavior disorder
trajectories
FDG-PET
SSM/PCA
Learning vector quantization
Relevance learning

ABSTRACT

Background and Objectives: ¹⁸F-fluorodeoxyglucose (FDG) positron emission tomography (PET) combined with principal component analysis (PCA) has been applied to identify disease-related brain patterns in neurodegenerative disorders such as Parkinson's disease (PD), Dementia with Lewy Bodies (DLB) and Alzheimer's disease (AD). These patterns are used to quantify functional brain changes at the single subject level. This is especially relevant in determining disease progression in idiopathic REM sleep behavior disorder (iRBD), a prodromal stage of PD and DLB. However, the PCA method is limited in discriminating between neurodegenerative conditions. More advanced machine learning algorithms may provide a solution. In this study, we apply Generalized Matrix Learning Vector Quantization (GMLVQ) to FDG-PET scans of healthy controls, and patients with AD, PD and DLB. Scans of iRBD patients, scanned twice with an approximate 4 year interval, were projected into GMLVQ space to visualize their trajectory.

Methods: We applied a combination of SSM/PCA and GMLVQ as a classifier on FDG-PET data of healthy controls, AD, DLB, and PD patients. We determined the diagnostic performance by performing a ten times repeated ten fold cross validation. We analyzed the validity of the classification system by inspecting the GMLVQ space. First by the projection of the patients into this space. Second by representing the axis, that span this decision space, into a voxel map. Furthermore, we projected a cohort of RBD patients, whom have been scanned twice (approximately 4 years apart), into the same decision space and visualized their trajectories.

Results: The GMLVQ prototypes, relevance diagonal, and decision space voxel maps showed metabolic patterns that agree with previously identified disease-related brain patterns. The GMLVQ decision space showed a plausible quantification of FDG-PET data. Distance traveled by iRBD subjects through GMLVQ space per year (i.e. velocity) was correlated with the change in motor symptoms per year (Spearman's $\rho = 0.62$, $P = 0.004$).

Conclusion: In this proof-of-concept study, we show that GMLVQ provides a classification of patients with neurodegenerative disorders, and may be useful in future studies investigating speed of progression in prodromal disease stages.

© 2022 The Authors. Published by Elsevier B.V.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

* Corresponding author.

E-mail address: rick.van.veen@scch.at (R. van Veen).

1. Introduction

Due to the aging of the population, neurodegenerative diseases such as Alzheimer's disease (AD) and Parkinson's disease (PD) will become the second leading cause of death worldwide after cancer by 2040 [1]. Accurate and early diagnosis of these conditions is needed to develop prevention strategies and precision therapeutic measures. However, this poses a significant challenge in clinical practice because neurodegenerative diseases are diagnosed based on clinical features. In the case of PD, motor symptoms in the early stages are subtle, and the distinction between PD and other disorders of the parkinsonian clinical spectrum can be difficult [2,3]. Early AD diagnosis and differentiation with other neurodegenerative dementias is equally challenging [4]. Furthermore, overlap exists between parkinsonian and dementing conditions. For instance, many (but not all) patients with PD will develop dementia in the course of their disease. In addition, a separate entity referred to as Dementia with Lewy Bodies (DLB) shares clinical and neuropathological characteristics with PD and AD. Compared with PD patients, DLB patients have an earlier onset of severe cognitive impairments, often including visual hallucinations. Compared with AD patients, memory functions are relatively preserved in DLB. Sometimes it is not clear, clinically, whether a patient has PD-dementia, DLB, or a combination of PD and AD. It is recognized that PD, DLB, and AD are part of a neuropathological (and clinical) spectrum, with considerable heterogeneity at the individual level [5]. Still, it is essential to provide a correct diagnosis because these conditions require specific clinical management [6].

New therapeutic interventions for neurodegenerative disorders may have a better chance of success if initiated in the earliest disease stages. Patients with 'idiopathic Rapid Eye Movement sleep behavior disorder (iRBD) [7] may be ideal candidates for clinical trials. This sleep condition can be diagnosed accurately with polysomnography. Longitudinal studies have shown that > 80% of patients initially diagnosed with iRBD developed PD, DLB, or incidentally multiple system atrophy (MSA) in the following decades [8,9]. Patients with iRBD, by definition, have not yet developed the motor or cognitive symptoms and represent the earliest stage of PD or DLB that can be diagnosed. However, the speed of progression from preclinical or prodromal to full clinical stages varies among patients and cannot be reliably predicted on the individual level. Clinical trials will require patients with a predictable conversion within a limited time window.

Following from the above, it is clear that biomarkers are needed to confirm the presence of neurodegenerative disease in the early stages, provide information regarding lead time and speed of progression, and finally that can differentiate between neurodegenerative conditions. When combined with advanced computational algorithms, imaging the brain with ^{18}F -2-fluoro-2-deoxy-D-glucose Positron Emission Tomography (FDG-PET) may provide a solution. The radiotracer FDG provides an index for the cerebral metabolic rate of glucose, which is strongly associated with neuronal activity [10]. Due to local pathology, regions with decreased FDG uptake can reflect impaired neuronal function. However, regional hypometabolism in a brain region unaffected by the pathology can also be secondary to dysfunction of another region if these two regions are organized in the same functional brain network. More severe changes in regional cerebral glucose metabolism can often be appreciated visually, which makes FDG-PET a useful ancillary investigation in the routine clinical workup of neurodegenerative disorders [11,12].

A (semi-)quantitative analysis of FDG-PET images involves calculating average FDG uptake in pre-defined regions or a voxel-by-voxel comparison (a *t*-test) between patients and controls. However, such methods disregard any interactions that may exist between brain regions. In contrast, the covariance relation-

ships between voxels can provide helpful information to identify the affected brain networks and quantify individual subjects. Spatial covariance analysis of FDG-PET images can be performed using Scaled Subprofile Model and Principal Component Analysis (SSM/PCA). SSM/PCA reduces the large number of voxels for every subject to a limited number of orthogonal dimensions (eigenvectors) that explain the major sources of variance in the data. A disease-related pattern (or 'network') is identified among the eigenvectors that discriminate between controls and patients [13,14]. Apart from visualizing the involved brain regions, disease-related patterns can also be used to quantify FDG-PET scans of new subjects, which has allowed an objective assessment of disease activity in individual subjects.

The PD-related pattern (PDRP) has been identified in several independent cohorts of approximately 20 controls and 20 PD patients with SSM/PCA. Its pattern topography is consistent across scanning platforms, and expression of the PDRP (i.e., PDRP subject scores) is consistently higher in PD patients than in controls and increases with disease progression [15]. The PDRP is also significantly expressed in individuals with iRBD, especially in those who phenoconvert to manifest PD within a few years of scanning [16–21]. Similarly, the AD-related pattern (ADRP; [22–25]) and DLB-related pattern (DLBRP; [26]) have been determined. However, simply calculating subject scores for these spatial covariance patterns does not help discriminate between PD, DLB, and AD, due to considerable regional overlap between the PDRP, DLBRP, and ADRP. Indeed, DLB patients were shown to express the PDRP [16] and DLB patients also expressed the ADRP [26].

Consequently, there is a need for an approach to quantify and classify FDG-PET scans in a system that can deal with more than one disease. Machine learning approaches may provide a solution, and their application to FDG-PET data in neurodegenerative disorders is becoming increasingly popular (for review in PD see [27]; for review in AD, see [28]). However, these typically involve black-box methods, which are not transparent in what they learn. In clinical practice, neurologists will require an intuitive diagnostic model that points to similar cases and explains its decision process when classifying a single subject [29,30]. We previously considered transparent or white-box algorithms such as decision trees [31] in the classification of FDG-PET data. However, these are not always intuitive when working with such complex and high dimensional features [31,32].

Learning Vector Quantization (LVQ) is a supervised learning algorithm that can naturally deal with more than two classes and produces interpretable systems, making it an ideal candidate [33]. In LVQ, prototypes are defined that represent the categories in the data on which the system is trained. Predictions are made based on the receptive fields of these prototypes, i.e., a new sample is assigned the same category or class as the closest prototype. Advanced LVQ variants, such as Generalized Matrix LVQ (GMLVQ) [34], improve the performance and robustness of the system by additionally constructing a relevance matrix that weighs the features and their combinations [34,35]. An essential advantage of LVQ (and its variants) over other machine learning algorithms is its interpretability. Specifically, the relevance matrix of GMLVQ can be decomposed into a coordinate system (eigenvectors) that spans its decision space. This 'discriminative visualization' can be used to detect outliers and groups of similar patients [36–38].

In a previous study, we combined FDG-PET, SSM/PCA, and GM-LVQ to classify healthy controls and PD patients from different centers. We showed that GMVLQ could separate data derived from different centers and discriminate between healthy controls and PD [38]. The prototypes can be represented in voxel space, allowing inspection of the learned patterns. In addition, the PD-prototype was almost identical to the PDRP previously derived for each center separately [38,39].

Table 1

Mean and standard deviation are indicated unless otherwise specified. The minimum and maximum is indicated between brackets. H&Y= Hoehn and Yahr stage, indicating the severity of parkinsonian symptoms. MMSE = mini-mental-state examination. n/a = not applicable or not available. * Several screening instruments were used in the PD patients in the evaluation set: scales for outcomes in Parkinson's disease-cognition (SCOPA-COG; cut-off for dementia 24/25), The Mini-Mental State Examination (MMSE; cut-off 23) and the Montreal Cognitive Assessment (MoCA; cut-off 17). Based on the scores that were obtained around the time of FDG-PET imaging, we indicate the number of patients who may have suffered from PD-dementia at the time of imaging. Although only three had formally dementia at the time of imaging, several had cognitive complaints or mild cognitive impairments, with varying severity. ** For iRBD, the MoCA was obtained.

	Age at scanning	Gender, (n male)	Disease duration, (years)	H&Y (median)	MMSE	Clinical follow-up time, (years)
HC (n = 20)	65.3 ± 7.1 [51 – 78]	15	n/a	n/a	29 ± 0.8 [28 – 30]	n/a
PD (n = 20)	64.2 ± 8.5 [49 – 67]	15	4.7 ± 3.5 [2 – 11]	2[1 – 2.5]	28.3 ± 1.2 [26 – 30] (n=13)	10 ± 4.1 [1.3 – 14.2]
AD (n = 20)	68.5 ± 8.6 [55 – 84]	12	2.9 ± 2.4 [0.25 – 10]	n/a	23.8 ± 3.8 [15 – 29] (n = 19)	3.6 ± 3.3 [0.26 – 9.3]
(a) Reference cohort.						
HC (n = 49)	58.4 ± 11.6 [20 – 78]	17	n/a	n/a	n/a	n/a
PD (n = 21)	66.1 ± 11.5 [40 – 78]	14	5.5 ± 5.3 [1 – 22]	1.95 ± 0.7 [1 – 4]	3*	6.8 ± 5.8 [0.6 – 24.6]
AD (n = 36)	65.4 ± 6.4 [53 – 78]	23	3.4 ± 3.8 [1 – 15]	n/a	25.8 ± 3.0 [17 – 30]	2.3 ± 2.2 [0.2 – 9.2]
DLB (n = 23)	72.1 ± 7.6 [57 – 83]	18	2.3 ± 1.1 [1 – 4]	1.7 ± 0.9 [0 – 3]	n/a	2.2 ± 2.0 [0.3 – 9.5]
(b) Evaluation cohort.						
iRBD (n = 20)	66.4 ± 5.2 [53 – 75]	18	9.7 ± 5.6 [5 – 28]	0	27.9 ± 1.4 [25 – 30]**	3.7 ± 0.6 [2.5 – 4.5]
(c) Progression cohort.						

This current study aimed to use the combination of FDG-PET, SSM/PCA, and GMLVQ to discriminate between healthy controls, PD, DLB, and AD. In addition, we determined the positions and trajectories of 20 iRBD patients who were scanned twice (approximately 3.7 years apart) [17], in our GMLVQ space. Furthermore, to provide insight into the GMLVQ decision process, we transformed the prototypes and relevance matrix back to the voxel space, similarly to our previous work [38]. In addition, we also look at the voxel representations of the axes of the decision space.

2. Materials and methods

2.1. Patient inclusion and clinical characteristics

We included resting state FDG-PET scans of patients with a clinical diagnosis of PD (n = 41), AD (n = 56), or DLB (n = 23). Some of these data were published previously [16,24,40,41]. However, some new cases were also added after evaluating a historical database of FDG-PET scans performed in the UMCG in the context of clinical work-up of patients with a neurodegenerative disease between 2010–2018. In addition, we included FDG-PET scans of a total of 69 healthy controls (HC) that could be combined from previous studies [16,24,40,41].

According to diagnostic consensus criteria, the clinical diagnosis of PD, AD, or DLB was made by expert neurologists at the movement disorders and memory outpatient clinics of the University Medical Center Groningen (UMCG) [42–44]. Clinical information of each cohort is provided in Table 1. In some cases, additional biomarkers were used to support the clinical diagnosis (see table A.5 in the supplementary material). This included presynaptic dopaminergic brain imaging with ¹²³I-FP-CIT SPECT (DAT-SPECT) or ¹⁸F-dihydroxyphenylalanine (F-DOPA) PET in patients with suspected PD or DLB. If AD was considered, in some cases, imaging of cerebral amyloid-β deposits with ¹¹C-labelled Pittsburgh Compound B (PiB) PET was performed, or cerebral spinal fluid analysis.

This included presynaptic dopaminergic brain imaging with ¹²³I-FP-CIT SPECT (DAT-SPECT) or ¹⁸F-dihydroxyphenylalanine (F-DOPA) PET in patients with suspected PD or DLB. If AD was considered, in some cases, imaging of cerebral amyloid-β deposits with ¹¹C-labelled Pittsburgh Compound B (PiB) PET was performed, or cerebral spinal fluid analysis. Structural imaging of the brain was performed in all patients to exclude other causes or major vascular disease, usually by magnetic resonance imaging (MRI) of the brain. In some cases, only Computed Tomography (CT) scans of the brain were available.

Most PD and DLB patients were retrospectively included from our historical database. A FDG-PET scan was performed in those patients as part of the clinical workup because there was a diagnostic uncertainty, and other parkinsonian syndromes were considered. Sometimes this was done in the context of a second opinion in our expertise center. Based on sufficiently long follow-up, a definite clinical diagnosis of PD or DLB could be made in each case. In the historical cohort, the clinical data were extracted from the patients' charts. Several screening instruments for cognitive symptoms were used for PD and DLB: scales for outcomes in Parkinson's disease-cognition (SCOPA-COG; cut-off for dementia 23/43), The Mini-Mental State Examination (MMSE; cut-off 23/30) and the Montreal Cognitive Assessment (MoCA; cut-off 17/30). Based on the scores obtained around the time of FDG-PET imaging, we indicate the number of PD patients who may have suffered from PD-dementia at the time of imaging (n = 3). Although only three had dementia formally at the time of imaging, most had a history of cognitive complaints with varying severity. Due to the retrospective nature of these data, the degree of PD-MCI could not be quantified.

Of the AD patients, 16 had mild cognitive impairment (MCI) at the time of FDG-PET imaging based on a more extensive battery of neuropsychological tests. This indicates that cognitive symptoms were not severe enough to diagnose Alzheimer's dementia. However, a diagnosis of MCI as an early disease stage of AD

could be made due to a clear clinical progression over time and/or other supporting biomarkers (positive PiB PET or cerebrospinal fluid markers compatible with AD).

Although AD is usually a primary amnesic syndrome, several subtypes exist that can present differently and make a diagnosis challenging. In our study, six patients had the posterior cortical atrophy (PCA) subtype of AD. AD-PCA is characterized by prominent visuospatial dysfunction, whereas amnesic symptoms may be less pronounced. Two patients had the primary progressive aphasia (logopenic) variant (AD-PPA), characterized by specific language problems.

Patients with a diagnosis of idiopathic RBD ($n = 20$), confirmed with a video-polysomnography, were included from a previous study [17]. These subjects underwent two FDG-PET brain scans approximately 3.7 years apart, with baseline and repeated motor and cognitive testing. Four subjects converted to PD during follow-up.

All subjects gave their written informed consent to participate in the study. The study was approved by the local ethics committee and was conducted in agreement with the Declaration of Helsinki and subsequent revisions.

2.2. FDG-PET imaging and preprocessing

FDG-PET imaging was performed on a Siemens Biograph mCT64 or mCT40 PET/CT camera (Siemens, Munich, Germany) at the UMCG. Images were reconstructed with OSEM3D (3 iterations, 21 subsets), time-of-flight, point-spread-function, Gaussian 8 mm full-width at-half-maximum spatial filter, and matrix size 256 (corresponding to a voxel size of 2 mm \times 3.18 mm \times 3.18 mm). Central nervous system depressants such as benzodiazepines were discontinued in all subjects for at least 24 h before imaging, and dopaminergic medication was not routinely withheld.

All FDG-PET images were spatially normalized to an FDG-PET template in Montreal Neurological Institute (MNI) brain space [45] using SPM5 software (Wellcome Centre for Human Neuroimaging, London, UK) implemented in MATLAB (version R2019a; MathWorks, Natick, MA, USA).

2.3. Definition of cohorts

The data were divided into three cohorts. The reference cohort (20 HC, 20 PD and 20 AD; Table 1a) was used to define the SSM/PCA feature space. The evaluation cohort (49 HC, 36 21 PD, 36 AD, and 23 DLB; Table 1b) can be considered the ‘training’ data. These data were transformed using the principal components of the *reference* group and given as input to the LVQ system to train on. We aimed to include the most typical PD and AD patients for the reference cohort. Thus, PDD, AD-MCI, AD-PCI, or AD-PPA subtypes were *not* included in the *reference* group but were included in the *evaluation* group. The third cohort consists only of 20 iRBD patients (Table 1c). Two FDG-PET scans are available for these patients: one at baseline and one approximately 3.7 years later. This data was not used for training but was later projected into the discriminative space created by GMLVQ (Section 2.5.1).

2.4. Definition of SSM/PCA feature space

First, a 35% threshold of the whole-brain intensity maximum was applied to each FDG-PET image in the *reference* cohort (Table 1a) to remove out-of-brain voxels. These were multiplicatively combined to create one common mask that included only (m) non-zero values for all subjects. This mask was applied to all images. Masked images were log-transformed, and the subject mean was removed. Additionally, each voxel is centered around the mean of the healthy controls included in the reference group. These actions result in a ($m \times 1$) subject residual profile denoted

by \mathbf{r} , in which a positive value indicates a higher relative FDG uptake than the average healthy control, and a negative value a lower FDG uptake.

Principal component analysis (PCA) was applied to the residual profiles of the *reference* group ($n = 60$) in voxel space, resulting in a set of $k = n - 1$ principal components denoted by G ($m \times k$), each explaining a percentage of the variance across all reference subjects. The mask from the *reference* cohort was also applied to the normalized FDG-PET scans in the *evaluation* and *progression* cohorts. All images were log-transformed, and the subject and group mean (i.e., the mean from the *reference* healthy controls) were removed [38]. Subsequently, subject scores (\mathbf{s}) were calculated for each residual profile on each of the 59 principal components determined on the reference cohort using

$$\mathbf{s} = G^T \mathbf{r}. \quad (1)$$

The subject scores of the training data were used in the GMLVQ model. We refer to the scores of a single subject as a feature vector where the expression of a single PC may be referred to as a feature.

2.5. Generalized matrix learning vector quantization

Generalized matrix learning vector quantization (GMLVQ), by Schneider et al. [34], employs an adaptive distance variant denoted by

$$d^\Lambda(\mathbf{w}_j, \mathbf{s}_i) = (\mathbf{s}_i - \mathbf{w}_j)^\top \Lambda (\mathbf{s}_i - \mathbf{w}_j), \quad (2)$$

with \mathbf{w}_j a prototype of the size ($k \times 1$) and $\Lambda = \Omega^\top \Omega$ a symmetric positive semidefinite $k \times k$ matrix. During training the values of the matrix (Ω) are updated in such a way that the distance from the closest correct prototype is decreased and the distance from the closest wrong prototype becomes larger [34]. Furthermore, by weighting every pair of features, GMLVQ is able to account for correlations of dimensions by implicit scaling and rotation of the data, resulting in robust performance [34,35].

GMLVQ was trained on the *subject scores* of the *evaluation* cohort to distinguish between the four classes (HC, PD, AD, and DLB) using the open-source Python package sklvq [46]. Performance was estimated using ten times repeated randomized ten-fold cross-validation. Each system was initialized with a single prototype per class and the initial relevance matrix, the identity matrix, not to bias any features. Optimal values for parameters such as the activation function, solvers, and related parameters have been found using a grid search.

Initially, the classes containing less than half the majority class samples (the class with most samples) were oversampled. Oversampling was done randomly such that the ratio with the majority class is 0.5. In a second step, each other class, i.e., each class that started with more than half the samples compared to the majority class, was randomly undersampled to contain the same number of samples as the oversampled classes. Each class that started with exactly the right number of subjects was left unchanged. By balancing the data, the system will learn not to bias any classes based on their availability.

In addition, each feature was transformed by subtracting the mean and dividing by the standard deviation of the data. This z-transformation was based on the training data within each cross-validation fold and applied to the testing data. Although a z-transformation is not strictly necessary for GMLVQ to find a system with good performance, the interpretation of the relevance matrix improves as it does not need to compensate for any differences in the magnitude of the feature values. For the projections shown in the results section, the z-transformation has been reversed, such that the origin in the plots corresponds to the average healthy control expression from the *reference* group.

The final GMLVQ system is chosen based on the average balanced accuracy determined over the test sets created by the cross-validation procedure. Balanced accuracy is defined as the average of the recall obtained per class and reduces to the conventional accuracy for balanced datasets [47,48]. Additionally, we report the area under the receiver operating characteristic curve (AUC). We use the simple generalization for the multi-class AUC, i.e., the macro-averaged *one vs. one* class AUC [49]. This means that for each binary problem, the AUCs are determined, which are then averaged with equal weights.

2.5.1. Discriminative visualization

After the system is trained the relevance matrix of GMLVQ can be used to construct a low dimensional discriminative visualization of the data [36–38]. We can use this type of visualization, e.g., to identify outliers and find similar data samples [36–38]. Note that ‘similar’ is defined within the context of the data and the classification task. In the case of SSM/PCA it refers to patients that show similar activity patterns.

The adaptive matrix Ω , can be interpreted as a linear transformation of the data and the prototypes. However, given a Λ , the matrix Ω is not uniquely defined. Geometrically, the distance measure (Eq. (2)) is invariant under reflections or rotations so many solutions may exist [36]. The specific outcome of the training will depend on the initialization and the randomized order by which GMLVQ learns from the examples [37]. However, we can define a canonical, unique solution $\hat{\Omega}$ with the eigenvalue decomposition of Λ . Thereby, we determine the orthonormal eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_N$ of Λ , corresponding to the k ordered non-zero eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N$ and define $\hat{\Omega}$ as:

$$\hat{\Omega} = [\sqrt{\lambda_1}\mathbf{v}_1, \sqrt{\lambda_2}\mathbf{v}_2, \dots, \sqrt{\lambda_k}\mathbf{v}_k]. \quad (3)$$

It has been shown analytically [37] and observed empirically that GMLVQ has a strong tendency to yield singular matrices Λ of very low rank [34,36,37,50], so the number of non-zero eigenvalues will be small. Furthermore, because Λ is symmetrical, the eigenvectors are orthogonal. The eigenvectors correspond to orthogonal combinations of features that each describe a part of the label-dependent differences in the data that are relevant for the classification. The corresponding eigenvalues indicate the weight of each of these combinations (eigenvectors).

Given the results of the ten times repeated ten-fold cross-validation procedure of the best model [51], we can compute the orthogonal $\hat{\Omega}$ parametrization of the average Λ and define a new coordinate system. We project the subject scores of both the *evaluation* and *progression* cohorts into the discriminative decision space constructed by GMLVQ similarly to Eq. (1)

$$\hat{\mathbf{s}} = \hat{\Omega}^\top \mathbf{s}. \quad (4)$$

The projection of the *evaluation* cohort (Table 1b) is used to verify the data and model, detect clusters of interesting patients, and outliers. We plot the data from the *progression* cohort (Table 1c) containing two scans of 20 iRBD patients at baseline and follow-up (approximately 3.7 years later) to study their trajectory over time.

2.5.2. Voxel representation

The merit of GMLVQ is that its prototypes and relevance matrix can be interpreted. The relevance matrix provides a weight to each feature, indicating the importance for the classification. This relevance information is essential as all PCs (and their combinations) may play a role in the discrimination, regardless how much variance in the data they explain (Eq. (2)). To aid in interpretation, the prototypes and relevance matrix can be reconstructed in voxel-space. To this end, the PC transformation (Eq. (1)) that defines the

Table 2

Average training and testing performance of GMLVQ on the differential diagnosis problem (AD, DLB, HC, and PD). Averages and standard deviations are extracted from the the cross-validation procedure.

	Training	Testing
AUC, mean (std)	0.996 (0.002)	0.862 (0.091)
Balanced accuracy (%), mean (std)	93.94 (1.66)	66.44 (14.55)

feature space, can be *approximately* reversed to obtain the prototypes

$$\tilde{\mathbf{w}}_j = \mathbf{G}\mathbf{w}_j, \quad (5)$$

for each included diagnosis $j \in \{HC, AD, PD, DLB\}$ in voxel space. Similarly to Eq. (5), the voxel representation of the diagonal of the relevance matrix and the axes of the discriminative plots, i.e., the discriminative space Section 2.5.1) can be obtained [38].

To aid visualization, each voxel in these 3D images was z-transformed and overlaid on a T1 MRI brain template. Images were arbitrarily thresholded at $|z| > 0.5$ to remove near-zero values and ease interpretation.

2.6. Projection of iRBD patients in GMLVQ space

The normalized FDG-PET scans (at baseline and follow-up) of the 20 iRBD patients were masked, log-transformed and double-demeaned similarly to the evaluation groups’ data. These subjects were not used for training of the GMLVQ model, but were projected into the average relevance space of the GMLVQ systems to visualize their position in this space at baseline and follow-up, and to study each individual’s trajectory; especially of those subjects that converted to PD.

The (Euclidean) distances were calculated within the decision space of GMLVQ. Finally, we evaluated the correlation (Spearman’s rho) between distance that each iRBD individual traveled in the relevance space of the GMLVQ system and the following clinical metrics: duration of RBD symptoms, age, the change in MoCA score per year and the change on the Unified Parkinson’s Disease Rating Scale Part III (UPDRS-III motor examination) per year. A (uncorrected) p -value of < 0.05 was considered statistically significant.

3. Results

The best performing model was constructed using waypoint averaged gradient descent optimization [52]. The three most recent updates were used to compute the tentative average update and the optimization was stopped after 50 epochs. The initial step sizes were set to 0.1 and 0.01 for the prototypes and relevance matrix, respectively. The best performing activation function was soft+ ($\beta = 1$) [53]. Other parameters were left to their default values.¹

The average performance extracted from the ten times repeated ten fold cross-validation procedure is included in Table 2. We observe a difference between the training and testing AUC and balanced accuracy (Table 2).

The decision space of GMLVQ can be represented by three eigenvectors (axes) with the first two the most relevant (fig. A.9). Fig. A.9 also shows that the residual relevance is divided over the remaining features that accounts for slightly less than half the accumulated relevance. The average diagonal of the relevance matrix has been included in the supplementary section (fig. A.8) and shows that all PCs (Section 2.4) are used for the classification.

Fig. 1 shows the average confusion matrix based on the test datasets from the cross-validation procedures. Healthy controls

¹ <https://sklvq.readthedocs.io/en/0.1.2/api.html>.

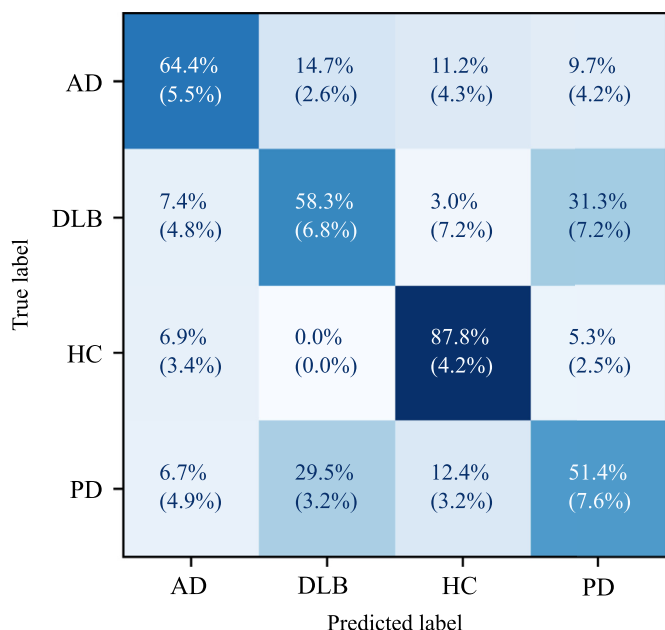


Fig. 1. Average testing confusion matrix of the differential diagnostics problem between (AD, DLB, HC and PD). Averages and standard deviations are extracted from the cross-validation procedure.

were correctly classified in 87.8% of cases, and were misclassified as AD or PD in roughly similar percentages. None of the healthy controls were classified as DLB. AD patients were correctly classified in 64.4% of cases; the remainder of AD were misclassified as DLB, HC and PD in approximately equal proportions. A little over half of the DLB cases were classified as DLB; approximately 30% was classified as PD, and 15% as AD. Performance in PD was low; approximately 50% was correctly classified, 30% as DLB, 10% as AD and 5% as healthy controls.

3.1. Clinical interpretation of patient clusters in GMLVQ decision space

Each participant (Table 1b) is plotted (see Fig. 2) in the GMLVQ space defined by the first eigenvector (y-axis) and second eigenvector (x-axis). To understand why cases were misclassified, we manually identified several clusters in this plot and evaluated the clinical information of these cases.

Cluster 1a consists of PD patients who are close to the healthy control space. These patients were relatively young at the time of scanning (average of 55 years). In fact, three patients had young-onset PD (they were 40, 44 and 51 at the time of scanning and 38, 42 and 47 when they first experienced symptoms). Of the six patients in this cluster, three received a full neuropsychological evaluation which showed only mild cognitive impairments. The other three patients were only evaluated with short screening instruments for cognition and did not appear to have severe cognitive symptoms. Based on their charts and histories taken by the treating neurologist they did not develop severe cognitive complaints at follow-up. Disease duration at the time of scanning was variable, most patients in this cluster had symptoms for less than 4 years at the time of scanning; one patient had motor symptoms for 11 years and was scanned in the context of a work-up for deep-brain stimulation.

Cluster 1b comprises PD patients who are close to the DLB space. Six of these patients had some cognitive symptoms which did not interfere with daily living at the time of the FDG PET scan, but progressed to PD-dementia at follow-up. Two patients had PD-

Table 3

Table includes the age difference between the baseline and follow-up scan of the RBD patients (Table 1c). Additionally, it includes the 2D distance, i.e., the direct observable (euclidean) distance in the progression plot (Fig. 3) as well as the 3D distance, i.e., the distance traveled within the three most important axes for the classification. The true distance is the distance traveled within the full relevance space. The velocity is the true distance divided by the age difference between the scan moments.

Age difference	2D distance	3D distance	True distance	Velocity
3.99	3.17	3.22	3.63	0.91
4.02	4.97	5.00	6.34	1.58
3.91	0.47	0.47	1.04	0.26
3.91	0.69	0.69	1.65	0.42
3.77	0.45	0.51	0.87	0.23
2.57	0.29	0.31	0.59	0.23
2.57	1.28	1.31	2.16	0.84
3.43	1.03	1.12	2.22	0.65
4.04	0.39	0.39	1.12	0.28
4.6	1.78	1.79	2.65	0.58
3.68	0.11	0.11	0.56	0.15
3.68	0.23	0.31	1.19	0.32
4.29	2.02	2.21	2.74	0.64
2.99	3.82	3.84	5.39	1.80
2.99	0.72	0.74	1.55	0.52
3	0.49	0.52	0.73	0.24

(a) The patients who did not develop PD.

Age difference	2D distance	3D distance	True distance	Velocity
3.98	4.03	4.61	5.92	1.49
4.03	2.84	2.85	3.52	0.87
3.92	5.88	5.93	7.09	1.81
4.05	9.38	9.63	11.48	2.83

(b) The patients who developed PD.

dementia at the time of imaging. The average age of patients in this cluster was 75 years.

Cluster 2a includes three AD patients who ended up in DLB space. One patient received the diagnosis AD-PCA. The second patient in this cluster had a more typical amnesic syndrome but also experienced delusions and visual hallucinations. The third patient did not have a specific AD-subtype. Of note, outside of this cluster, another patient with AD-PCA was an outlier (*) in the DLB space. However, the other 4 AD-PCA patients were classified in the middle of the AD space.

Cluster 2b describes six patients, all of whom suffered from MCI at the time of imaging. There is an outlier in AD-space (*), who was scanned in an advanced disease stage (judging by the degree of brain atrophy, especially of the parietal lobes on MRI brain imaging). This patient had cognitive symptoms for 15 years before he was seen in our expertise center at age 53. At that point, his neuropsychology exam showed a severe amnesic syndrome. He had a positive PiB-PET scan. Unfortunately, genetic testing was not performed.

3.2. Projection of iRBD baseline and follow-up data

Fig. 3 shows the projection of the baseline (RBD1) and follow-up data (RBD2) of iRBD patients (Table 1c) in the GMLVQ model. For each patient, the two data points are connected with an arrow. Most patients move from the healthy control space towards the PD and DLB prototypes.

Four patients converted to PD during follow-up, these cases are marked by a circle. Of these four patients, one patient was in the PD discriminative space at baseline, and traveled a distance towards the DLB prototype. The other four patients started in the healthy control space, but all traveled towards the disease prototypes.

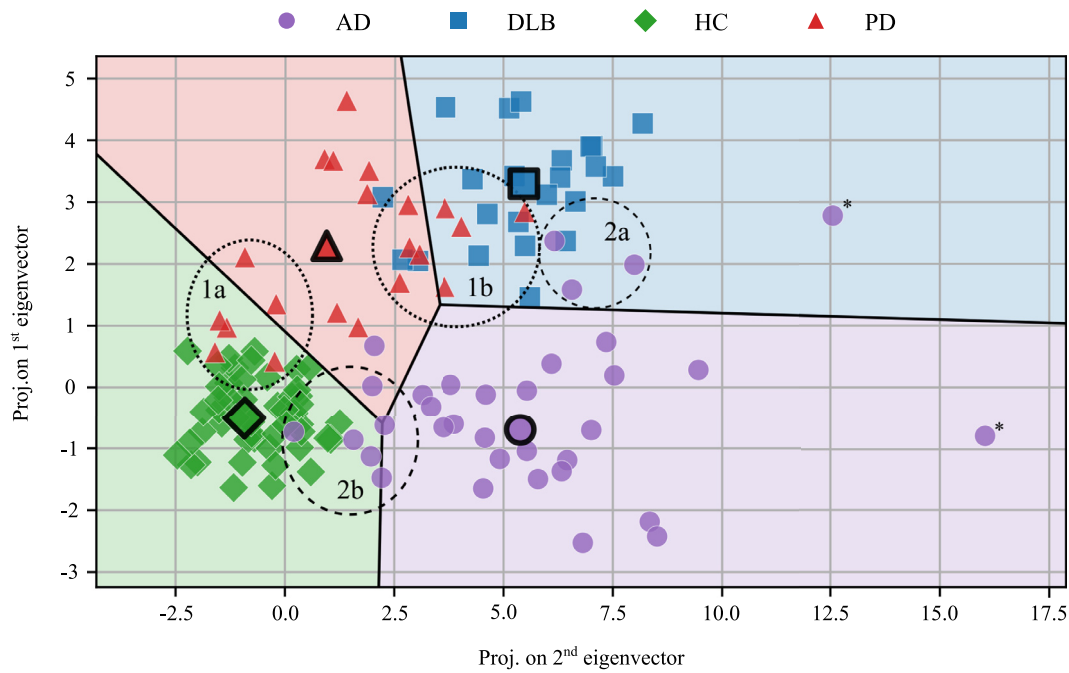


Fig. 2. Projection of the *evaluation* (training) data (z-transform was reversed, to ensure the HCs locate around the origin), labeled according to the diagnosis. The data has been projected on the eigenvectors of the relevance matrix, scaled by the square root of their eigenvalues.

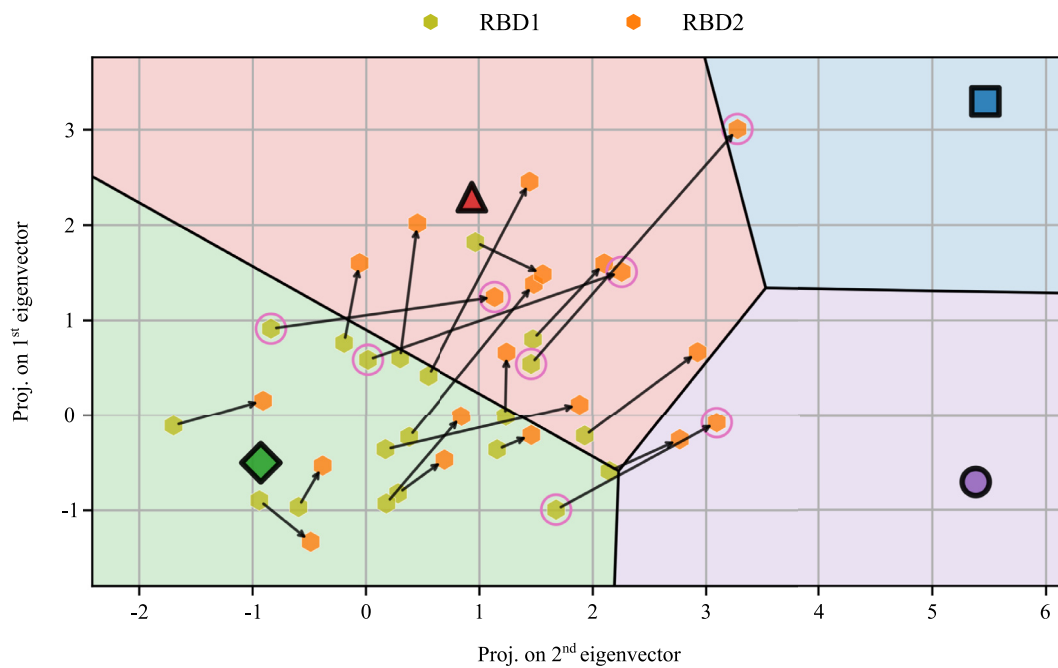


Fig. 3. Projection of the iRBD patients' (progression group, Table 1c) data (z-transform was reversed, to ensure the HCs locate around the origin) within the same space as Fig. 2 but zoomed in. The labeling RBD1 and RBD2 indicate the scans at baseline and follow-up respectively. Scans belonging to the same patient have been connected with an arrow. The four iRBD patients that converted to PD during follow-up are outlined with a pink circle.

Table 3 includes the visible and total distances traveled by the RBD patients in between scans in the GMLVQ relevance space. The differences between the visible and total distance traveled are explained by how much accumulated discriminativeness is captured by the respective number of eigenvectors. In this case, there are three eigenvectors (see Fig. A.9) that, combined, explain most of the discriminative power, and thus going beyond three dimensions does not add substantially to the distance. Additionally,

Table 3 shows that the difference in time between the scans varies between subjects.

The distance traveled by the iRBD subjects through the GMLVQ space per year correlated positively with the change in the UPDRS-III score ($\rho=0.56, p=0.009$). This indicates that patients who develop motor symptoms over time, which is associated with phenoconversion to manifest PD/DLB, traveled further from the healthy control prototype towards the disease discriminative space

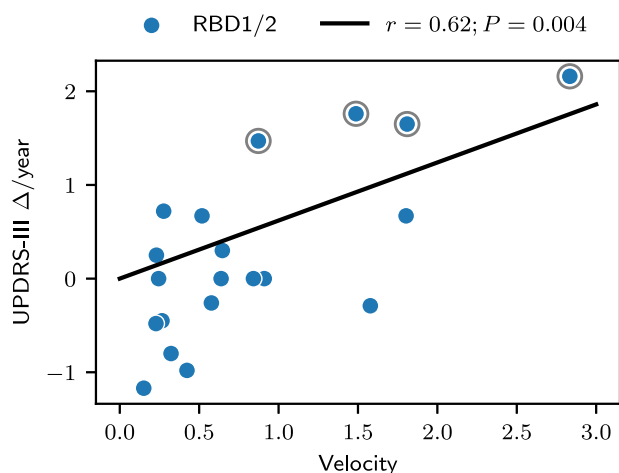


Fig. 4. Scatter plot of the velocity versus UPDRS-III/year of the RBD patients (Table 1c). The line indicates the significant ($P < 0.05$) correlation between the two ($r = 0.62$; $P = 0.004$). The four iRBD patients that converted to PD during follow-up are outlined with a gray circle.

(PD/DLB). There was also a positive correlation between the velocity of each subject in GMLVQ space (distance/yr) and the UPDRS-III change per year ($\rho = 0.62$, $p = 0.004$; Fig. 4). The distances or velocities were not significantly correlated with the change in MoCA score, duration of RBD symptoms, or age.

3.3. Voxel representations

In Fig. 2 it can be appreciated that patients with PD/DLB have a coordinate in GMLVQ space with a positive value on the y -axis (the first eigenvector), whereas most AD patients have a negative value. Thus, the first eigenvector determines, to some extent, the difference between AD and PD/DLB. The voxel representation of the first eigenvector is shown in Fig. 5a. This pattern is characterized by negative voxel weights in the thalamus and putamen, occipital cortex, parietal cortex, lateral frontal cortex and anterior cingulate. Positive regions include the temporal lobes, the posterior cingulate and the sensorimotor cortex.

The distinction between groups based on the x -axis (the second eigenvector) relies on the difference between controls (low or negative x -value) and patients. Especially patients with dementia have a higher value for x . The voxel representation for the second eigenvector is shown in Fig. 5b. This pattern is characterized by negative voxel weights in the temporal cortex, the parietal cortex and lateral frontal cortex. Regions with increased voxel weights include the cerebellum, thalamus, putamen, brainstem, orbitofrontal cortex and (to lesser extent) the sensorimotor cortex. The voxel representation of the third eigenvector (z -axis) can be found in the supplementary material (Fig. A.10).

The voxel representation of the relevance diagonal is shown in Fig. 6. The relevance diagonal provides a summary of all relevant features (PCs) for the classification problem and consists of only positive values. In contrast to the interpretation of the prototypes the voxel map of the relevance diagonal does not indicate where difference between the diagnoses can be found but indicates which differences, which can be small or large, are important or 'relevant' for the classification.

The voxel representation for each prototype is presented in Fig. 7. Positive and negative values are with respect to the mean of the reference healthy controls. The prototypes of PD, DLB and AD all show relatively increased metabolism in the cerebellum and brainstem, the thalamus, putamen, globus pallidus, insula, orbitofrontal cortex and sensorimotor cortex. Relatively decreased

metabolism was seen in all three prototypes in the parietal cortex, lateral frontal cortex and caudate nucleus. In PD/DLB the occipital cortex was also involved, whereas in AD the hypometabolism was more extensive in the temporal cortex. The healthy control prototype is difficult to interpret but is, to some degree, the inverse of the disease prototypes.

4. Discussion

In this study, we show that FDG-PET analysis with principal component analysis and GMLVQ can be used to discriminate between PD, DLB and AD. More importantly, we show that most iRBD patients who were scanned twice with FDG-PET (approximately 3.7 years apart), traveled through the GMLVQ space, away from the healthy control receptive field, and towards the disease prototypes. Four patients with iRBD developed motor symptoms (parkinsonism) during follow-up and were diagnosed with PD. Each of these four cases indeed escaped the healthy-control receptive field. In support of this, severity of motor symptoms (as measured with the UPDRS-III) was positively correlated with the distance traveled through the GMLVQ space. Equally important, six iRBD subjects remained within the healthy control receptive field during follow-up, which may indicate that these subjects are not prone for phenotypic conversion in the short term. Some iRBD subjects traveled only a short distance, which may indicate only subtle metabolic changes and relatively slow disease progression. The distance traveled did not correlate significantly with age, indicating that the observed changes did not simply signify age difference.

The main merit of GMLVQ is that the coordinate system of the model, as well as each prototype, can be reconstructed in voxel space. This allows inspection of the regional brain changes that contributed to the model and its classifications of patients. As could be expected, the patterns of relative hypo- and hypermetabolism of each prototype are consistent with previously identified SSM/PCA disease-related patterns and also with univariate studies [15,22–26]. This means that the brain metabolic changes that underlie the PD, DLB and AD conditions are consistently identified, independent of the analytical method that is used. These prototypes showed considerable regional overlap, just like the disease-related patterns.

The axes of the GMLVQ space determine the position of the training data (and the prototypes). These axes can also be reconstructed in voxel space to understand the classifications. The y -axis (eigenvector 1) determined the separation of Lewy body disorders (PD/DLB) and AD. Typical features of this eigenvector included hypometabolism of the occipital cortex and the frontal cortex. PD and DLB patients had a high value on the y -axis, which indicates that they matched this pattern, whereas AD patients had a negative value (indicating a mismatch with the described pattern). This is in line with the literature: Hypometabolism of the occipital cortex is a well-known specific (but less sensitive) signature of DLB [54,55]. With progression of disease and the development of cognitive impairment, PD patients can also develop occipital hypometabolism [56]. ^{18}F -FDG PET studies in autopsy-confirmed DLB and AD patients showed comparable degrees of metabolic reductions in the parietotemporal and frontal association cortices, but only DLB patients showed severe metabolic reductions in the occipital cortex [57,58].

The x -axis (eigenvector 2) determined the separation of controls versus disease. The voxel representation of eigenvector 2 included regions that the prototypes of all conditions (PD, DLB and AD) had in common: Relative hypermetabolism of the cerebellum, brainstem, orbitofrontal cortex, basal ganglia, thalamus, and sensorimotor cortex, and temporo-parietal hypometabolism. Note that the occipital cortex was not implicated in this pattern.

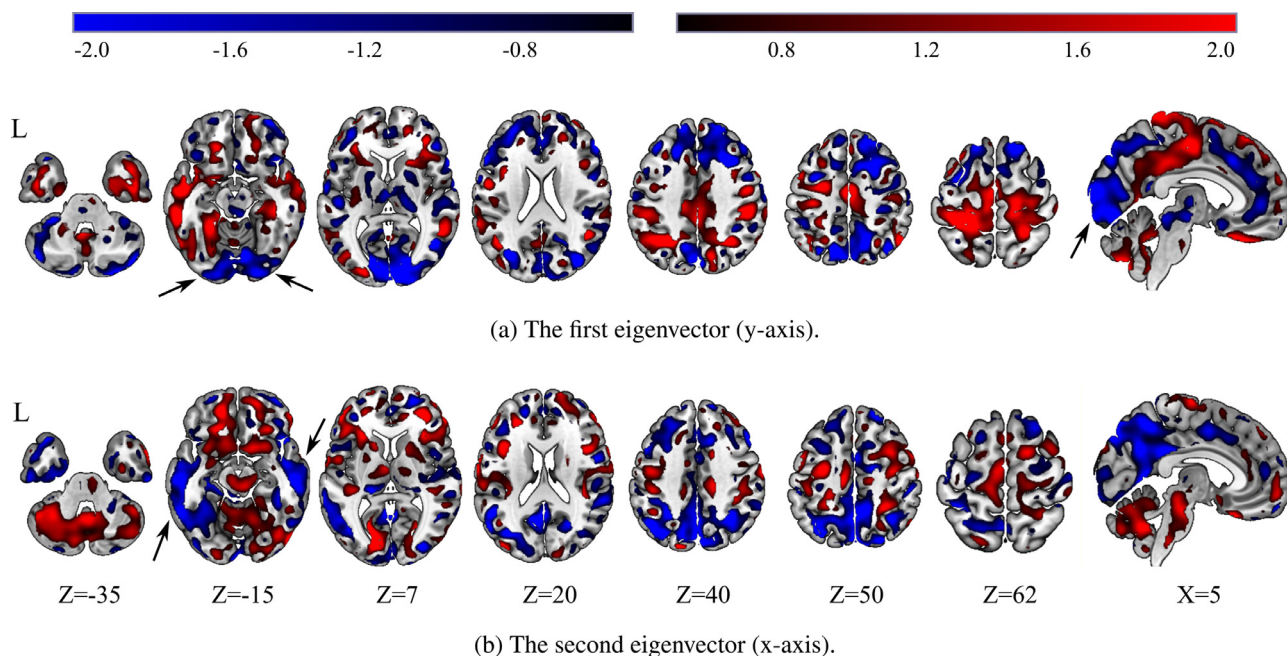


Fig. 5. This figures shows the voxel representation (patterns) of the eigenvectors defining the space as illustrated in Fig. 2. The values indicate activation above or below the average healthy control from the *reference* group. To aid visualization, each voxel in these 3D images was z-transformed and overlaid on a T1 MRI brain template. Images were arbitrarily thresholded at $|z > 0.5|$ to remove near-zero values.

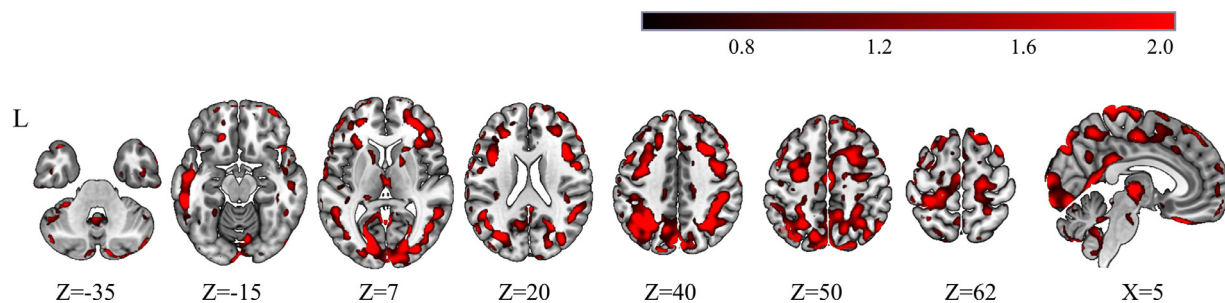


Fig. 6. Voxel representation (pattern) of the diagonal of the relevance matrix. The values indicate how relevant the voxels are for the classification task. To aid visualization, each voxel in these 3D images was z-transformed and overlaid on a T1 MRI brain template. Images were arbitrarily thresholded at $|z > 0.5|$ to remove near-zero values.

Following from the above, we can understand why some subjects were misclassified, and we can understand why some cases are outliers. The GMLVQ system allows further inspection of difficult cases. This could potentially aid clinical practice. For instance, if a patient is close to a prototype that is in agreement with the clinical diagnosis, this confirms the clinical diagnosis. If a case is in between two or more prototypes and closer to the decision boundary the diagnosis is uncertain, and further investigation or follow-up is necessary. Finally, if a case lies close to a prototype which has a different label than the current clinical diagnosis, there is a mismatch and both clinical information and the imaging data should be revisited.

The PD subjects who were in the healthy control receptive field were relatively young and also did not appear to develop cognitive impairment upon clinical follow-up after scanning. Thus, they can be expected to have limited degrees of cortical hypometabolism [59]. In line with this, PD patients who were in, or close to, the DLB receptive field were demented or developed dementia upon clinical follow-up. The metabolic patterns for PD-dementia and DLB are known to be very similar [56]. Patients with AD-PCA have atrophy (and thus also hypometabolism) of the posterior brain regions, which is similar to DLB. Two of the six AD-PCA subjects in our study indeed ended up in the DLB receptive field. Finally, six patients with AD-MCI were in, or close to, the

healthy control receptive field. Patients with MCI have only mild cognitive complaints and typically have limited degrees of cortical hypometabolism. Just like iRBD patients, they represent a very early stage of neurodegeneration. It is therefore not surprising that some AD-MCI patients are still considered normal by the GMLVQ system. Multiple measurements may also be necessary for MCI to determine the distance traveled through the model as a proxy for disease progression.

When considering the confusion matrix, the diagnostic performance of the GMLVQ model seems poor. However, the performance needs to be evaluated in the context of the clinical diagnoses that were included. The distinction between healthy controls and patients was good. Patients that were misclassified as controls included 6.94% of AD subjects, which was solely due to the MCI subjects, and 5.31% of PD patients. The misclassification of PD subjects could be due to levodopa therapy. It has been shown that levodopa causes a normalization of abnormal hyperactivity in the posterior putamen, globus pallidus, ventral thalamus and dorsal pons [60]. In our study, levodopa therapy was not routinely withheld before scanning. This could have influenced results (details are provided in the supplemental material).

The model was also able to provide a relatively good distinction between DLB and AD. 14.7% of AD cases were misclassified as DLB, which may have included the AD-PCA cases. In addition, only 7.4%

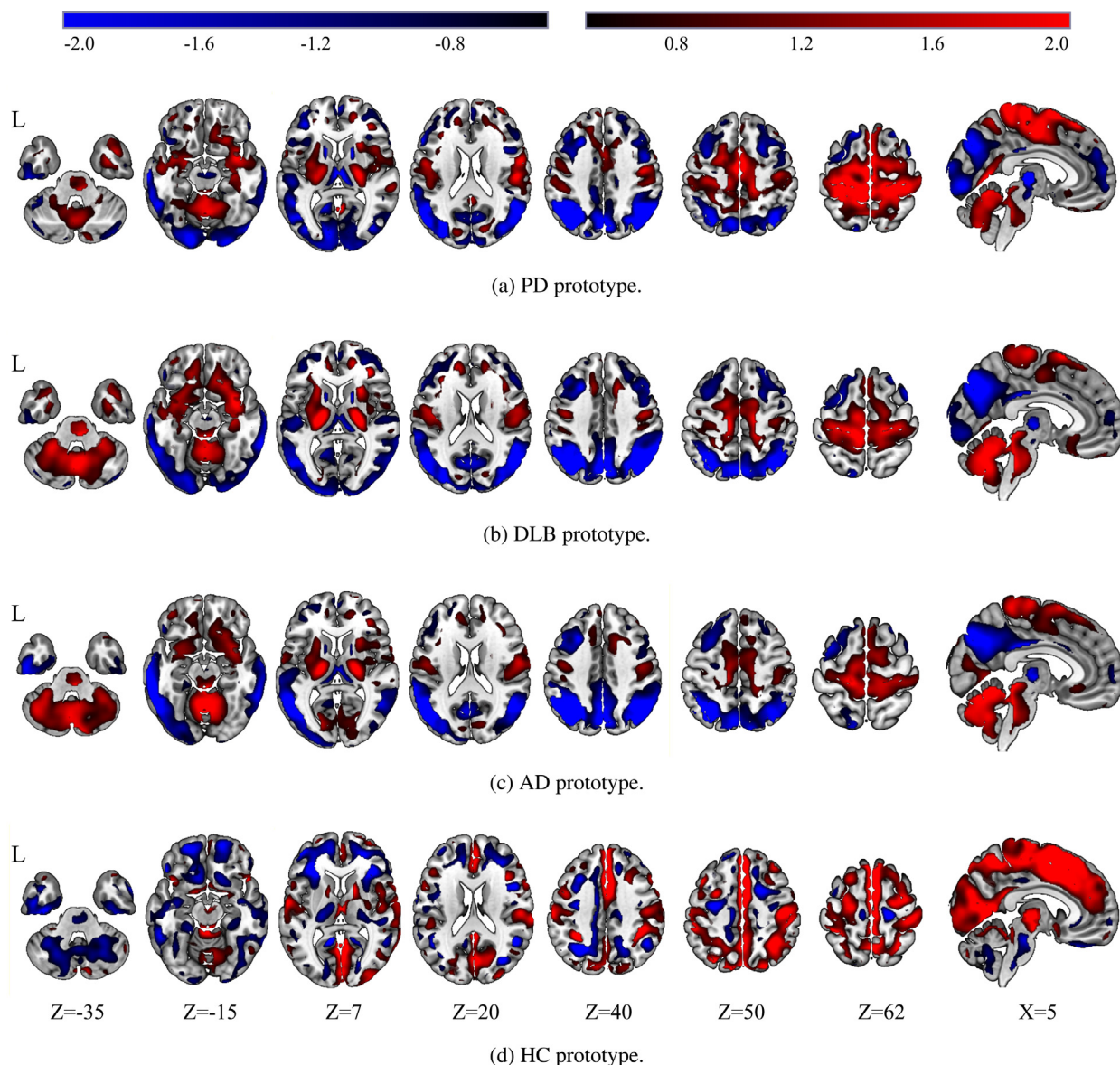


Fig. 7. This figure shows the voxel representations (patterns) of the average prototypes, as extracted from the ten times repeated ten-fold cross-validation. The values indicate activation above or below the average healthy control from the *reference* group. To aid visualization, each voxel in these 3D images was z-transformed and overlaid on a T1 MRI brain template. Images were arbitrarily thresholded at $|z| > 0.5$ to remove near-zero values.

of DLB cases were confused as AD. A perfect separation between AD, DLB and PD will be impossible based on the pathophysiology of these conditions. It is estimated that over half of patients with a clinical DLB diagnosis [61] and approximately one third of all PD cases have concurrent AD pathology. This is more pronounced in PD patients with dementia but is also found in cognitively intact PD patients [62]. Finally, PD and DLB were most often confused but this is not surprising as these conditions are part of the same spectrum. However, it would be useful to repeat this analysis in a larger cohort of PD, AD and DLB patients where the diagnostic process was consistently enriched with other biomarkers (specifically PiB PET and CSF) to further disentangle cases with a pure alpha-synucleinopathy from those with overlapping pathologies.

The difference between the training and testing performance indicates that the model overfits the training data. This can be explained by the amount of the diagnosis-dependent inter-patient variance in these data. These inequalities in patients can be observed in the discriminative plots in Fig. 2 as well. Overfitting for our specific use case might not be a big issue, as the projections

of the GMLVQ systems will produce more distinct visualizations of the data, therefore, providing a clearer image of potential difficult cases. For the interpretation of these projections, though, one must also consider the predictive performance of these systems.

5. Conclusion

We wish to emphasize that our study is a proof of concept, indicating that GMLVQ can be a useful tool in the analysis of FDG PET data in multi-class problems. So far, our study does not allow any firm conclusions concerning the diagnostic value in clinical practice or prediction of phenoconversion in iRBD. However, we have illustrated the potential of this method and its merit over the classical SSM-PCA patterns and other ‘black box’ machine-learning algorithms. In future studies, a GMLVQ model trained with larger numbers of HC, PD, DLB and MSA will be more intuitive to evaluate the disease progression and differential diagnosis in iRBD patients, as iRBD patients do not develop AD. That said, the AD-PD space did not hamper application to iRBD, possibly because the PD- and

AD-related patterns overlap. A GMLVQ system may also be considered as an aid in the differential diagnosis of patients presenting with parkinsonism (motor symptoms), i.e. PD, MSA, DLB, progressive supranuclear palsy, and corticobasal degeneration. Similarly, a system for patients presenting with cognitive impairment could be constructed, including HC, AD, DLB and Frontotemporal dementia. The latter approach could be useful to predict the diagnosis in patients presenting with mild cognitive impairment (MCI). In patients with iRBD and MCI, future studies should investigate the relationship between distance traveled (and velocity) in GMLVQ space and disease progression.

Ethical approval

The data used in this study has been approved by the ethical review board of the UMCG.

Funding

This study was supported by Dutch “StichtingParkinsonFonds”, the German “ParkinsonFonds Deutschland”, and partly funded by the Michael J. Fox Foundation (ID 17081), BMK, BMDW, and the State of Upper Austria in the frame of SCCH, part of the COMET Program managed by FFG.

Declaration of Competing Interest

The authors declare no competing or financial interest in this work.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.cmpb.2022.107042](https://doi.org/10.1016/j.cmpb.2022.107042).

References

- [1] K. Gammon, Neurodegenerative disease: brain windfall, *Nature* 515 (7526) (2014) 299–300, doi:[10.1038/nj7526-299a](https://doi.org/10.1038/nj7526-299a).
- [2] G. Rizzo, M. Copetti, S. Arcuti, D. Martino, A. Fontana, G. Logroscino, Accuracy of clinical diagnosis of parkinson disease: a systematic review and meta-analysis, *Neurology* 86 (6) (2016) 566–576.
- [3] K.A. Jellinger, G. Logroscino, G. Rizzo, M. Copetti, S. Arcuti, D. Martino, A. Fontana, Accuracy of clinical diagnosis of parkinson disease: a systematic review and meta-analysis, *Neurology* 87 (2) (2016) 237–238.
- [4] B.T. Hyman, C.H. Phelps, T.G. Beach, E.H. Bigio, N.J. Cairns, M.C. Carrillo, D.W. Dickson, C. Duyckaerts, M.P. Frosch, E. Masliah, et al., National institute on aging–alzheimer’s association guidelines for the neuropathologic assessment of Alzheimer’s disease, *Alzheimer’s Dementia* 8 (1) (2012) 1–13.
- [5] W.R. Galpern, A.E. Lang, Interface between tauopathies and synucleinopathies: a tale of two proteins, *Ann. Neurol.* 59 (3) (2006) 449–458, doi:[10.1002/ana.20819](https://doi.org/10.1002/ana.20819).
- [6] J.G. Goldman, C.G. Goetz, M. Brandabur, M. Sanfilippo, G.T. Stebbins, Effects of dopaminergic medications on psychosis and motor function in dementia with Lewy bodies, *Mov. Disord.* 23 (15) (2008) 2248–2250, doi:[10.1002/mds.22322](https://doi.org/10.1002/mds.22322).
- [7] S. Heinzel, D. Berg, T. Gasser, H. Chen, C. Yao, R.B. Postuma, Update of the MDS research criteria for prodromal Parkinson’s disease, *Mov. Disord.* 34 (10) (2019) 1464–1470, doi:[10.1002/mds.27802](https://doi.org/10.1002/mds.27802).
- [8] C.H. Schenck, B.F. Boeve, M.W. Mahowald, Delayed emergence of a parkinsonian disorder or dementia in 81% of older men initially diagnosed with idiopathic rapid eye movement sleep behavior disorder: a 16-year update on a previously reported series, *Sleep Med.* 14 (8) (2013) 744–748, doi:[10.1016/j.sleep.2012.10.009](https://doi.org/10.1016/j.sleep.2012.10.009).
- [9] R.B. Postuma, A. Iranzo, M. Hu, B. Högl, B.F. Boeve, R. Manni, W.H. Oertel, I. Arnulf, L. Ferini-Strambi, M. Puligheddu, E. Antelmi, V.C.D. Cock, D. Arnaldi, B. Mollenhauer, A. Videnovic, K. Sonka, K.-Y. Jung, D. Kunz, Y. Dauvilliers, F. Provini, S.J. Lewis, J. Buskova, M. Pavlova, A. Heidebreder, J.Y. Montplaisir, J. Santamaria, T.R. Barber, A. Stefani, E.K.S. Louis, M. Terzaghi, A. Janzen, S. Leu-Semenescu, G. Plazzi, F. Nobili, F. Sixel-Doering, P. Dusek, F. Bes, P. Cortelli, K.E. Martens, J.-F. Gagnon, C. Gaig, M. Zucconi, C. Trenkwalder, Z. Gan-Or, C. Lo, M. Rolinski, P. Mahlknecht, E. Holzkecht, A.R. Boeve, L.N. Teigen, G. Toscano, G. Mayer, S. Morbelli, B. Dawson, A. Pelletier, Risk and predictors of dementia and parkinsonism in idiopathic REM sleep behaviour disorder: a multicentre study, *Brain* 142 (3) (2019) 744–759, doi:[10.1093/brain/awz030](https://doi.org/10.1093/brain/awz030).
- [10] M. Reivich, D. Kuhl, A. Wolf, J. Greenberg, M.A. Phelps, T. Ido, V. Casella, J.V. Fowler, E. Hoffman, A. Alavi, et al., The [18F] fluorodeoxyglucose method for the measurement of local cerebral glucose utilization in man, *Circ. Res.* 44 (1) (1979) 127–137.
- [11] V. Garibotto, K. Herholz, M. Boccardi, A. Picco, A. Varrone, A. Nordberg, F. Nobili, O. Ratib, Geneva Task Force for the Roadmap of Alzheimer’s Biomarkers, Clinical validity of brain fluorodeoxyglucose positron emission tomography as a biomarker for Alzheimer’s disease in the context of a structured 5-phase development framework, *Neurobiol. Aging* 52 (2017) 183–195.
- [12] Z. Walker, F. Gandolfo, S. Orini, V. Garibotto, F. Agosta, J. Arbizu, F. Bouwman, A. Drzezga, P. Nestor, M. Boccardi, et al., Clinical utility of FDG PET in Parkinson’s disease and atypical parkinsonism associated with dementia, *Eur. J. Nucl. Med. Mol. Imaging* 45 (9) (2018) 1534–1545.
- [13] D. Eidelberg, Metabolic brain networks in neurodegenerative disorders: a functional imaging approach, *Trends Neurosci.* 32 (10) (2009) 548–557.
- [14] S.K. Meles, J.G. Kok, R.J. Renken, K.L. Leenders, From positron to pattern: a conceptual and practical overview of 18F-FDG PET imaging and spatial covariance analysis, in: *PET and SPECT in Neurology*, Springer International Publishing, 2020, pp. 73–104, doi:[10.1007/978-3-030-53168-3_4](https://doi.org/10.1007/978-3-030-53168-3_4).
- [15] K.A. Schindlbeck, D. Eidelberg, Network imaging biomarkers: insights and clinical applications in Parkinson’s disease, *Lancet Neurol.* 17 (7) (2018) 629–640.
- [16] S.K. Meles, D. Vadasz, R.J. Renken, E. Sittig-Wiegand, G. Mayer, C. Depboylu, K. Reetz, S. Overeem, A. Pijpers, F.E. Reesink, et al., FDG PET, dopamine transporter SPECT, and olfaction: combining biomarkers in REM sleep behavior disorder, *Mov. Disord.* 32 (10) (2017) 1482–1486.
- [17] R.V. Kogan, A. Janzen, S.K. Meles, E. Sittig, R.J. Renken, V. Gurvits, G. Mayer, K.L. Leenders, W.H. Oertel, R.W. Group, et al., Four-year follow-up of [18F] fluorodeoxyglucose positron emission tomography-based Parkinson’s disease-related pattern expression in 20 patients with isolated rapid eye movement sleep behavior disorder shows prodromal progression, *Mov. Disord.* 36 (1) (2020) 230–235.
- [18] F. Holtbernd, J.-F. Gagnon, R.B. Postuma, Y. Ma, C.C. Tang, A. Feigin, V. Dhawan, M. Vendette, J.-P. Soucy, D. Eidelberg, J. Montplaisir, Abnormal metabolic network activity in REM sleep behavior disorder, *Neurology* 82 (7) (2014) 620–627, doi:[10.1212/wnl.0000000000000130](https://doi.org/10.1212/wnl.0000000000000130).
- [19] E.J. Yoon, J.-Y. Lee, H. Nam, H.-J. Kim, B. Jeon, J.M. Jeong, Y.K. Kim, A new metabolic network correlated with olfactory and executive dysfunctions in idiopathic rapid eye movement sleep behavior disorder, *J. Clin. Neurol.* 15 (2) (2019) 175, doi:[10.3988/jcn.2019.15.2.175](https://doi.org/10.3988/jcn.2019.15.2.175).
- [20] P. Wu, H. Yu, S. Peng, Y. Dauvilliers, J. Wang, J. Ge, H. Zhang, D. Eidelberg, Y. Ma, C. Zuo, Consistent abnormalities in metabolic network activity in idiopathic rapid eye movement sleep behaviour disorder, *Brain* 137 (12) (2014) 3122–3128, doi:[10.1093/brain/awu290](https://doi.org/10.1093/brain/awu290).
- [21] J.H. Shin, J.-Y. Lee, Y.-K. Kim, E.J. Yoon, H. Kim, H. Nam, B. Jeon, Parkinson disease-related brain metabolic patterns and neurodegeneration in isolated REM sleep behavior disorder, *Neurology* 97 (4) (2021) E378–E388, doi:[10.1212/wnl.0000000000012228](https://doi.org/10.1212/wnl.0000000000012228).
- [22] C. Habeck, N.L. Foster, R. Perneczky, A. Kurz, P. Alexopoulos, R.A. Koeppe, A. Drzezga, Y. Stern, Multivariate and univariate neuroimaging biomarkers of Alzheimer’s disease, *Neuroimage* 40 (4) (2008) 1503–1515.
- [23] P.J. Mattis, M. Niethammer, W. Sako, C.C. Tang, A. Nazem, M.L. Gordon, V. Brandt, V. Dhawan, D. Eidelberg, Distinct brain networks underlie cognitive dysfunction in parkinson and alzheimer diseases, *Neurology* 87 (18) (2016) 1925–1933.
- [24] L.K. Teune, R.J. Renken, B.M. de Jong, A.T. Willemsen, M.J. van Osch, J.B. Roerdink, R.A. Dierckx, K.L. Leenders, Parkinson’s disease-related perfusion and glucose metabolic brain patterns identified with PCASL-MRI and FDG-PET imaging, *Neuroimage* 5 (2014) 240–244.
- [25] S.K. Meles, M. Pagani, D. Arnaldi, F. De Carli, B. Dessi, S. Morbelli, G. Sambucetti, C. Jonsson, K.L. Leenders, F. Nobili, The Alzheimer’s disease metabolic brain pattern in mild cognitive impairment, *J. Cereb. Blood Flow Metab.* 37 (12) (2017) 3643–3648.
- [26] T. Iizuka, M. Kameyama, Spatial metabolic profiles to discriminate dementia with Lewy bodies from Alzheimer disease, *J. Neurol.* 267 (7) (2020) 1960–1969, doi:[10.1007/s00415-020-09790-8](https://doi.org/10.1007/s00415-020-09790-8).
- [27] S. Peng, P.G. Spetsieris, D. Eidelberg, Y. Ma, Radiomics and supervised machine learning in the diagnosis of parkinsonism with FDG PET: promises and challenges, *Ann. Transl. Med.* 8 (13) (2020), doi:[10.21037/atm.2020.04.33](https://doi.org/10.21037/atm.2020.04.33). 808–808
- [28] S. Rathore, M. Habes, M.A. Iftikhar, A. Shacklett, C. Davatzikos, A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer’s disease and its prodromal stages, *Neuroimage* 155 (2017) 530–548, doi:[10.1016/j.neuroimage.2017.03.057](https://doi.org/10.1016/j.neuroimage.2017.03.057).
- [29] Z. Lipton, The mythos of model interpretability, *Queue* 16 (3) (2018) 30:31–30:57, doi:[10.1145/3236386.3241340](https://doi.org/10.1145/3236386.3241340).
- [30] Z. Lipton, The doctor just won’t accept that!, *arXiv preprint arXiv:1711.08037* (2017).
- [31] D. Mudali, L. Teune, R. Renken, K. Leenders, J. Roerdink, Classification of parkinsonian syndromes from FDG-PET brain data using decision trees with SSM/PCA features, *Comput. Math. Methods Med.* 2015 (2015) 10, doi:[10.1155/2015/136921](https://doi.org/10.1155/2015/136921).
- [32] D. Williams, D. Mudali, H. Buddelmeijer, P. Noorishad, S. Meles, R. Renken, N. Leenders, E. Valentijn, J. Roerdink, Visualization of decision tree state for the classification of Parkinson’s disease, *J. Biomed. Eng. Med. Imaging* 3 (3) (2016), doi:[10.14738/jbemi.33.1858](https://doi.org/10.14738/jbemi.33.1858).
- [33] T. Kohonen, The self-organizing map, *Proc. IEEE* 78 (9) (1990) 1464–1480, doi:[10.1109/5.58325](https://doi.org/10.1109/5.58325).

- [34] P. Schneider, M. Biehl, B. Hammer, Adaptive relevance matrices in learning vector quantization, *Neural Comput.* 21 (12) (2009) 3532–3561, doi:10.1162/neco.2009.11-08-908.
- [35] M. Biehl, P. Schneider, D.J. Smith, H. Stiekema, A.E. Taylor, B.A. Hughes, C.H.L. Schackleton, P.M. Stewart, W. Arlt, Matrix relevance LVQ in steroid metabolomics based classification of adrenal tumors, in: *20th European Symposium on Artificial Neural Networks (ESANN)*, 2012, pp. 423–428.
- [36] K. Bunte, P. Schneider, B. Hammer, F.M. Schleif, T. Villmann, M. Biehl, Limited rank matrix learning, discriminative dimension reduction and visualization, *Neural Netw.* 26 (2012) 159–173, doi:10.1016/j.neunet.2011.10.001.
- [37] M. Biehl, K. Bunte, F.M. Schleif, P. Schneider, T. Villmann, Large margin linear discriminative visualization by matrix relevance learning, in: *The 2012 International Joint Conference on Neural Networks (IJCNN)*, 2012, pp. 1–8, doi:10.1109/IJCNN.2012.6252627.
- [38] R. van Veen, V. Gurvits, R. Kogan, S. Meles, J. de Vries, R. Renken, M. Rodriguez-Oroz, R. Rodriguez-Rojas, D. Arnaldi, S. Raffà, B. de Jong, K. Leenders, M. Biehl, An application of generalized matrix learning vector quantization in neuroimaging, *Comput. Methods Programs Biomed.* 197 (2020) 105708, doi:10.1016/j.cmpb.2020.105708.
- [39] S.K. Meles, R.J. Renken, M. Pagani, L.K. Teune, D. Arnaldi, S. Morbelli, F. Nobili, T. van Laar, J.A. Obeso, M.C. Rodríguez-Oroz, K.L. Leenders, Abnormal pattern of brain glucose metabolism in Parkinson's disease: replication in three european cohorts, *Eur. J. Nucl. Med. Mol. Imaging* 47 (2) (2019) 437–450, doi:10.1007/s00259-019-04570-7.
- [40] S.K. Meles, J.G. Kok, B.M. De Jong, R.J. Renken, J.J. de Vries, J.M. Spikman, A.L. Ziengs, A.T. Willemsen, H.J. van der Horn, K.L. Leenders, et al., The cerebral metabolic topography of spinocerebellar ataxia type 3, *Neuroimage* 19 (2018) 90–97.
- [41] D.E. Peretti, R.J. Renken, F.E. Reesink, B.M. de Jong, P.P. De Deyn, R.A. Dierckx, J. Doorduyn, R. Boellaard, D.V. García, Feasibility of pharmacokinetic parametric pet images in scaled subprofile modelling using principal component analysis, *Neuroimage* 30 (2021) 102625.
- [42] R.B. Postuma, D. Berg, M. Stern, W. Poewe, C.W. Olanow, W. Oertel, J. Obeso, K. Marek, I. Litvan, A.E. Lang, et al., MDS clinical diagnostic criteria for Parkinson's disease, *Mov. Disord.* 30 (12) (2015) 1591–1601.
- [43] I.G. McKeith, B.F. Boeve, D.W. Dickson, G. Halliday, J.-P. Taylor, D. Weintraub, D. Aarsland, J. Galvin, J. Attems, C.G. Ballard, et al., Diagnosis and management of dementia with Lewy bodies: fourth consensus report of the DLB consortium, *Neurology* 89 (1) (2017) 88–100.
- [44] C.R. Jack Jr., D.A. Bennett, K. Blennow, M.C. Carrillo, B. Dunn, S.B. Haeblerlein, D.M. Holtzman, W. Jagust, F. Jessen, J. Karlawish, et al., NIA-AA research framework: toward a biological definition of Alzheimer's disease, *Alzheimer's Dementia* 14 (4) (2018) 535–562.
- [45] P.A.D. Rosa, C. Cerami, F. Gallivanone, A. Prestia, A. Caroli, I. Castiglioni, M.C. Giardi, G. Frisoni, K. Friston, J. Ashburner, et al., A standardized [18 F]-fdg-pet template for spatial normalization in statistical parametric mapping of dementia, *Neuroinformatics* 12 (4) (2014) 575–593.
- [46] R. van Veen, M. Biehl, J. de Vries, sklvq: Scikit learning vector quantization, *J. Mach. Learn. Res.* 22 (231) (2021) 1–6.
- [47] L. Mosley, A balanced approach to the multi-class imbalance problem, Ph.D. thesis, 10.31274/etd-180810-3375
- [48] J. Kelleher, B.M. Namee, A. D'arcy, *Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies*, MIT press, 2020.
- [49] D. Hand, Measuring classifier performance: a coherent alternative to the area under the ROC curve, *Mach. Learn.* 77 (2009) 103–123, doi:10.1007/s10994-009-5119-5.
- [50] M. Biehl, B. Hammer, T. Villmann, Prototype-based models in machine learning, *Wiley Interdiscip. Rev. Cognit. Sci.* 7 (2) (2016) 92–111, doi:10.1002/wics.1378.
- [51] E. Elpaydin, *Introduction to Machine Learning*, MIT press, 2020.
- [52] G. Papari, K. Bunte, M. Biehl, *Waypoint Averaging and Step Size Control in Learning by Gradient Descent*, Tech. Rep. Mlr-2011-06, Leipzig University, 2011.
- [53] T. Villmann, J. Ravichandran, A. Villmann, D. Nebel, M. Kaden, Investigation of activation functions for generalized learning vector quantization, in: A. Vellido, K. Gibert, C. Angulo, J.D.M. Guerrero (Eds.), *Advances in Self-Organizing Maps, Learning Vector Quantization, Clustering and Data Visualization*, Springer International Publishing, Cham, 2020, pp. 179–188.
- [54] M. Trošt, M. Perovnik, Z. Pirtošek, Correlations of neuropsychological and metabolic brain changes in Parkinson's disease and other α -synucleinopathies, *Front. Neurol.* (2019) 10, doi:10.3389/fneur.2019.01204.
- [55] T. Yousaf, G. Dervenoulas, P.-E. Valkimadi, M. Politis, Neuroimaging in Lewy body dementia, *J. Neurol.* 266 (1) (2018) 1–26, doi:10.1007/s00415-018-8892-x.
- [56] D. Garcia-Garcia, P. Clavero, C.G. Salas, I. Lamet, J. Arbizu, R. Gonzalez-Redondo, J.A. Obeso, M.C. Rodriguez-Oroz, Posterior parietooccipital hypometabolism may differentiate mild cognitive impairment from dementia in Parkinson's disease, *Eur. J. Nucl. Med. Mol. Imaging* 39 (11) (2012) 1767–1777, doi:10.1007/s00259-012-2198-5.
- [57] S. Minoshima, N.L. Foster, A.A.F. Sima, K.A. Frey, R.L. Albin, D.E. Kuhl, Alzheimer's disease versus dementia with Lewy bodies: cerebral metabolic distinction with autopsy confirmation, *Ann. Neurol.* 50 (3) (2001) 358–365, doi:10.1002/ana.1133.
- [58] M. Higuchi, M. Tashiro, H. Arai, N. Okamura, S. Hara, S. Higuchi, M. Itoh, R.-W. Shin, J.Q. Trojanowski, H. Sasaki, Glucose hypometabolism and neuropathological correlates in brains of dementia with Lewy bodies, *Exp. Neurol.* 162 (2) (2000) 247–256, doi:10.1006/exnr.2000.7342.
- [59] A. Pilotto, E. Premi, S.P. Caminiti, L. Presotto, R. Turrone, A. Alberici, B. Paghera, B. Borroni, A. Padovani, D. Perani, Single-subject SPM FDG-PET patterns predict risk of dementia progression in Parkinson disease, *Neurology* 90 (12) (2018) E1029–e1037, doi:10.1212/wnl.0000000000005161.
- [60] K. Asanuma, Network modulation in the treatment of Parkinson's disease, *Brain* 129 (10) (2006) 2667–2678, doi:10.1093/brain/awl162.
- [61] R. Ossenkoppele, W.J. Jansen, G.D. Rabinovici, D.L. Knol, W.M. van der Flier, B.N.M. van Berckel, P. Scheltens, P.J. Visser, S.C.J. Verfaillie, M.D. Zwan, S.M. Adriaanse, A.A. Lammertsma, F. Barkhof, W.J. Jagust, B.L. Miller, H.J. Rosen, S.M. Landau, V.L. Villemagne, C.C. Rowe, D.Y. Lee, D.L. Na, S.W. Seo, M. Sarazin, C.M. Roe, O. Sabri, H. Barthel, N. Koglin, J. Hodges, C.E. Leyton, R. Vandenberghe, K. van Laere, A. Drzezga, S. Forster, T. Grimmer, P. Sánchez-Juan, J.M. Carril, V. Mok, V. Camus, W.E. Klunk, A.D. Cohen, P.T. Meyer, S. Hellwig, A. Newberg, K.S. Frederiksen, A.S. Fleisher, M.A. Mintun, D.A. Wolk, A. Nordberg, J.O. Rinne, G. Chételat, A. Lleo, R. Blesa, J. Fortea, K. Madsen, K.M. Rodrigue, D.J. Brooks, Prevalence of amyloid PET positivity in dementia syndromes, *JAMA* 313 (19) (2015) 1939–1950, doi:10.1001/jama.2015.4669.
- [62] C. Smith, N. Malek, K. Grosset, B. Cullen, S. Gentleman, D.G. Grosset, Neuropathology of dementia in patients with Parkinson's disease: a systematic review of autopsy studies, *J. Neurol. Neurosurg. Psychiatry* (2019) 1234–1243, doi:10.1136/jnnp-2019-321111.