Computer Methods and Programs in Biomedicine 197 (2020) 105708

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

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

An application of generalized matrix learning vector quantization in neuroimaging

Rick van Veen[a,*], Vita Gurvits[c], Rosalie V. Kogan[c], Sanne K. Meles[e], Gert-Jan de Vries[b], Remco J. Renken[d], Maria C. Rodriguez-Oroz[f], Rafael Rodriguez-Rojas[f], Dario Arnaldi[g,h], Stefano Raffa[i,j], Bauke M. de Jong[e], Klaus L. Leenders[c], Michael Biehl[a]

[a] Bernoulli Institute for Mathematics, Computer Science and Artificial Intelligence, University of Groningen, the Netherlands
[b] Philips Research - Healthcare, the Netherlands
[c] Department of Nuclear Medicine & Molecular Imaging, University Medical Center Groningen, the Netherlands
[d] Department of Biomedical Sciences of Cells & Systems, Cognitive Neuroscience Center, University Medical Center Groningen, the Netherlands
[e] Department of Neurology, University Medical Centre Groningen, the Netherlands
[f] Clinica Universidad de Navarra and Centre for Applied Medical Research, Pamplona, Spain
[g] Department of Neuroscience, University of Genoa, Italy
[h] Neurology Clinic, IRCCS Ospedale Policlinico San Martino, Genoa, Italy
[i] Department of Health Sciences, University of Genoa, Italy
[j] Nuclear Medicine Unit, IRCCS Ospedale Policlinico San Martino, Genoa, Italy

ARTICLE INFO

Article history:
Received 27 January 2020
Accepted 8 August 2020

Keywords:
Parkinson's disease (PD)
[18F]Fluorodeoxyglucose positron emission tomography (FDG-PET)
Scaled sub-profile scaling model principal component analysis (SSM/PCA)
generalized matrix learning vector quantization (GMLVQ)

ABSTRACT

Background and objective: Neurodegenerative diseases like Parkinson's disease often take several years before they can be diagnosed reliably based on clinical grounds. Imaging techniques such as MRI are used to detect anatomical (structural) pathological changes. However, these kinds of changes are usually seen only late in the development. The measurement of functional brain activity by means of [18F]fluorodeoxyglucose positron emission tomography (FDG-PET) can provide useful information, but its interpretation is more difficult. The scaled sub-profile model principal component analysis (SSM/PCA) was shown to provide more useful information than other statistical techniques. Our objective is to improve the performance further by combining SSM/PCA and prototype-based generalized matrix learning vector quantization (GMLVQ).

Methods: We apply a combination of SSM/PCA and GMLVQ as a classifier. In order to demonstrate the combination's validity, we analyze FDG-PET data of Parkinson's disease (PD) patients collected at three different neuroimaging centers in Europe. We determine the diagnostic performance by performing a ten times repeated ten fold cross validation. Additionally, discriminant visualizations of the data are included. The prototypes and relevance of GMLVQ are transformed back to the original voxel space by exploiting the linearity of SSM/PCA. The resulting prototypes and relevance profiles have then been assessed by three neurologists.

Results: One important finding is that discriminative visualization can help to identify disease-related properties as well as differences which are due to center-specific factors. Secondly, the neurologist assessed the interpretability of the method and confirmed that prototypes are similar to known activity profiles of PD patients.

Conclusion: We have shown that the presented combination of SSM/PCA and GMLVQ can provide useful means to assess and better understand characteristic differences in FDG-PET data from PD patients and HCs. Based on the assessments by medical experts and the results of our computational analysis we conclude that the first steps towards a diagnostic support system have been taken successfully.

* Corresponding author.
E-mail address: rick.van.veen@rug.nl (R. van Veen).

https://doi.org/10.1016/j.cmpb.2020.105708

1. Introduction

Neurodegenerative diseases are often life-threatening. Besides lowering life expectancy, they impact the patients in ways that are life-changing for both the patients and their families. The two disorders far most prevalent are Parkinson's Disease (PD) and Alzheimer Disease (AD) [1,2]. Affecting approximately four and seven million people in Europe, respectively [3,4].

These two diseases together with a number of other less frequent similar conditions usually develop very slowly, often taking several years before they become diagnostically clear. The initial complaints of the patients and the clinical findings are often unclear or not severe enough to differentiate between the different diseases. This includes findings from neurocognitive examinations, neuroimaging scans using CT or MRI and other tests. The commonly used scans in clinical practice are performed in order to demonstrate or exclude anatomical (structural) pathological changes. If these structural changes are present indeed, those scans fulfill their role well. However, pathological anatomical brain changes are usually seen only late in the development of the disorders. Functional brain alterations as a result of slowly progressing brain diseases are more difficult to obtain and their interpretation is for clinicians often not possible or is much less reliable than structural lesions. A thorough clinical evaluation combined with imaging techniques, able to provide patterns of neuronal dysfunction which are specific to a particular disease, might become an indispensable approach to assist the accurate diagnosis and choice of appropriate treatment [5].

Currently, there is still a need to classify patients as early as possible. On the one hand the patients and family themselves wish to be informed about what is happening to them. This is especially important for the number of subjects who present with clinical alarm signs but do not develop brain degeneration and recover. On the other hand, the management of these brain diseases differ, depending on which disease will be most likely. Furthermore, there is a growing need to select patients at early or very early stages to be able to test experimental therapies. For that purpose one does not need a large group of subjects, but a highly specific group of patients. Experimental therapies will not be easily allowed on people if their diagnosis is uncertain and if a very large study group is required to determine a useful effect. In addition, clinical studies with large patient groups are difficult to handle and can be extremely costly. These negative hurdles could be avoided by methods with high specificity and sensitivity that allows to recruit on short notice a small group of patients for whom the therapy is justified. Lastly, there is a need of a method which can quantify the severity of the brain pathology directly from the cerebral tissue measurement by which progression or regression can be measured without being dependent solely on the subjective wellbeing of the patient. For this and similar purposes, machine learning methods have become increasingly popular within the health care domain. However, many of them are black box methods which are not transparent in how they work. Moreover, as illustrated above and discussed in literature [6,7], medical decision makers expect a diagnostic model to provide intuition by pointing to similar cases and explaining its decision process. Transparent or white box algorithms such as decision trees [8] exist. However, when working with complex high dimensional features these are not necessarily intuitive [8,9].

The method presented in this paper aims to fulfill the characteristics mentioned above by combining properties of the scaled sub-profile model principal component analysis (SSM/PCA) [10,11] with those of generalized matrix learning vector quantization (GMLVQ) [12]. We show how this combination allows us to construct *post-hoc visualizations* of the data and the model by combining the linearity of the PCA transformation with the prototypes and relevance matrix GMLVQ constructs during training. First, a discriminative visualization that allows to discover differences or similarities between scans in an understandable number of dimensions (Section 2.3.1). Second, a visualization of the internal representations (prototypes) of GMVLQ and its relevance matrix in the original voxel space (Section 3.2), helping domain experts to interpret the inner workings of the automated diagnostic process. In order for a model to be justifiable it needs to be in line with existing domain knowledge [13]. To validate the presented approach, we look at data from healthy controls (HC) and PD patients collected at three different neuroimaging centers of which the differences are known, see Section 2.4 for more details considering the data.

The SSM/PCA is a data driven spatial covariance method designed to reduce the high dimensional voxel data in a much smaller set of covariance patterns of which the expression scores can be computed on any functional brain data from the same imaging modality [14,15]. In our case the data was collected using [^{18}F]-fluoro-deoxyglucose positron emission tomography (FDG-PET). Previous studies have shown the reproducibility and stability of the PD related (covariance) patterns (PDRP) based on FDG-PET [16,17]. In our work, we deviate from the traditional way of constructing and applying these patterns, see Section 2.1.

The distance and prototype based GMLVQ employs a full relevance matrix in the distance metric and is thereby able to account for correlations of dimensions and rotations of the feature space. [12]. Previous work has demonstrated GMLVQ's superior or at least competitive performance to that of support vector machines and decision trees in similar diagnostic problems relating Parkinsonian syndromes and SSM/PCA [18,19]. Furthermore, the method has shown competitive performance and provided useful insights due to analysis of its relevance matrix in a multitude of other biomedical applications [20–24]. A brief review of GMLVQ is included in Section 2.2

The paper is structured following the classical machine learning pipeline. The methods and materials section introduces the theory, interpretation, and mathematics behind, what can be considered, the pre-processing and feature generation method SSM/PCA (Section 2.1), highlighting the difference with the traditional way of applying SSM/PCA. Section 2.2 provides a review of GMLVQ followed by Section 2.3 in which we discuss the intuitive combination of GMLVQ and SSM/PCA, resulting in the voxel representations of the prototypes and relevance matrix. To show this is not purely a theoretical exercise we validate the method by applying it to data from HC and PD patients described in Section 2.4. The visualizations are evaluated by three neurologists, the results and a discussion are presented in Section 3. The paper is concluded in Section 4.

2. Methods and materials

Here we provide the background information, relevant mathematical formalism and notation, as well as some intuition behind SSM/PCA and GMLVQ in Sections 2.1 and 2.2. This background is required for the combination of the two, presented in Section 2.3. Furthermore we introduce the practical application of our approach and include a description of the dataset at hand in Section 2.4.

2.1. Scaled sub-profile model/PCA

The scaled sub-profile model (SSM) [10,11], is a spatial covariance method based on principal component analysis (PCA). In contrast to structural data from, e.g., magnetic resonance imaging, SSM/PCA is applied to metabolic activity, i.e., functional data from FDG-PET.

A fundamental property of the brain lies at the basis of the procedure. The brain uses only glucose as the source of the cellular energy which the nerve cells in the brain need to perform and maintain their function. Glucose in the brain is metabolized using oxygen and by this metabolic process energy rich biological substances are produced which are then available for nerve and other cells in the brain. If for any reason the nerve cells are lost or are diseased they need less biological energy and thus consume less glucose. The overall glucose consumption in any particular brain region reflects therefore directly the locally averaged functional state of the brain. These glucose use changes may be subtle but significant and often cannot be seen visually. Due to the brain's high degree of connectivity and fixed neuronal interconnections between thousands of regions a æpatterng of altered glucose use may arise. It is assumed that each neurodegenerative brain disease expresses itself in specific patterns of regional metabolic covariation [10,11].

In order to detect the glucose uptake in the brain FDG-PET scanning uses the radiotracer FDG (fluoro-deoxy-glucose). This glucose variant behaves like native glucose but can be radiolabeled with the physiological positron emitting radio-isotope [^{18}F]. The PET camera registers the accumulation of the radio-isotope everywhere in the brain and obtains a total æfootprintg of the nervous tissue condition resulting in metabolic activity data per voxel.

The SSM is able to characterize the regional covariance structures by pre-processing of the measured voxel values, i.e., by working with the *subject residual profiles* (SRP) [25]. After determining the SRPs a PCA is applied to identify the metabolic covariance patterns, also known as the *group invariant sub-profiles* (GIS). The GIS are determined on the data from what we call the *space defining reference group*. This group contains both healthy controls and a disease specific patient subgroup, e.g., Parkinson's disease patients in our case (Section 2.4). Furthermore, one can calculate the *subjects' expression scores* of the obtained GIS patterns for unseen patients. The amount that each of the GIS contribute to a subject's cerebral metabolic rate is quantified by the expression scores which vary from subject to subject. This variability in subject scores has been shown to have high discriminative power in a number of cases [26–29]. A detailed theoretical description of the SSM/PCA can be found in [10,11]. The following subsections review the relevant computational steps involved, following a similar process described in [15].

2.1.1. Spatial preprocessing

Starting with raw FDG-PET volumes we first normalize the data to the Montreal Neurological Institute (MNI) common space, such that there is a direct correspondence between voxel coordinates. Next, the volumes are masked in order to remove, low-values, noise-related artifacts, and the areas not involving brain activity (e.g., the eyes). Initially, a set of individual masks are created by computing per volume the maximum value and marking, in the mask of that volume, each voxel above 35% of this maximum with a one and every other voxel with zero. These individual masks are then multiplicatively combined to create one common mask that will include only non-zero values for all subjects.

2.1.2. Subject residual profiles

To compute the subject residual profiles (SRP), each volume is reshaped into a one-dimensional vector ($m \times 1$) and concatenated to form a matrix with each row representing a subject. The data is logarithmically transformed and thereafter each subject is centered with its respective mean value. By centering logarithmically transformed data, subject global scaling effects are reduced. These effects may be caused by, e.g., differences in dosage or measuring parameters [15]. Finally, the subject residual profiles are the result of column (voxel) centering the data by subtracting the group mean profile (GMP), i.e., the mean voxel values computed over the healthy controls. As a consequence, the sign of the SRP values indicate if a subject has a higher (positive) or lower (negative) activity value with respect to the healthy controls' average. Alternatively, one could choose to use the entire group including the PD patients instead, but then the interpretation of the SRP values depend on the more heterogeneous PD group's attributes, decreasing interpretability.

2.1.3. Group invariant sub-profiles

The group invariant sub-profiles (GIS) are given by the eigenvalue decomposition of the SRP's voxel by voxel covariance matrix. The GIS correspond to the orthonormal eigenvectors obtained from this procedure. The details of how to perform such a PCA are described and discussed in [15,25]. The GIS ($m \times n$) are computed on the *space defining reference group* which limits the number of GIS patterns to the number of subjects (n) included in this group, with the size of the patterns equal to the number of voxels (m) that are left after applying the mask. Traditionally, the different GIS patterns are combined into a single disease related pattern, we refer the interested reader to [15,30] for examples on how to do this. However, in contrast to the traditional way of using SSM/PCA we do not combine any GIS patterns beforehand. Instead, we use all non-zero eigenvalue GIS to compute the subject scores that serve as input to the learning vector quantization (LVQ) system (Section 2.2) and let the model decide which scores or combination of scores best discriminate the different classes of data.

2.1.4. Subject scores

The forward application of SSM/PCA, also known as the topographic profile rating [31] provides the subject scores of an unseen subject. The same mask as calculated for the space defining reference group has to be used to ensure the selection of the same coordinates in the novel volumes. The SRP of the new subject is then computed using the steps described above, with the only difference being that the GMP is the one of the space defining reference group. A vector of subject scores ($\boldsymbol{\xi}$) of size $n \times 1$ quantifies the amount that each of the GIS patterns contributes to the subject's cerebral metabolic rate. It is computed as:

$$\boldsymbol{\xi} = G^{\top}\boldsymbol{r}, \tag{1}$$

where G a $m \times n$ matrix containing the GIS patterns of the space defining reference group and $\boldsymbol{r}$ is the subject residual profile ($m \times 1$) of a subject. Here, $\top$ is the transpose operator. As the matrix G is an orthonormal matrix, multiplying by the transpose of the GIS matrix can be used to approximately reverse the operation and get back the SRP:

$$\boldsymbol{r} = G\boldsymbol{\xi} \tag{2}$$

The transformation in Eq. (2) is especially important, as it enables the visualization described in Section 2.3.2.

2.2. Learning vector quantization

Learning vector quantization (LVQ) [32] is a family of machine learning algorithms that attempt to find typical patterns (prototypes) that represent the different labels in the data, i.e., the different groups of patients and healthy controls. The prototypes are constructed during the training phase based on the available samples. During this phase the model is presented with a patient's expression scores as described in Section 2.1.4. The prototypes are updated by pushing the closest correct prototype towards the considered patient's scores and the closest wrong prototype away from the considered patient's scores. After training, an unseen patient's data are classified by computing the distance, for instance the squared Euclidean distance, between the patient's scores and the

prototypical scores. The label of the new patient is then determined using the nearest prototype classification (NPC) scheme, i.e., it is assigned the label of the closest prototype.

2.2.1. Generalized matrix LVQ

Generalized matrix learning vector quantization (GMLVQ), introduced by Schneider et al. [12], is an extension to the cost function of the GLVQ algorithm [33] and the idea of incorporating relevance factors into the distance measure [34]. GMLVQ employs a full and adaptive matrix in the distance measure which is, together with the prototypes, optimized during training. The adaptive distance is denoted mathematically by:

$$d^{\Lambda}(\mathbf{w}, \boldsymbol{\xi}) = (\boldsymbol{\xi} - \mathbf{w})^{\top} \Lambda (\boldsymbol{\xi} - \mathbf{w}), \tag{3}$$

where $\boldsymbol{w}$ is a prototype of size $n \times 1$ and $\Lambda = \Omega^{\top}\Omega$ is a symmetric positive semidefinite $n \times n$ matrix, with n the number of features. 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 [12]. 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 a more robust performance [12,21]. Initial prototypes are generally set to be equal to the class conditional mean with small random offset. Initially, the matrix Ω is set equal to the identity matrix in order to not introduce any bias towards any of the dimensions. The updates of prototypes and relevance matrix are given by gradient descent optimization of an appropriate cost function, see [12] for details. During this procedure the algorithm is presented with the training data for a number of epochs until the model converges. Here, we consider the cost function introduced in [33]:

$$S = \sum_{i=1}^{N} f[\mu^{\Lambda}(\boldsymbol{\xi}_i)], \tag{4}$$

with $\mu^{\Lambda}(\boldsymbol{\xi}_i)$ being the relative distance function and the monotonically increasing function $f(x) = x$. The relative distance function,

$$\mu^{\Lambda}(\boldsymbol{\xi}_i) = \frac{d^{\Lambda}(\boldsymbol{\xi}_i, \boldsymbol{w}_L) - d^{\Lambda}(\boldsymbol{\xi}_i, \boldsymbol{w}_K)}{d^{\Lambda}(\boldsymbol{\xi}_i, \boldsymbol{w}_L) + d^{\Lambda}(\boldsymbol{\xi}_i, \boldsymbol{w}_K)}, \tag{5}$$

is a is a bounded quantity in the range $[-1, 1]$. A negative value indicates a correct classification, as $d^{\Lambda}(\boldsymbol{\xi}_i, \boldsymbol{w}_L)$ denotes the distance between the closest prototype with the same label ($\boldsymbol{w}_L$) as the presented example ($\boldsymbol{\xi}_i$). Consequently, a positive value means an incorrect classification as $\boldsymbol{w}_K$ denotes the closest prototype with a different label.

The eigenvectors of Λ provide the directions in which the data from different classes are well separated, while keeping the prototypes close to the data they represent. This property can be leveraged by projecting the data in feature space onto the leading eigenvectors of Λ enabling the user to view a discriminative low-dimensional representation of the data, see Section 2.3.1. This is similar in spirit to methods such as Linear Discriminant Analysis (LDA) [35], see [36] for a comparison of the concepts.

2.3. Visualization

In this section we discuss the two techniques that combine properties of SSM/PCA and GMLVQ in order to expand our understanding of the data. In particular, the visualizations can be used to more easily find similar or dissimilar patients (Section 2.3.1) or to visualize the inner representations that the GMLVQ model holds of the data (Section 2.3.2).

2.3.1. Discriminative visualization

Using the following technique we can get a low dimensional discriminative visualization of the data which can be used, e.g., to identify outliers and or find similar patients. It is important to note that "similar" is defined within the context of the data and the classification task and refers to patients that show similar activity patterns.

The adaptive matrix Ω as discussed in Section 2.2.1, can be interpreted as a linear transformation of the data and the prototypes. However, given a specific matrix Λ, the corresponding Ω is not necessarily uniquely defined. Intuitively, the distance measure is invariant under reflections or rotations so many solutions may exist. The found solution will depend on the initialization and the randomized order by which GMLVQ learns from the examples [36]. However, we can define a canonical, unique solution $\hat{\Omega}$ with the eigenvalue decomposition of Λ. Thereby, we determine the orthonormal eigenvectors $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_M$ of Λ, corresponding to the M ordered non-zero eigenvalues $l_1 \geq l_2 \geq \cdots \geq l_M$ and define $\hat{\Omega}$ as:

$$\hat{\Omega} = \left[\sqrt{l_1}\mathbf{e}_1, \sqrt{l_2}\mathbf{e}_2, \cdots, \sqrt{l_M}\mathbf{e}_M\right]^{\top}. \tag{6}$$

It has been shown analytically [36] and observed empirically that the GMLVQ approach has a strong tendency to yield singular matrices Λ of very low rank [12,36–38], so the number of non-zero eigenvalues will be small. Furthermore, because Λ is symmetrical the eigenvectors can be orthonormalized. Intuitively, the eigenvectors correspond to linearly independent sources causing the label dependent variance in the data, with the corresponding eigenvalues indicating the magnitude of this effect. For the remainder of this paper $\hat{\Omega}$ refers to the orthonormal eigenvectors of Λ. The method described here is applied after training. However, variants of LVQ exist that explicitly control the rank of the matrix during training [37].

2.3.2. Voxel space representation

We present here the technique that can be used to visualize the internal representation the GMLVQ model creates during the training phase. It is possible to look at and explain the different covariance patterns in voxel space, i.e., when the expression score for a specific pattern increases the activation in the positively weighted voxels regions increases as well, whereas activation in the negatively valued regions decreases [30]. However, in practice comparing the expression of multiple covariance patterns of a new scan with that of the prototypes is non-trivial. It would be easier if the prototypes would be in the same representation as the new scan. In the Section 2.2.1 we described that GMLVQ produces a prototype for each of the different classes in the data. By combining the two concepts, on one side the prototypes as representative values of a class and on the other side the relation (Eq. (2)) between the subject scores (feature space) and the subject residual profiles, it is possible to construct the prototypical subject residual profiles $\tilde{\boldsymbol{w}}$ in voxel space:

$$\tilde{\boldsymbol{w}} = G\boldsymbol{w}, \tag{7}$$

with $\boldsymbol{w}$ the prototype as trained in feature space and G the GIS matrix.

Similarly, the relevance matrix can be visualized [24]. As the GIS matrix $G^{\top}$ is a linear mapping of the high dimensional residual profiles (voxel space) to the lower dimensional subject scores, we can rewrite the distance measure in feature space of the form Eq. (3) as:

$$(\mathbf{x} - \tilde{\mathbf{w}})^{\top} G \Lambda G^{\top} (\mathbf{x} - \tilde{\mathbf{w}}), \tag{8}$$

with $\boldsymbol{x}$ the SRP. Hence, even though training and classification is formulated in feature space, we can compute the relevance matrix in voxel space $\tilde{\Lambda} = G\Lambda G^{\top}$. However, the visualization of the whole

Table 1
General features of the space defining reference group.

	UMCG	
	HC (n = 17)	PD (n = 19)
Age, mean(std) in years	61.3(7.5)	63.8(7.5)
Male gender, n(%)	12(70.6)	13(68.4)
Disease evolution, mean(std) in years	-	3(2)

matrix is not sensible. Therefore, we consider the diagonal of Λ only, which can be used as a summary of the relevance of the features or in voxel space the most relevant areas to consider for the classification task.

2.4. Datasets

In this setion we introduce the data used to test our methods. We consider a set of brain images acquired using [^{18}F]fluorodeoxyglucose positron emission tomography (FDG-PET) from both healthy controls (HC) and Parkinson's patients (PD). The data have been collected at three different centers: the Movement Disorder Unit of the Clinica Universidad de Navarra (CUN) [39], the University Medical Center Groningen (UMCG) [16], and the University of Genoa and IRCCS AOU San Martino-IST (UGOSM) [40]. A fourth disjoint group, see Table 1, of healthy controls and Parkinson's patients from the UMCG is included as our space defining reference group, as discussed in Section 2.1. The details of the different study setups can be found in the respective publications. However, in Table 2 some relevant features of the participating subjects are provided for convenience. Observe from the data in Table 2 that the PD patients from the three centers are in different stages of the disease evolution. The subjects from the UGOSM are at an early stage of PD. In contrast, the patients from the CUN are in a much later stage of the disease. Finally, the group of patients from the UMCG are at a slightly later stage than the patients from the UGOSM, resulting in a diverse group of patients with respect to disease evolution. Previously published works show that SSM/PCA created PDRPs can capture differences with the progression of Parkinson's disease [41] as correlates with clinical symptoms and treatments [42,43]. The healthy controls from UGOSM and CUN are relatively similar with respect to age. In contrast, the UMCG healthy group contains a more varied collection of relatively young individuals compared to the other two centers. In the centers, different scanners, setting, and scanning protocols are used [16,39,40] which has previously been observed to have a negative effect on the performance of GMLVQ trying to build a universally (center independent) applicable model [19]. In Table 1 the information of the space defining reference group is included. This is the group on which the GIS and GMP are calculated and therefore define the space the subject scores are calculated in and the models are trained in.

3. Results and discussion

In order to be able to assess about the usefulness of the visualizations it is relevant to know how well the model performs, as the performance of the model is essential for gaining trust [6].

We trained GMLVQ to distinguish between the healthy controls and Parkinson's patients using the data of each center individually as well as on all data together, see Table 3. The problem where the data of all centers is taken together has been labeled and trained in two ways. In the "2-class" setting the source of the data is not taken into account and the model has been trained to distinguish between HCs and PD patients. For the "6-class" problem we labeled every HC and PD patient in combination with the center of origin, resulting in six classes. However, the performance reported for the 6-class problem is that of the binary problem, i.e., how well the model distinguishes between PD patients and HC. Table 3 reports the mean and standard deviation of a ten times repeated randomized ten fold cross validation. To create a balanced dataset, underrepresented classes were oversampled by including copies of randomly selected patients in the training set. By balancing the training sets we decrease the bias in performance indicators, but in addition this increases the representativeness of the prototypes as the prototypes of underrepresented classes are not pushed away by other over represented groups. This is especially important for the interpretations of the voxel space representations of the prototypes and relevance matrix (Section 3.2). Balancing is furthermore justified because in the dataset the availability of data for Parkinson's disease does not represent the a priori possibility of a person having the disease. The prototype locations and relevance matrix were found by eight epochs of stochastic gradient descent optimization using the GMLVQ toolbox that can be found at [44], with the initial relevance matrix the identity matrix to not introduce any bias towards any of the GIS patterns and the prototypes (one per class) initialized on the class conditional means with a small random offset. All other parameters were set to the provided default values.

The results in Table 3a show the model is better at distinguishing between PD patients and HCs using the data from CUN and UMCG than it is at separating the patients from the UGOSM. This is likely explained by the observation that the patients of the three centers are at different stages of the disease, with the patients from UGOSM at the very beginning of the disease, UMCG at an early stage, and CUN at a very late stage. Furthermore, we assume homogeneous healthy controls, even though they are not necessarily healthy in an absolute sense, they do not have a neurodegenerative or any other brain disease.

In the performance results of the individual centers we see an average performance with relatively high standard deviation, which means that the performance depends heavily on which part of the data the algorithm is presented with during training. The variability of the AUC (area under the ROC curve) decreases when we consider the data of all the centers together as compared to data from single centers. Even though the standard deviation is still relatively high for the other metrics they can possibly be made more consistent between cross validation runs by introducing a bias towards one of the classes and shifting away from the NPC scheme along the ROC curve. Considering the 2-class problem, we do not know which source of variation GMLVQ is using to distinguish between the PD patients and HCs. It is possible that the model uses center dependent effects resulting in prototypes not describing a clean difference between HC and PD. This could be harmful for the interpretation of the prototypes and relevance matrix. For this reason, we also looked at the 6-class problem. Forcing

Table 2
General features of the different groups.

	UMCG		UGOSM		CUN	
	HC (n = 19)	PD (n = 20)	HC (n = 49)	PD (n = 38)	HC (n = 20)	PD (n = 68)
Age, mean(std) in years	56(14)	63(9)	67.8(11.6)	72(6.8)	67.9(3.1)	70.6(6.4)
Male gender, n(%)	9(47.4)	-	16(32.7)	24(63.2)	11(55)	37(54.4)
Disease evolution, mean(std) in years	-	3(2)	-	1.7(1.6)	-	13.6(5.1)

Table 3
Classification performance of HCs versus PD patients of the five different classification problems. Results of ten times repeated ten fold cross validation. Mean values with standard deviation within brackets. PPV and NPV are the positive and negative predictive values respectively. AUC is the area under the ROC curve.

Center(s)	All		Individual		
	2-class	6-class	UMCG	UGOSM	CUN
Accuracy (%)	83(7)	81(8)	87(16)	71(14)	86(11)
AUC	0.89(0.07)	0.81(0.09)	0.96(0.12)	0.79(0.14)	0.89(0.16)
Sensitivity (%)	86(8)	81(9)	82(29)	69(20)	88(12)
Specificity (%)	78(13)	81(15)	92(20)	75(22)	81(31)
PPV (%)	88(7)	89(8)	94(14)	80(14)	95(8)
NPV (%)	77(12)	71(11)	88(18)	69(16)	70(28)

(a) Performance of the approach, as described in the previous sections, when using all the GIS to generate the subjects' scores

Center(s)	All		Individual		
	2-class	6-class	UMCG	UGOSM	CUN
Accuracy (%)	74(9)	72(9)	83(18)	68(14)	70(15)
AUC	0.82(0.10)	0.31(0.09)	0.92(0.15)	0.76(0.16)	0.77(0.23)
Sensitivity (%)	74(12)	71(13)	77(29)	66(19)	71(15)
Specificity (%)	73(16)	75(16)	90(21)	70(22)	69(35)
PPV (%)	84(8)	84(9)	92(16)	76(17)	90(11)
NPV (%)	63(13)	61(12)	82(21)	62(18)	41(24)

(b) Performance of the approach when using the first GIS only to generate a single score per subject.

GMLVQ to separate the centers in combination with the diagnostic labels could help find prototypes that more clearly describe the difference. The performance measure considered for evaluating the system trained in the 6-class setting still refers to the problem of discriminating between PD patients and HCs.

Previous studies have shown that that the first GIS alone tends to describe the difference between PD and HCs [29]. One of the reasons to not reduce the number of GIS beforehand was in order to improve the reconstruction performance of the method described in Section 3.2, as more information would be available. Furthermore, one of the strengths of GMLVQ comes from the inclusion of the relevance matrix that is optimized to weigh the (relevance of) features and their combinations. Additionally, the relevance matrix enables to construct a discriminative visualization as described in Section 2.3.1. The relevance profiles of the 2-class and 6-class have been included in Fig. 1 and the profiles for the other centers can be found in Appendix B. To see how our approach performs when using the first GIS only we have run the same experiments on the same data, but with generating only a single score (feature) per subject using the first GIS of the space defining reference group. The results are included in Table 3b. The GIS are ordered by variance accounted for (VAF) which have been included in Appendix A, where the first GIS accounts for almost 18% of the variance. Note that GMLVQ reduces to a simple one dimensional threshold classifier for 2-class problems. The performance in Table 3 show a decrease in performance when using the first GIS. However, it does not drop far, which does indicate that the first GIS has high discriminative power with respect to HCs and PD patients. Especially the lower drop in performance of the UGOSM and UMCG compared to that of CUN might indicate that later signs of PD could be contained within different GIS. This is furthermore supported by the relevance profile of CUN (Fig. B.8), as this set contains only very late stage PD patients (Table 2), and shows that the first GIS is not the most relevant for the discrimination.

The relevance profiles of the 2-class and 6-class problem have been included in Fig. 1. The profile of the 2-class problem shows that the first GIS is the most relevant one, but clearly the other GIS carry relevant information that facilitates increased performance of the classifier. The relevance profile of the 6-class problem tells a similar story with the exception that the 35th GIS is the most relevant. Based on the previous findings this suggests that the first

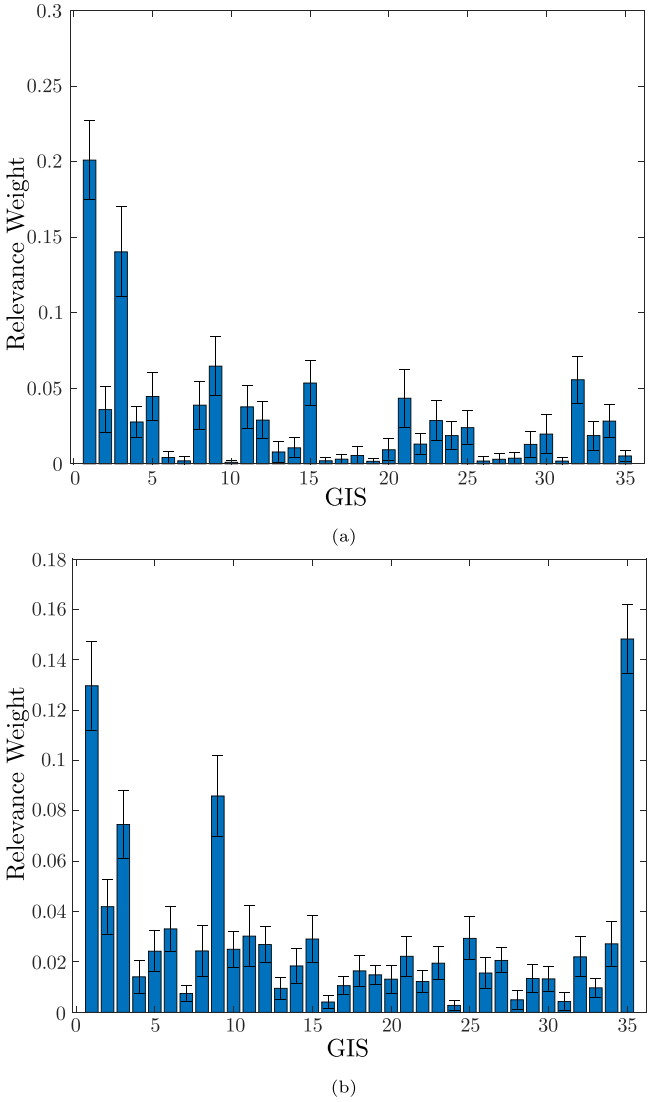

Fig. 1. Relevance profiles, i.e., the diagonal component of the relevance matrix Λ_{ii} (y-axis) of (a) the 2-class (b) the 6-class problems. The values are the average value found during the randomized ten times repeated ten fold cross validation, with the error bars the standard deviation.

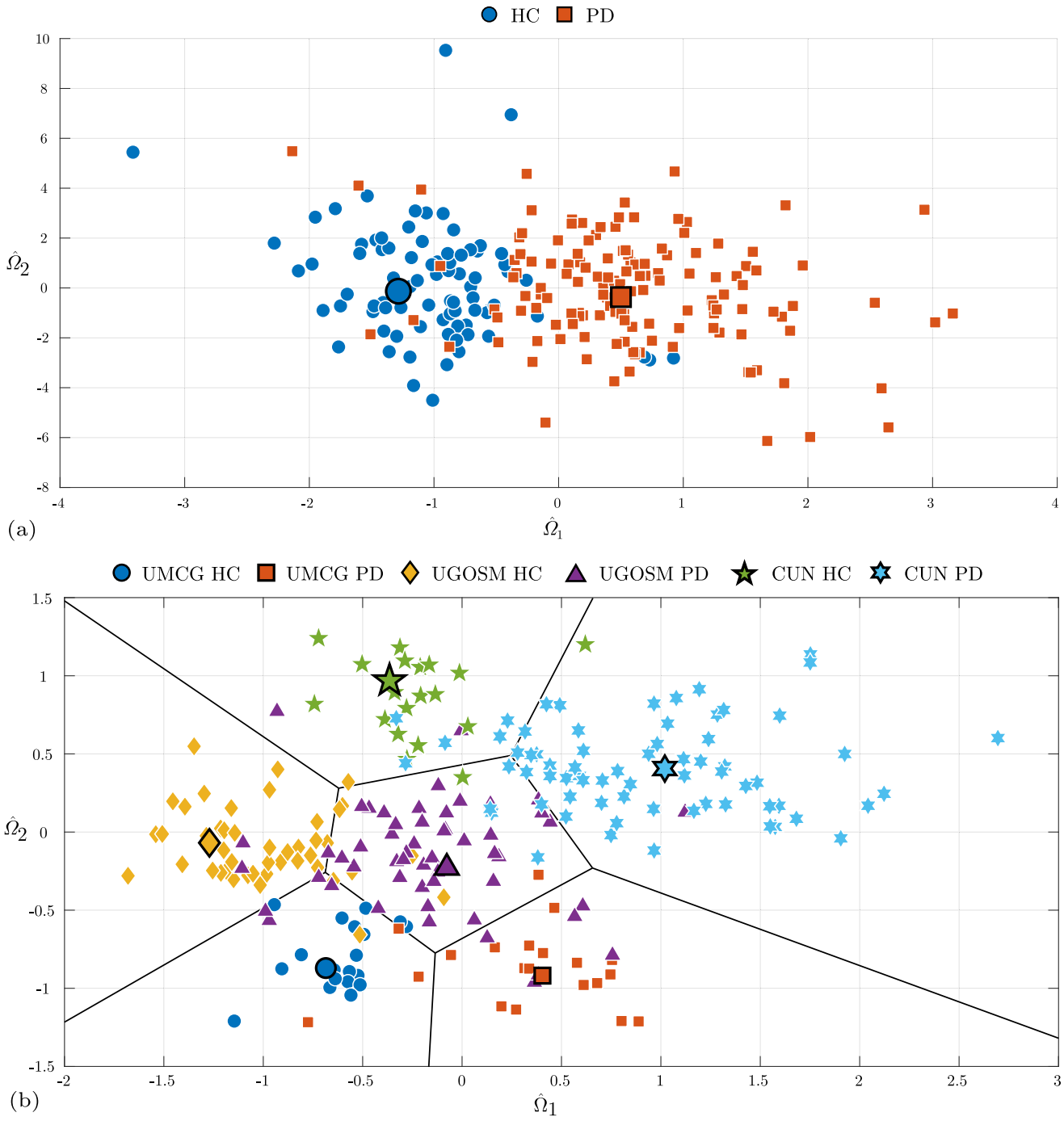

Fig. 2. Projection of the subject scores from all available centers and prototypes (depicted larger and with black border) onto the first two leading eigenvectors of the relevance matrix Λ, scaled by the square root of their eigenvalues (columns of $\hat{\Omega}$) for two different problem setting. (a) Labeling considering health status only. Eigenvalues for $\hat{\Omega}_1$ and $\hat{\Omega}_2$ are 1 and 0 respectively, reducing the 2D plot to a one dimensional one. (b) Labeling health status in combination with center of origin. Eigenvalues for $\hat{\Omega}_1$ and $\hat{\Omega}_2$ are 0.5331 and 0.2517 respectively.

GIS is mostly related to the difference between PD and HC. In particular, the 35th GIS appears to comprise information relevant for the discrimination of the centers.

3.1. Discriminative visualization

This section presents the results and discussion of the visualization from Section 2.3.1 that provides a technique to visualize the high dimensional voxel data in a scatter plot. Furthermore, it can help to understand what the model has learned.

In Fig. 2a we have included the two-dimensional discriminative plot of the 2-class (Fig. 2a) and 6-class problem (Fig. 2b). These plots were created using the method described in Section 2.3.1 after a single training run on all available data of GMLVQ following the same procedure and with the same hyperparameters as for the models in Table 3. Therefore, the space of the presented models is representative of a model with the same expected performance of distinguishing between PD and HC as can be found in Table 3. The plotted points represent single patient's subject scores from the training data projected onto the corresponding problems first two

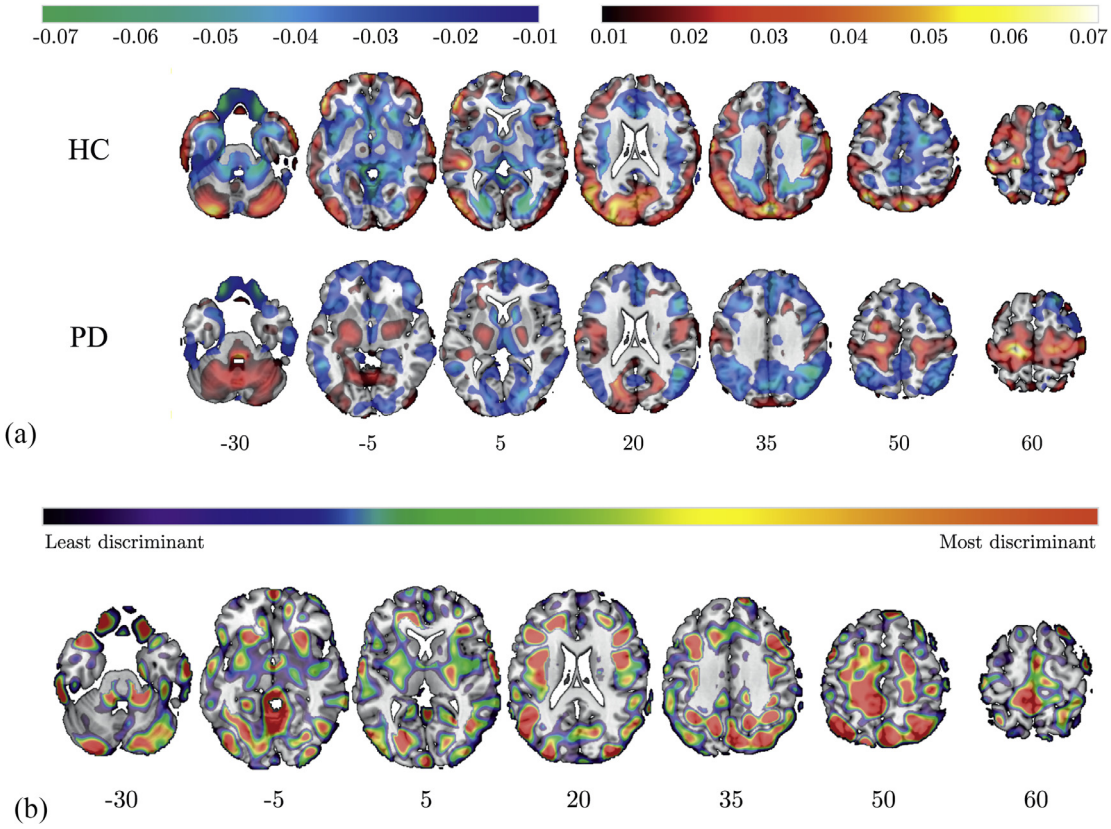

Fig. 3. The GMLVQ model of the 2-class problem projected back into voxel space: (a) The found prototypical SRPs (prototypes) of the two considered classes, with values relative to the healthy controls of the space defining reference group. (b) The diagonal of the relevance matrix, indicating relevant areas for discrimination between the two classes. Note that the prototypes are represented by the colors which are projected on the mni152 template [45,46].

columns of $\hat{\Omega}$. The prototypes are depicted larger and with a black and thicker border. Additionally, in Fig. 2b the decision boundaries are drawn between prototypes when using the NPC scheme. It is important to realize that the second eigenvector in Fig. 2a has been included for aesthetic reasons and will not count in the calculation of the distance as it's eigenvalue is close to zero, and therefore is unimportant for the classification. Similarly, it is possible in Fig. 2b that some points might seem to be misclassified but in fact are not due to the third non-zero eigenvalued eigenvector ($\hat{\Omega}_3$) not visible in the plot.

Looking at the projection in Fig. 2a we can see a relatively clear separation between the two classes, as expected considering the performance in Table 3 for the 2-class problem. With an eigenvalue of approximately one the only relevant axis is $\hat{\Omega}_1$, which thus describes the difference between the classes. As we believe a center dependent effect exist [19], we suspect that the direction of $\hat{\Omega}_1$ might not contain the difference between HC and PD subjects only, but might also use the difference between centers to distinguish between HCs and PD patients.

In Fig. 2b we have included the projection of the 6-class problem. The idea of the 6-class setting is to construct more representative prototypes by considering the center of origin of the data during training. The performance will not necessarily be better by doing so, as the model will try to separate the different centers and therefore will not be able to use any center dependent difference to distinguish between HC and PD. The plot in Fig. 2b confirms the existence of a center specific effect as all classes are separable, an effect that cannot be explained by difference in PD cohorts only (Table 2) under the assumptions that HC cohorts are homogeneous.

Considering the projection of the data on $\hat{\Omega}_2$, we would find clusters on a line mostly describing the difference between centers as HCs and PD patients from the same center would fall mostly within the same segment. Interestingly, the first eigenvector ($\hat{\Omega}_1$) still describes most of the difference between the HCs and PD patients, comparing to Fig. 2a and considering every center separately. At a first glance one might conclude that the PD patient clusters arrange from left to right corresponding with the disease evolution progression of the different cohorts, see Table 2. However, comparing HCs and PD subjects of each center we see that the HCs clusters shift together with the PD clusters indicating that the direction does not contain an effect explained only by the difference in PD disease progression. Under the assumption that the healthy controls generally do not differ, it is likely some of the center specific differences are included in this direction. If we would find a clean direction without any center specific differences we would be able, besides distinguishing between HC and PD, to compute a severity score along this direction indicating the progression of PD.

3.2. Voxel representation

For the validation of the interpretability of the prototypes and relevance matrices, we look at the five different classification problems as introduced in the beginning of Section 3 and perform the transformation described in Section 2.3.2 to obtain a prototypical residual profile and relevance profile in the original voxel space. This gives us fourteen prototypes and five relevance profiles.

Each of the prototypes and relevance profiles are presented to three medical specialists in the follow way: (i) In order to determine if the prototypes show activity patterns and areas of interest that compare to the knowledge of experts, we show the constructed images without providing any information about them, i.e., no knowledge about the considered disease, centers, and train-

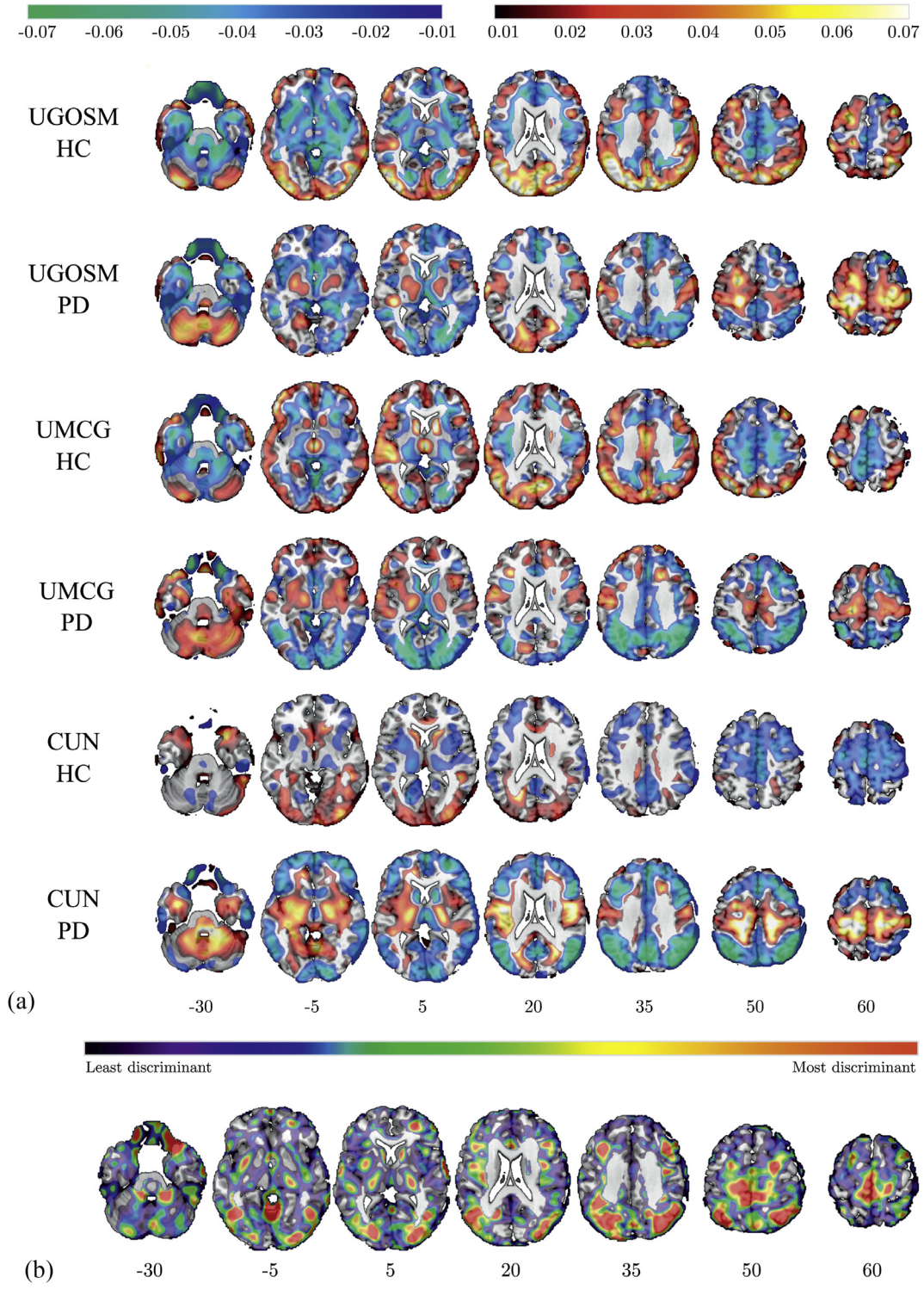

Fig. 4. The GMLVQ model of the 6-class problem projected back into voxel space: (a) The found prototypical SRPs (prototypes) of the six considered classes, with values relative to the healthy controls of the space defining reference group. (b) The diagonal of the relevance matrix, indicating relevant areas for discrimination between the six classes. Note that the prototypes are represented by the colors which are projected on the mni152 template [45,46].

ing settings. Furthermore, they are asked to describe what they see and possibly give a diagnosis. After this the classes are revealed to the experts, i.e., the general class of PD patient or healthy control, no specifics about the cohort's average disease evolution, age or gender. By including the prototypes trained on the data of multiple centers we try to see if this has any negative effects on the representativeness of the prototypes. (ii) In a second round, now that the classes are known, a number of prototypes that have significant differences in disease evolution are shown in pairs to see if this difference is captured by the prototypes. Initially without conveying this information to the assessors. By doing so, we hope they find a noticeable difference between prototypes representing different stages of the same disease, and without bias are able to tell which prototypes represent a further progressed patient. (iii)

Finally, the relevance profiles are presented to the assessors to see if the important areas according to LVQ are similar to the ones traditionally considered to be important by the domain experts.

The medical specialists observed, in the prototypes in Figs. 3 and 4, an increased metabolism in the globus pallidus and putamen, thalamus, cerebellum, pons, and sensorimotor cortex and relative decreases in the lateral frontal and parietooccipital areas, which are similar activities to the to the previously found PD-related patterns (PDRP) in literature [26]. Although exact difference in network activity between patients at different disease evolution stages are not known we do know that the expression of the PDRP does correlate with clinically observable symptoms and different stages of the disease [41,42]. We see similar signs in the prototypes shown in Fig. 2b, the patterns of activity do not change. However, there is an increase of the activity in the relevant areas for patients from CUN, of which the patients are much further in the disease evolution, compared to the patients from the UMCG and UGOSM.

Comparing the visual representations in Fig. 4 it can be seen that there are differences between the prototypical distributions of the subject residual profiles. In the CUN and the UMC group the parieto-occipital regions show clear reductions in metabolic activity but this less so in the UGOSM group. This could be due to the differences of group compositions. The UGOSM group included many more subjects with rather short disease durations. This might suggest that PD patients who have been diseased for a longer period of time do show more pathology. The parieto-occipital region is suggested to be involved in Parkinsonian subjects who develop cognitive deficits or even dementia. However, this study was not designed to investigate the involvement of the regional changes in relation to the clinical signs and thus conclusions in this respect are not warranted.

It is important to note here that hard conclusions are difficult to make because of the presence of the variability between centers as well as between HC and PD. Interestingly, within this setting the relevance profiles Fig. 2a and b do still capture similar areas of importance as indicated by the medical specialists, with a higher relevance or discriminative power given to the basal ganglia, thalamus, oxipital cortex, and motor cortex.

4. Conclusion

We have shown that the techniques presented in this paper can be used to improve the interpretability of the GMLVQ model in combination with SSM/PCA and in the context of the diagnosis of Parkinson's disease. The low-dimensional visualization of the subjects opens up the potential to project unseen subjects onto a low-dimensional plot. Moreover, this can help visualize potential issues with the data, such as multiple sources of variation which can not all be contributed to the difference in pathology. Most importantly, the voxel representation of the model makes it possible to follow the decision process of GMLVQ by looking at images medical experts are more familiar with. Additionally, we have shown that very similar to previous work the first GIS are considered most relevant by GMLVQ, but do not provide the same level of performance compared to using all GIS as relevant information is present in later GIS as well. Future studies looking at the relevance profiles of GMLVQ might prove insightful with respect to which GISs would most likely capture certain sources of variance, either the variance due to disease evolution stage or difference due to protocols or processing materials.

By looking at a relatively well understood disease, i.e., PD, we have shown that the model can produce prototypical activity profiles capturing the pathological differences between subjects. The goal is to use our approach on other datasets including multiple pathologies that are clinically relevant and more difficult to distinguish. The various diseases which are accompanied by clinical Parkinsonism may thus be separated on a consistent pathophysiological basis. Not including any other sources of variability. Another possible application would be to project an unseen disease group such as idiopathic REM sleep behavior disorder (RBD), which is considered a pre-stage of PD, into the eigenvectors of a model trained to distinguish between HC and PD subjects. Potentially, this can provide some insight into the difference between RBD patients that will and which will not develop Parkinson's disease.

Ethical approval

The data in this study has been reused from previous clinical studies. No new clinical data was gathered or used. There is thus no need for an ethical approval.

Funding

This project was (partially) funded by: The Michael J. Fox Foundation with grant ID 17081, the European Unions Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No. 676157, and the Dutch 'Stichting Parkinson-Fonds'.

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.2020.105708.

References

[1] European commission, Public health, (http://ec.europa.eu/health/). [Online; accessed 11-April-2018].
[2] S. Przedborski, M. Vila, V. Jackson-Lewis, Series introduction: neurodegeneration: what is it and where are we? J. Clin. Invest. 111 (1) (2003) 3–10, doi:10.1172/JCI17522.
[3] European commission, Website of EU joint programme - neurodegenerative disease research, (http://www.neurodegenerationresearch.eu/). [Online; accessed 11-April-2018].
[4] M.J. Firbank, A.J. Yarnall, R.A. Lawson, G.W. Duncan, T.K. Khoo, G.S. Petrides, J.T. O'Brien, R.A. Barker, R.J. Maxwell, D.J. Brooks, D.J. Burn, Cerebral glucose metabolism and cognition in newly diagnosed Parkinson's disease: ICICLE-PD study, J. Neurol. Neurosurg. Psychiatry 88 (4) (2017) 310–316, doi:10.1136/jnnp-2016-313918.
[5] S. Risacher, A. Saykin, Neuroimaging biomarkers of neurodegenerative diseases and dementia, Semin. Neurol. 33 (4) (2013) 386–416.
[6] Z.C. Lipton, The Mythos of model interpretability, Queue 16 (3) (2018) 30:31–30:57, doi:10.1145/3236386.3241340.
[7] Z.C. Lipton, The Doctor Just Won't Accept That!, arXiv:1711.08037(2017).
[8] D. Mudali, L.K. Teune, R.J. Renken, K.L. Leenders, J.B. 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.
[9] 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.
[10] J.R. Moeller, S.C. Strother, J.J. Sidtis, D.A. Rottenberg, Scaled subprofile model: a statistical approach to the analysis of functional patterns in positron emission tomographic data, J. Cereb. Blood Flow Metab. 7 (5) (1987) 649–658, doi:10.1038/jcbfm.1987.118.
[11] J.R. Moeller, S.C. Strother, A regional covariance approach to the analysis of functional patterns in positron emission tomographic data, J. Cereb. Blood Flow Metab. 11 (1_suppl) (1991) A121–A135, doi:10.1038/jcbfm.1991.47.
[12] 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.
[13] D. Martens, J. Vanthienen, W. Verbeke, B. Baesens, Performance of classification models from a user perspective, Decis. Support Syst. 51 (4) (2011) 782–793, doi:10.1016/j.dss.2011.01.013.

[14] D. Eidelberg, Metabolic brain networks in neurodegenerative disorders: a functional imaging approach, Trends Neurosci. 32 (10) (2009) 548–557, doi:10.1016/j.tins.2009.06.003.

[15] P.G. Spetsieris, D. Eidelberg, Scaled subprofile modeling of resting state imaging data in Parkinson's disease: methodological issues, NeuroImage 54 (4) (2011) 2899–2914, doi:10.1016/j.neuroimage.2010.10.025.

[16] L.K. Teune, A.L. Bartels, B.M. de Jong, A.T.M. Willemsen, S.A. Eshuis, J.J. de Vries, J.C.H. van Oostrom, K.L. Leenders, Typical cerebral metabolic patterns in neurodegenerative brain diseases, Mov. Disord. 25 (14) (2010) 2395–2404, doi:10.1002/mds.23291.

[17] 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) (2020) 437–450, doi:10.1007/s00259-019-04570-7.

[18] D. Mudali, M. Biehl, K.L. Leenders, J.B.T.M. Roerdink, LVQ and SVM classification of FDG-PET brain data, Advances in Intelligent Systems and Computing, vol. 428, Springer International Publishing, 2016, doi:10.1007/978-3-319-28518-4.

[19] R. van Veen, L. Talavera Martinez, R. Kogan, S. Meles, D. Mudali, J.B.T.M. Roerdink, F. Massa, M. Grazzini, J.A. Obeso, M. Rodriguez-Oroz, K.L. Leenders, R.J. Renken, J.J.G. de Vries, M. Biehl, Machine learning based analysis of FDG-PET image data for the diagnosis of neurodegenerative diseases, in: N. Petkov, N. Strisciuglio, C. Travieso-González (Eds.), Applications of Intelligent Systems, Frontiers in Artificial Intelligence and Applications, vol. 310, 2018, pp. 280–289, doi:10.3233/978-1-61499-929-4-280.

[20] W. Arlt, M. Biehl, A.E. Taylor, S. Hahner, R. Libé, B.A. Hughes, P. Schneider, D.J. Smith, H. Stiekema, N. Krone, E. Porfiri, G. Opocher, J. Bertherat, F. Mantero, B. Allolio, M. Terzolo, P. Nightingale, C.H.L. Shackleton, X. Bertagna, M. Fassnacht, P.M. Stewart, Urine steroid metabolomics as a biomarker tool for detecting malignancy in adrenal tumors, J. Clin. Endocrinol. Metab. 96 (12) (2011) 3775–3784, doi:10.1210/jc.2011-1565.

[21] 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.

[22] L. Yeo, N. Adlard, M. Biehl, M. Juarez, T. Smallie, M. Snow, C.D. Buckley, K. Raza, A. Filer, D. Scheel-Toellner, Expression of chemokines CXCL4 and CXCL7 by synovial macrophages defines an early stage of rheumatoid arthritis, Ann. Rheumatic Dis. 75 (4) (2016) 763–771, doi:10.1136/annrheumdis-2014-206921.

[23] G. Mukherjee, G. Bhanot, K. Raines, S. Sastry, S. Doniach, M. Biehl, Predicting recurrence in clear cell renal cell carcinoma: analysis of TCGA data using outlier analysis and generalized matrix LVQ, in: 2016 IEEE Congress on Evolutionary Computation, CEC 2016, 2016, pp. 656–661, doi:10.1109/CEC.2016.7743855.

[24] M. Biehl, Biomedical applications of prototype based classifiers and relevance learning, in: International Conference on Algorithms for Computational Biology, Springer, 2017, pp. 3–23.

[25] G.E. Alexander, J.R. Moeller, Application of the scaled subprofile model to functional imaging in neuropsychiatric disorders: a principal component approach to modeling brain function in disease, Hum. Brain Mapp. 2 (1–2) (1994) 79–94, doi:10.1002/hbm.460020108.

[26] S.K. Meles, L.K. Teune, B.M. De Jong, R.A. Dierckx, K.L. Leenders, Metabolic imaging in Parkinson disease, J. Nucl. Med. 58 (1) (2017) 23–28, doi:10.2967/jnumed.116.183152.

[27] M. Niethammer, C.C. Tang, A. Feigin, P.J. Allen, L. Heinen, S. Hellwig, F. Amtage, E. Hanspal, J.P. Vonsattel, K.L. Poston, P.T. Meyer, K.L. Leenders, D. Eidelberg, A disease-specific metabolic brain network associated with corticobasal degeneration, Brain 137 (11) (2014) 3036–3046, doi:10.1093/brain/awu256.

[28] C.C. Tang, K.L. Poston, T. Eckert, A. Feigin, S. Frucht, M. Gudesblatt, V. Dhawan, M. Lesser, J.P. Vonsattel, S. Fahn, D. Eidelberg, Differential diagnosis of parkinsonism: a metabolic imaging study using pattern analysis, Lancet Neurol. 9 (2) (2010) 149–158, doi:10.1016/S1474-4422(10)70002-8.

[29] P.G. Spetsieris, J.H. Ko, C.C. Tang, A. Nazem, W. Sako, S. Peng, Y. Ma, V. Dhawan, D. Eidelberg, Metabolic resting-state brain networks in health and disease, Proc. Natl. Acad. Sci. U. S. A. 112 (8) (2015) 2563–2568, doi:10.1073/pnas.1411011112.

[30] 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, doi:10.1016/j.neuroimage.2008.01.056.

[31] D. Eidelberg, J.R. Moeller, T. Ishikawa, V. Dhawan, P. Spetsieris, T. Chaly, W. Robeson, J.R. Dahl, D. Margouleff, Assessment of disease severity in parkinsonism with fluorine-18-fluorodeoxyglucose and pet, J. Nucl. Med. 36 (3) (1995) 378–383.

[32] T. Kohonen, The self-organizing map, Proc. IEEE 78 (9) (1990) 1464–1480, doi:10.1109/5.58325.

[33] A. Sato, K. Yamada, Generalized learning vector quantization, in: Advances in Neural Network Information Processing Systems, 1996, pp. 423–429.

[34] B. Hammer, M. Strickert, T. Villmann, Supervised neural gas with general similarity measure, Neural Process. Lett. 21 (1) (2005) 21–44, doi:10.1007/s11063-004-3255-2.

[35] E. Alpaydin, Introduction to Machine Learning, The MIT Press, second ed., pp. 128–133.

[36] 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.

[37] 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.

[38] 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/wcs.1378.

[39] D. Garcia-Garcia, P. Clavero, C. Gasca 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.

[40] D. Arnaldi, S. Morbelli, A. Brugnolo, N. Girtler, A. Picco, M. Ferrara, J. Accardo, A. Buschiazzo, F. de Carli, M. Pagani, F. Nobili, Functional neuroimaging and clinical features of drug naive patients with de novo Parkinson's disease and probable RBD, Parkinsonism Relat. Disord. 29 (2016) 47–53, doi:10.1016/j.parkreldis.2016.05.031.

[41] C. Huang, C. Tang, A. Feigin, M. Lesser, Y. Ma, M. Pourfar, V. Dhawan, D. Eidelberg, Changes in network activity with the progression of Parkinson's disease, Brain 130 (7) (2007) 1834–1846, doi:10.1093/brain/awm086.

[42] P. Tomše, L. Jensterle, M. Grmek, K. Zaletel, Z. Pirtošek, V. Dhawan, S. Peng, D. Eidelberg, Y. Ma, M. Trošt, Abnormal metabolic brain network associated with Parkinson's disease: replication on a new European sample, Neuroradiology 59 (5) (2017) 507–515, doi:10.1007/s00234-017-1821-3.

[43] R. Rodriguez-Rojas, J.A. Pineda-Pardo, R. Martinez-Fernandez, R.V. Kogan, C.A. Sanchez-Catasus, M. del Alamo, F. Hernández, L. García-Cañamaque, K.L. Leenders, J.A. Obeso, Functional impact of subthalamotomy by magnetic resonanceguided focused ultrasound in Parkinson's disease: a hybrid PET/MR study of resting-state brain metabolism, Eur. J. Nucl. Med. Mol.Imaging (2019), doi:10.1007/s00259-019-04497-z.

[44] M. Biehl, P. Schneider, K. Bunte, Relevance and matrix adaptation in learning vector quantization (GRLVQ, GMLVQ and LiRaM LVQ), (http://matlabserver.cs.rug.nl/gmlvqweb/web/ or direct dowload from https://www.cs.rug.nl/~biehl/LVQ_toolbox.tar.gz). [Online; accessed 09-July-2020].

[45] V. Fonov, A. Evans, R. McKinstry, C. Almli, D. Collins, Unbiased nonlinear average age-appropriate brain templates from birth to adulthood, NeuroImage 47 (2009) S102, doi:10.1016/S1053-8119(09)70884-5.

[46] V. Fonov, A.C. Evans, K. Botteron, C.R. Almli, R.C. McKinstry, D.L. Collins, Unbiased average age-appropriate atlases for pediatric studies, NeuroImage 54 (1) (2011) 313–327, doi:10.1016/j.neuroimage.2010.07.033.