Identification of Parkinson's and Alzheimer's diseases through the analysis of FDG-PET images using machine learning techniques

Lidia Talavera Martínez

Abstract

Alzheimers disease (AD) and Parkinsons disease (PD) are the two most common neurodegenerative brain disorders, which are incurable diseases characterized by a progressive nerve cells dysfunction. Since early diagnosis of these conditions is difficult if based on the clinical symptoms only, the clinical evaluation is combined with brain imaging techniques like [18F]-fluorodeoxyglucose positron emission tomography (FDG-PET). The neuroimaging techniques enable identification of disease-related changes in cerebral glucose metabolism. In this project, we use machine learning techniques to perform an automated classification of FDG-PET brain scans. This approach is mainly based on the extraction of features and their later classification. Scaled subprofile model (SSM) is applied to FDG-PET data to extract characteristics patterns of glucose metabolism. These features are used for distinguishing different PD subjects from AD subjects and healthy controls using two classification methods, Generalized Matrix Learning Vector Quantization (GMLVQ), and Support Vector Machine (SVM) with linear kernel.

I. Introduction

THE nervous system has two major divisions: the central nervous system (CNS), which consists of the brain and the spinal cord, and the peripheral nervous system (PNS), which consists of nerves and small concentrations of gray matter called ganglia. An adult brain contains about 100 billion neurons, of which part of them connect the CNS with the parts of the body that control the muscles and internal organs [1].

Neurodegenerative diseases represent more than 600 disorders which affect the nervous system. They are characterized by a progressive nerve cells dysfunction [2]. Although, the main cause that provokes the disease remains unknown, the combination of genetics and environmental factors, altogether with the natural process of ageing, seems to have a relevant role. Existing treatments for neurodegenerative diseases are very limited, they only focus on reducing symptoms to enable an improvement of the quality of life and a normal life expectancy [3].

Since age is one of the principal risk factors, the prevalence of these neurodegenerative brain disorders is expected to increase significantly by 2030, as the European population is rapidly ageing. Currently, 16% of the European population is over 65, and it is expected to reach 25% by 2030 [4].

Alzheimers disease (AD) and Parkinsons disease (PD) are the two most common neurodegenerative brain diseases. AD, with an average duration between 2 and 10 years, is a type of dementia that affects over 7 million people in Europe [4], causing problems with memory, thinking and behaviour. On the other hand, PD is manifested clinically by bradykinesia, muscular rigidity, rest tremor and postural instability. It affects over 4 million people and over 80% of patients with PD will eventually develop dementia (PDD) [5].

Differentiating between neurodegenerative brain disorders and diagnose them at early disease stages can be challenging on clinical grounds only, as they have overlapping clinical and pathological features. In fact, as described in [6], up to 15% of patients diagnosed with early PD at the time of recruitment into trials were later found to have other parkinsonian syndromes. A thorough clinical evaluation combined with imaging techniques, able to provide patterns of neuronal dysfunction which are specific to a particular disease, has become an indispensable approach to assist an accurate diagnosis and choice of an appropriate treatment to be followed by the patient.

The technological advances have allowed the development of computer aided algorithms, allowing a strong capacity of data processing. These algorithms facilitate the work of the physician and reduce the discomfort caused to the patient. In addition, they are, in many cases, more effective in making available much more information about the patient's history to the doctor.

Author: Lidia Talavera Martínez, talaveramartinez.l@gmail.com
Advisor: Michael Biehl, Department of Computing Science, University of Groningen
Thesis dissertation submitted: September 2017

The objectives of this project are:

- Study the use of machine learning techniques in the analysis of FDG-PET image data.
- Identify the center where a specific sample comes from. This is important as we would like to build a universal classifier able to classify any sample as HC, PD or AD regardless of whether the classifier has been trained with data from the same center or not.
- Discriminate among the different types of subjects by center.
- Compare the similarity of the centers by training a classifier with data of one center and a later testing into data from other center.

This work is structured as follows. In the next chapter we establish the contextual basis in which the project is immersed. Then, in section III, we describe the functionality and implementation of the approach used in the classification of FDG-PET images. The next two sections IV and V establish the experimental framework in which we describe the experiments done and the results obtained, respectively. Finally, in section VI, we reflect on our findings and present conclusions.

II. State of the art

This chapter will outline some aspects of the three main areas addressed in this work: neurodegenerative brain diseases, neuroimaging and computer aided diagnosis.

A. *Neurodegenerative brain diseases*

Neurodegenerative brain diseases are incurable diseases strongly linked with age, that result in progressive loss of structure or function of nerve cells in the human brain. Neurons, responsible for sending and receiving nerve impulses or messages, normally do not reproduce themselves, so when they become damaged or die they cannot be replaced [4]. Figure 1 illustrates how neurons interact via synaptic connections. This will be helpful to understand how neurodegenerative brain diseases affect normal functions [7].

Changes of synaptic activity in the brain are accompanied by proportional changes in local glucose consumption. Hence, loss of neurons may result in decreased glucose consumption in distant brain regions by deafferentiation, whereas also increased regional glucose consumption by increased activation of afferent neurons can occur. Thus producing degeneration of specific brain regions and development of specific patterns of metabolic brain activity [8].

Thereby, neurodegenerative diseases which cause problems with movement or mental functioning, grouping them as parkinsonian syndromes and dementia, respectively. Within parkinsonian syndromes we can differenciate between *Parkinson's disease*

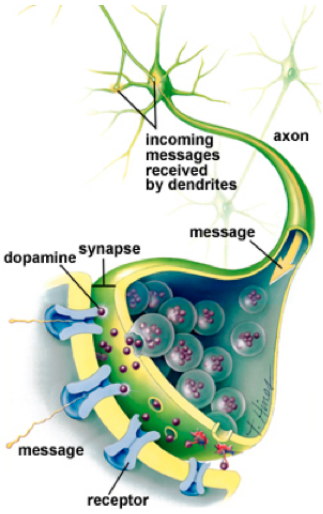


Figure 1: Neurons interaction via synaptic connections [1]. Incoming messages from the dendrites are passed to the axon where the nerve cell is stimulated to release neurotransmitters (dopamine) into the synapse. The neighboring nerve cell receptors pick up these chemical messengers and effectively transmit the message onto the next nerve cell. Image taken from [7].

[1] Structure that permits a neuron to pass an electrical or chemical signal to another neuron.

(PD), *Multiple System Atrophy* (MSA), *Progressive Supranuclear Palsy* (PSP) and *Corticobasal degeneration* (CBD). In the case of dementia we can distinguish among *Alzheimer's Disease* (AD), *Dementia with Lewy Bodies* (DLB) and *Frontotemporal Dementia* (FTD).

Even though these diseases have different pathophysiology and clinical evolution, PD, MSA, and DLB share the pathological feature of disturbed α-synuclein protein while PSP and CBD show disturbances in τ protein handling [8]. This overlapping of feature makes more difficult to diagnose different diseases in the early stages. More than 60% of cases with a final clinical diagnosis of a parkinsonian syndrome different than PD had their diagnosis changed during the course of their illness [9].

About 85% of people with Parkinsonism have PD [10]. Whilst AD is the most common cause of dementia accounting for 50-60% of all such cases [8].

Nowadays, existing treatments for neurodegenerative diseases are very limited, and address only the symptoms rather than the cause [4].

1) ***Parkinson's disease (PD)****:* is the most common degenerative parkinsonism, usually begins in late middle age and its prevalence is likely to increase in the next decades as the population is aging. Mainly, it is characterized by loss of dopamine producing nerve cells that emanate from the substantia nigra pars compacta in the midbrain, and project to the striatum (putamen and caudate nucleus), that regulate movement, see Figure 2, [11][12]. Another characteristic would be the presence of depositions of misfolded α-synuclein (Lewy bodies) in the surviving and eventually died neurons [13][14]. Decease of the nerve cells continues very slowly and it is only when 80% of these cells are lost that the main motor symptoms of Parkinson appear [7][15]. As the degeneration advances, it also spreads to other neurotransmitter systems, which gives rise to other nonmotor symptoms the body.

The clinical diagnosis of PD is based on a person's medical history and a neurological examination, varying from person to person, as does the rate of progression [7]. Nevertheless, in general terms, it relies on the presence of multisystem disorders with motor and non-motor symptoms. Parkinson's symptoms usually begin gradually and get worse over time [17].

The motor dysfunctions hallmarks are bradykinesia (slow movements), rigidity (muscle stiffness), rest tremor and postural instability. Furthermore, there are also multiple nonmotor evidences that leads to cognitive impairment, mood disorders (depression and anxiety), sleep disorders, olfactory dysfunction, and autonomic disorders (constipation, genitourinary dysfunction). Some of these nonmotor symptoms, which increases in number and severity with advancing disease, can precede motor features as prodromal symptoms, having thus implications in the diagnosis and development of preventive strategies. At advanced disease stages, the presence of dementia in PD is estimated to be approximately 40% [13]. Being associated with a higher risk in PD patients with sleep behavior disorder (RBD). In contrast, patients with a tremor-dominant phenotype rarely develop dementia [11][18][14].

Cognitive dysfunction may appear in 15-20% of even early stage, untreated PD patients and eventually be found in over 80%

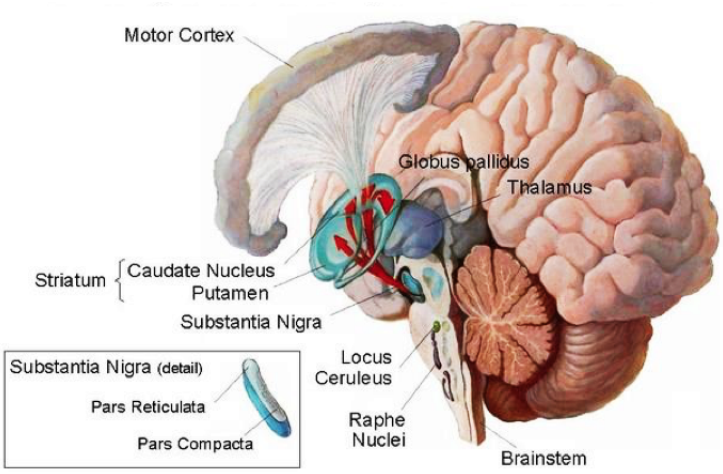


Figure 2: Dopamine pathway for the motor system. Dopamine signals travel from the substantia nigra to brain regions including the corpus striatum, the globus pallidus, and the thalamus in order to control movement and balance. Figure from [16].

of patients during long-term follow-up [18]. Sometimes 85% of patients with PD have olfactory loss, which can be measurable very early in disease, 2-5 years before motor symptoms [19]. Depression has been reported in 45 to 75% of cases. Around 80% of patients with rapid eye movement (REM) RBD may convert to PD within 10 to 15 years [18].

Although the exact cause of why dopamine producing cells become lost remains unknown, researchers focus on how the natural process of aging affects cell death, a combination of genetic mutations and environmental factors, side effects of medications and traumatic brain injury [18][7].

Nowadays, treatments focus on reducing symptoms to enable a more active lifestyle and a normal life expectancy. Primarily, PD can be managed with medication, diet, exercise, and deep brain stimulation surgery [7].

2) ***Alzheimer's disease (AD)****:* is the most common cause of dementia and accounts for 50-60% of all these cases [8], being the fourth leading cause of death in individuals over 65 years [20]. It is characterized by progressive, accumulative nerve cell death and tissue loss throughout the brain, which over time shrinks dramatically [21], see Figure 3.

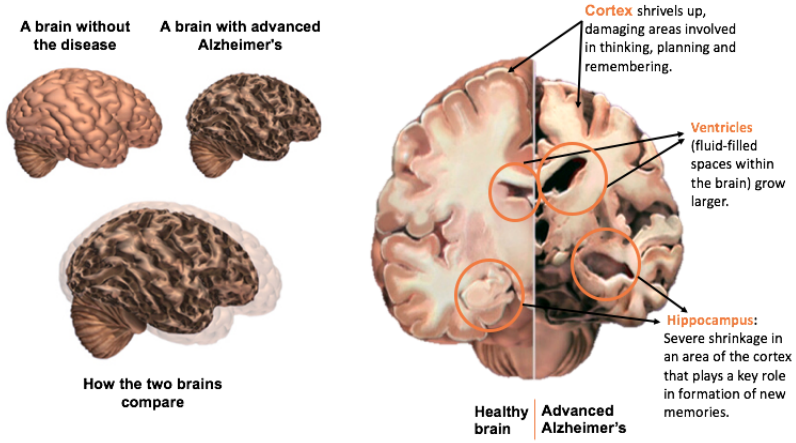


Figure 3: How massive cell loss changes the whole brain in advanced Alzheimer's disease. Figure taken from [21].

This is thought to be due to the presence of abnormal clusters of β-amyloid protein fragments built up between nerve cells (Plaques) and tangles formed from twisted strands of τ protein contained among dead and dying neurons [21]. The damage at neuronal structures and interconnectivity (synapses), disrupts the transmitters meant to carry messages in the brain. In Figure 4, we can see how plaques and tangles spread through the brain at the different stages of the AD, affecting particularly those functions which are related with memory, thinking, language, behaviour and emotion. Thus memory impairment together with progressive cognitive decline such as impaired judgement, decision making and orientation are the main hallmarks in AD [22][20].

The diagnosis of the disease is not straightforward as the symptoms vary from person to person. In general, the only definitive way to diagnose it is to find out whether plaques and tangles exist in brain tissue [21].

As it is explained in [22], at the moment there is no curative treatment for dementia, although many of the associated problems such as restlessness and depression can be treated. There are some drugs available for people with mild to moderate AD that may temporarily slow down the progression of symptoms.

B. Neuroimaging

The differential diagnosis of neurodegenerative brain diseases may be difficult on clinical grounds only, especially at an early disease stage where there are overlapping features among some diseases [8]. Because of the diagnostic challenges, the development of neuroimaging techniques provided considerable advancement in obtaining imaging data and deriving significant patterns of changes in brain activity. In addition, the combination of the use of imaging techniques together a thorough clinical examination led to gain in diagnostic accuracy [15][23].

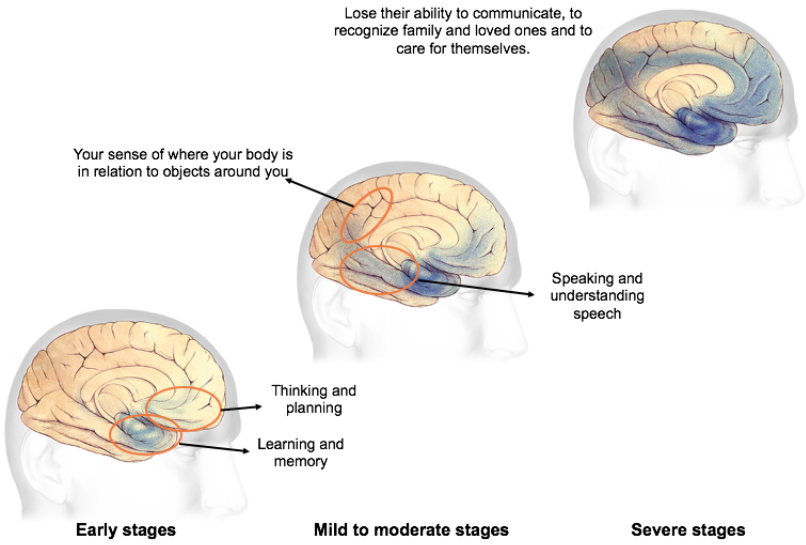


Figure 4: How plaques and tangles begin to form in different brain areas depending on the stage of the disease and what involve. Figure taken from [21].

Distinctive patterns of brain glucose metabolism, network activity and brain structure have been found in various neurodegenerative brain diseases, such as parkinsonian syndromes (PD, MSA, PSP, CBD) and dementia (AD, DLB, FTD), characterised by the degeneration of specific regions of the brain [24].

A disease specific pattern for PD was first described by Eidelberg. The related pattern is represented by relatively increased metabolism in the globus pallidus and putamen, thalamus, cerebellum, pons, and sensorimotor cortex and relative decreases in the lateral frontal and parieto-occipital areas [11].

Regarding the AD related pattern, it is represented by decreased metabolic activity in the angular gyrus and other parieto-temporal regions including precuneus, extending to the posterior and middle cingulate gyrus. As well as, relatively increased metabolic activity in some white matter regions, cerebellum, pons and sensorimotor cortex [8]. In addition, metabolism is typically relatively preserved in the visual and anterior cingulate cortices, basal ganglia and thalamus [20].

Diagnosis of neurodegenerative disease at an early disease is important because treatment and also the expectancy of future motor or cognitive problems depend on it. Therefore, early disease detection can be aided by brain structural evaluation with magnetic resonance imaging (MRI), and functional and molecular assessment of cerebral metabolism with nuclear imaging methods such as 18F-FluoroDeoxyGlucose Positron Emission Tomography (FDG-PET) and Single Photon Emission Computed Tomography (SPECT). These techniques help to understand the mechanisms underlying in the development neurodegenerative diseases [14][24][15][23].

Structural imaging is in general not very helpful because functional changes (i.e. changes in synaptic activity of specific brain regions) precede structural changes (i.e. neuronal death and brain atrophy). However, the latter are extremely useful for excluding structural abnormalities at later disease stages. These abnormalities can be mass lesions, infarcts, or hydrocephalus, which may produce symptoms mimicking neurodegenerative disease [15].

On the other hand, PET is a nuclear imaging technique, that allows in vivo estimation of metabolic processes. This method provides an index for regional neuronal activity by assessing specific biochemical activities of the human brain. FDG is a radiopharmaceutical analogue of glucose, the main energy source of the brain, and it is transported from the blood to the brain. As it is not a substrate for any enzyme present in brain tissue, it remains trapped for some time and therefore can be used as an imaging marker [8]. Consequently, FDG-PET is useful for imaging regional glucose consumption in the brain, where FDG uptake increases with synaptic activity and decreases with neural dysfunction in a pattern like manner, following the neuronal systems that are affected [11].

Comparative studies have shown FDG-PET to be superior to regional cerebral blood flow SPECT. This is likely due to the higher intrinsic spatial resolution of PET and the improved attenuation correction algorithms used with 3D tomographic reconstruction [13][20].

In Figure 5, we can observe typical images of FDG-PET imaging of a healthy control (HC), PD, MSA, PSP, CBD, DLB, AD and FTD subject.

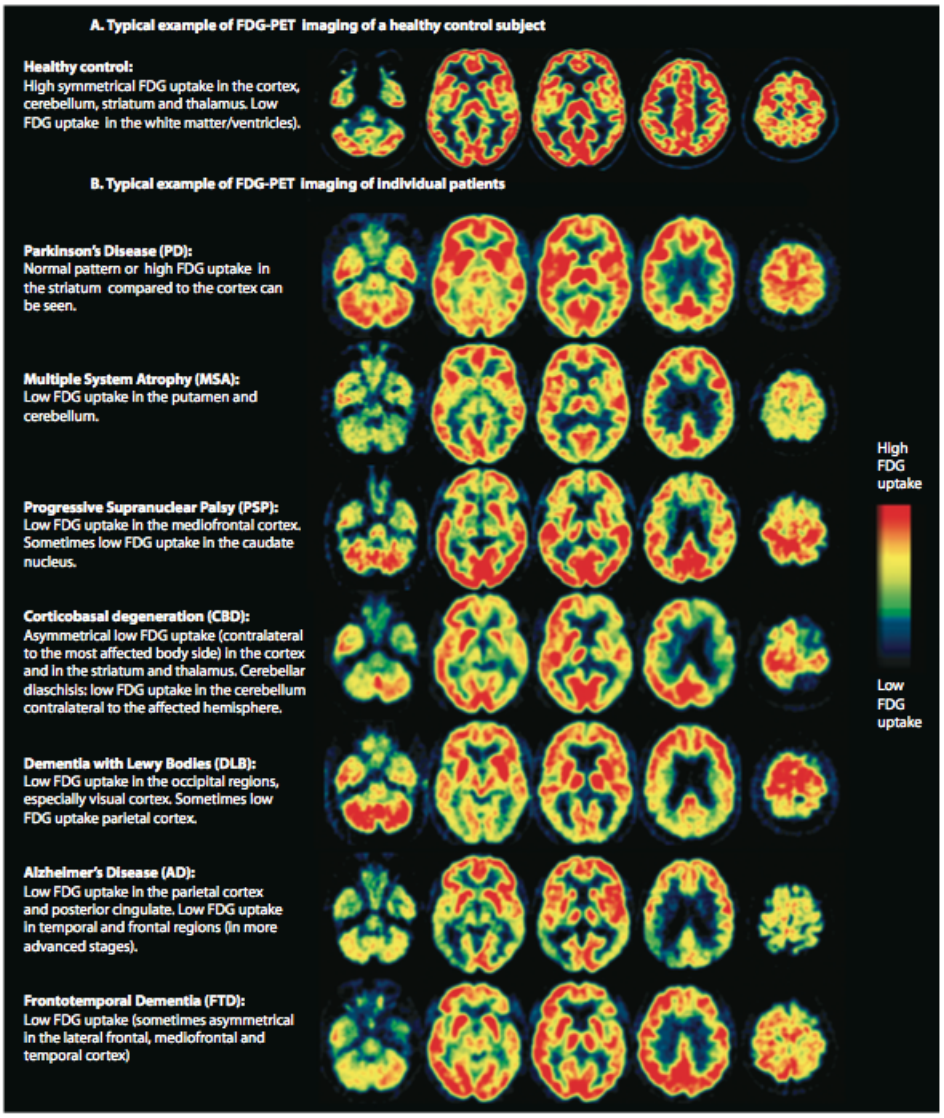


Figure 5: Examples of FDG-PET imaging of HC subjects and individual patients. [8].

C. Computer aided diagnosis

Distinguishing different neurodegenerative disorders based only on the baseline clinical assessment may be difficult, but the automated analysis of medical imaging data sets has led to an advancement in their evaluation. Aiming a more reliable and objective evaluation of the results, a greater storage and a better observation of diverse characteristics.

This is feasible thanks to computer-aided diagnosis (CAD), defined in [9] as a system that integrates data processing, mathematics and statistics into computerized techniques to maximize the information that may be extracted from data. One of the goals of CAD is to assist the clinicians in the differential diagnosis between several conditions with overlapping clinical features, such as neurodegenerative brain diseases. CAD techniques are mainly based on the extraction of disease patterns and their classification, involving four principal steps: data acquisition (neuroimaging), feature extraction, feature selection, and classification.

Concerning feature extraction, automated voxel-based statistical mapping such as statistical parametric mapping (SPM)[2] is

[2]Software package for the analysis of brain imaging data. See http://www.fil.ion.ucl.ac.uk/spm/software/

used for an initial normalization step. Then scaled subprofile model principal component analysis, SSM/PCA [25][26], would be an appropriate method to extract characteristics patterns of certain diseases from brain imaging data [27]. Besides being an example of biomarker indicator, also helps to reduce data dimensions. Finally, the resulting subject scores from the SSM/PCA are used as input features for the classifier [11][23][8].

We consider some machine learning approaches to perform an automated image based classification.

The majority of published work is based on support vector machine (SVM) as trainable classifiers. SVM classification generates a hyperplane as boundary that separates different classes of data points in the feature space. The decision boundary should maximize the distance, called margin, between the hyperplane and the support vectors [23]. Other state-of-the-art classifiers are decision trees (DT) [28] and generalized matrix relevance LVQ (GMLVQ) [29] that have shown good performance on the discrimination of patients with several diseases and HC. DT is a supervised technique which builds a classifier from a set of training samples with a list of features and a class label. The algorithm splits a set of training samples into subsets based on an optimal threshold value of a single feature. The result is a tree in which each leaf carries a class name and each interior node specifies a test on a particular feature [23]. On the other hand, GMLVQ is also a supervised system, but it is based on learning vector quantization [30], that can be applied to multi-class classification problems and where classes are represented by prototypes. GMLVQ differs from normal LVQ, it the sense that GMLVQ relies on an adaptive distance measure to optimize the discrimination of classes, instead of standard Euclidean distance.

In [31], the authors used three different classification methods, C4.5 decision tree [32], SVM with linear kernel and GMLVQ, to distinguish advanced PD stage subjects from HC. To assess the classifiers performance, they apply leave-one-out cross validation (LOOCV). As a result, they show how SVM and GMLVQ classifiers provide a significant improvement as compared to the DT classifier. However, even DT classifiers with a performance around 87% show an improvement regarding the previous results in [23], where the poor results were attributed to the small data set or that the data was obtained at an early disease stage.

III. METHOD

This chapter focuses first on the description of the subjects that form part of the study, then on how the data set was obtained and preprocessed. Finally, we specify the steps followed for the implementation of different algorithms that have been applied to perform the classification of FDG-PET images. In Figure 6, we can see the flow diagram of process that has been followed.

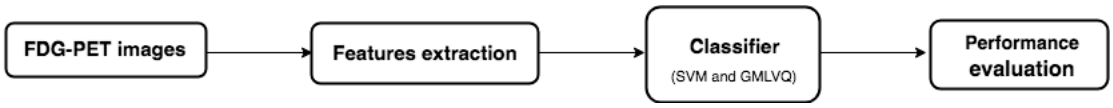


Figure 6: Work flow for the application and comparison of SVM and GMLVQ classifiers.

A. *Group data*

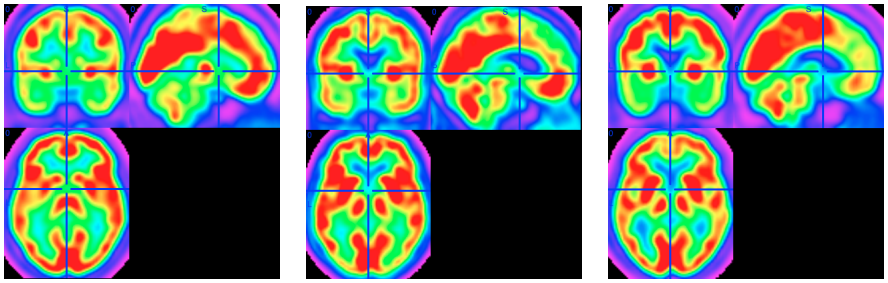
Subject brain images were acquired from three hospitals. First, from the Movement Disorders Unit of the Clinica Universidad de Navarra (CUN), Spain, we consider 49 patients diagnosed with PD whose clinical and metabolic data was published in [33]. Also, 19 HC subjects were included. Second, 20 PD subjects, 21 AD subjects and 19 HC were obtained from the University Medical Center Groningen (UMCG), for more details see [34]. Finally, we got 58 PD subjects, 55 AD subjects and 44 HC from the University of Genoa and IRCCS AOU San Martino-IST (UGOSM). In the patients from these three centers there is not a significant difference in the age, gender and disease duration of PD and AD patients.

Table I shows the data sets from the Clinica Universidad de Navarra (CUN), the University Medical Center Groningen (UMCG) and the University of Genoa and IRCCS AOU San Martino-IST (UGOSM).

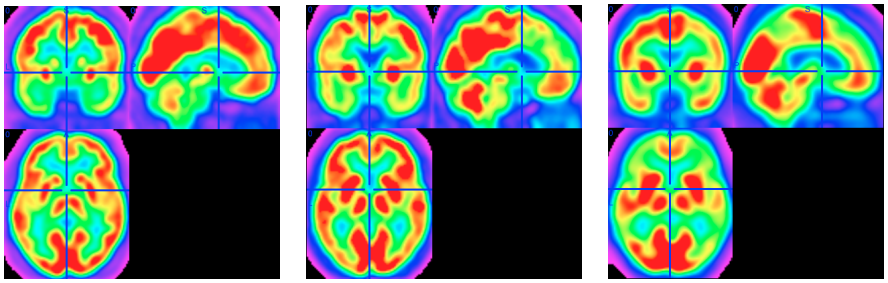
	Subjects		
Source	HC	PD	AD
CUN	19	68	-
UGOSM	44	58	55
UMCG	19	20	21

Table I: Distribution of the subjects from their respective sources.

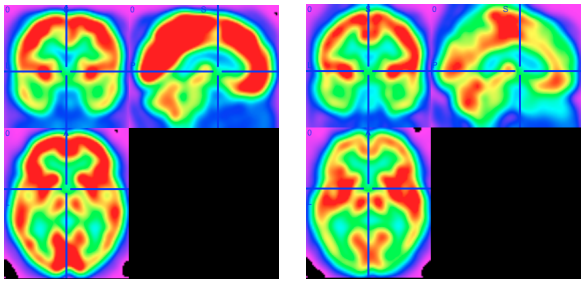
In Figure 7 we can see an example of the FDG-PET images that can be found in the different data sets with which we have worked.



(a) HC UMCG **(b)** PD UMCG **(c)** AD UMCG



(d) HC UGOSM **(e)** PD UGOSM **(f)** AD UGOSM



(g) HC CUN **(h)** PD CUN

Figure 7: Example of the FDG-PET images that can be found in the different data sets with which we have worked.

*B. **Feature extraction***

In order to extract features, subject scores, from the FDG-PET brain images, we apply the implementation of SSM/PCA method in Matlab used in [31], and which is based on the description by Spetsieris and Eidelberg [27].

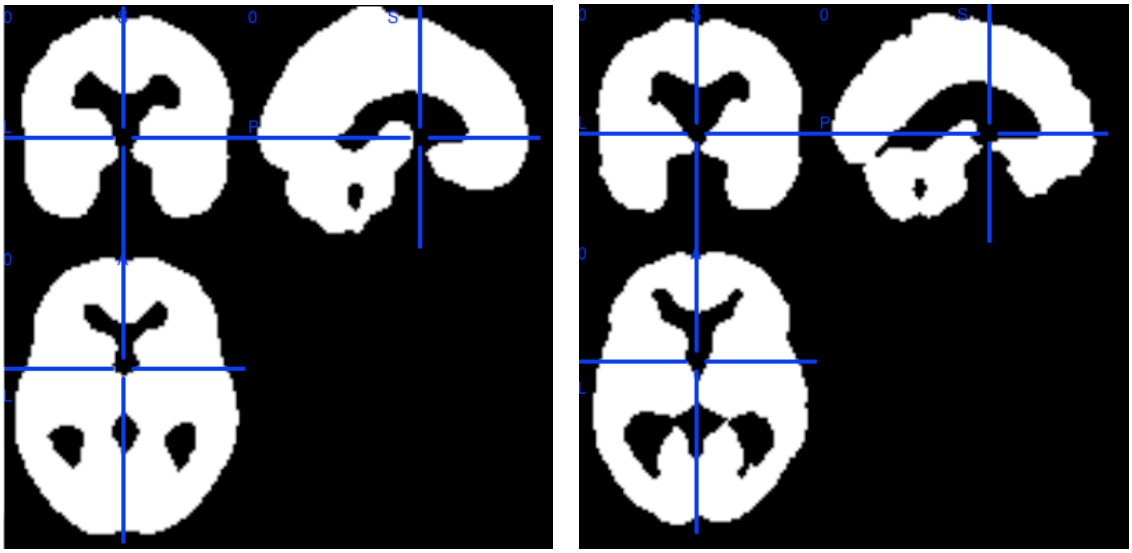
Our approach can be decomposed in the following steps:

1) Consider all the FDG-PET images as a combined group of different patient cohorts.
2) Use SPM software, a voxel based approach that employs classical inference, to characterize the functional brain anatomy of every subject, $s = (1, ..., M)$. For each subject's image the voxel values are stringed together to form a single row vector $v = (1, , N)$, creating a $M \times N$ data matrix, P_{sv}.
3) Brain data,P_{sv}, is masked to reduce low values and noise. This has been done by using a mask of the specific disease-related pattern (PDRP and ADRP), see Figure 8, which was obtained by applying SSM/PCA to FDG-PET data of HC and disease patients of UMCG. The mask can be applied to other centers due to the invariability and specificity of the disease-related pattern, becoming a true descriptor of the abnormal regional interactions that characterize the disorder.
4) The masked data is logarithmically transformed to separate irrelevant subject and regional scaling effects.

$$P_{sv} \rightarrow LogP_{sv}$$

5) Double centering is applied to remove the mentioned scaling effects. It is based on the subtraction of the mean data from, both, subjects and voxels data in $LogP_{sv}$. This normalization ensures that the analysis will focus on relevant subject-by-voxel interation effects being invariant to subject and regional scaling effects.

$$Q_{sv} = LogP_{sv} - mean_{vox}(LogP_{sv})$$



(a) **(b)**

Figure 8: Masks of a) PDRP and b) ADRP.

$$SRP_{sv} = Q_{sv} - mean_{sub}(Q_{sv})$$

6) The subject scores are obtained by the projection (inner product) of the principal components of the disease-related brain pattern into the log-transformed and doubly centered matrix,SRP_{sv}.

All the extracted subject scores were provided as features to build the SVM and GMLVQ classifiers for purposes of differential classification, and thus of diagnosis.

C. Support Vector Machine (SVM)

The SVM is a well known method introduced by Cortes and Vapnik [35]. It can be defined as a binary classification method whose solution is a linear classifier obtained from a quadratic optimization problem based on finding the maximum margin between the two hyperplanes containing the support vectors, defined as the samples of both classes closest to the boundary [36].

In this way, the approach only would work for linearly separable classes, being impossible to find a hyperplane able to classify data that are non-linearly separable. So, in order to make the classifier more efficient in classifying data sets where classes overlap or can only be separated by a non-linear decision boundary, two modifications have been introduced.

The first one which is called soft margin, allows the relaxation of the maximal margin condition. It implements a tolerance to errors, understood as vectors violating the margin condition. This tolerance of misclassification is controlled by slack variables, which carry a regularization factor that controls the balance between margin maximization and error minimization in the training samples.

The second adjustment known as kernel trick, aims to solve the problem of non-linearly separable data by mapping the feature space into a higher dimensional space in which the data set is linearly separable. For that it is not necessary to define the mapping function between the original space and the higher dimension space. It is only needed to define the scalar product (kernel) for satisfying that the distance defined in the transformed space has a relationship to the distance in the original space.

Both, the kernel parameters and the regularization factor introduce additional parameters, and their values have to be adjusted during the training stage, for instance by means of cross-validation techniques.

D. Learning Vector Quantization (LVQ)

Introduced by Kohonen in 1986, LVQ constitutes a prototype-based supervised classification algorithm, that can be applied to multi-class classification problems [37]. The prototypes represent the different classes of the input data and are defined

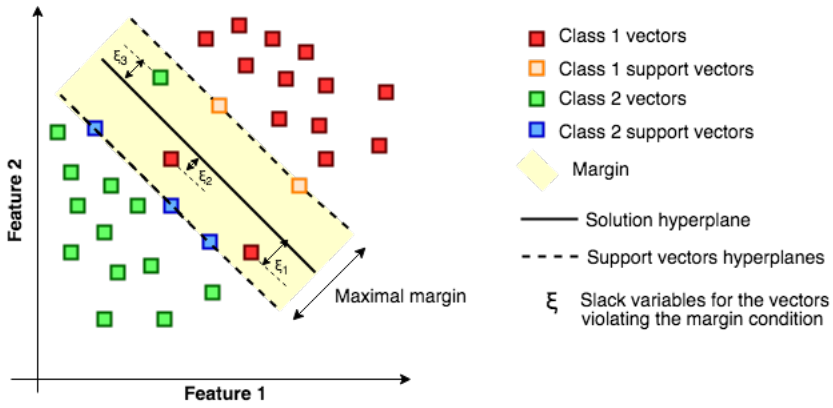
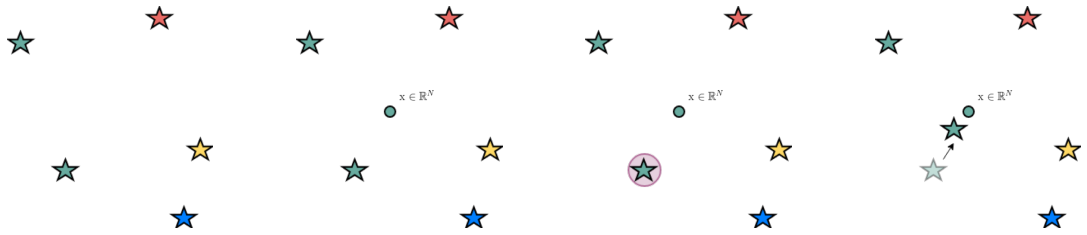


Figure 9: SVM model.

in the same feature space as the observed samples, facilitating a direct interpretation of the classifier. At least one prototype per class needs to be defined and in most of the cases their initialization is based on the mean value of the training samples belonging to each class.

In LVQ1, which is the first and basic training scheme, the classifier selects a set of labelled prototypes, and in combination with a distance measure, parameterize a nearest prototype classification (NPC), that performs a winner-takes-all decision. The position of this winner prototype is adapted, attracted if it correctly classifies the sample and repelled if not. After training, when an incoming unknown sample needs to be labelled, its distances to all the prototypes are computed and it is assign the label of the closest one. See Figure 10 for a graphical representation of the algorithm.

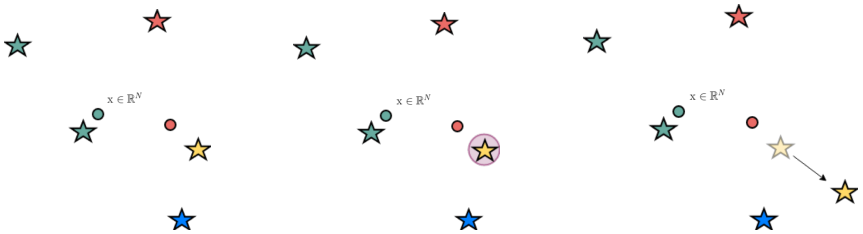


(a) Initialize prototype vectors for different classes.

(b) Present a new sample.

(c) Identiy the winner (closest prototype).

(d) Move the winner closer towards the data (same class).



(e) Present a new sample.

(f) Identify the winner.

(g) Move the winner away from the data (different class).

Figure 10: Illustration of the basic LVQ training scheme: LVQ1.

Regarding the training process, initially heuristic update schemes were used until 1996 when Sato and Yamada proposed a cost function based on stochastic gradient descent (SGD) to train the network ensuring convergence and establishing the Generalized LVQ (GLVQ) approach [38].

An important issue in LVQ and its extensions is the choice of an appropriate distance measure to evaluate the (dis-)similarity between prototypes and feature vectors in training and classification stages. Most frequently, predefined metrics are used. The by far most popular choice is the standard Euclidean distance, implicitly assuming that the data can be represented by isotropic clusters. However, several LVQ variants, as Relevance LVQ [39], have allowed to learn metric parameters (Λ) that are adapted to the data in the training phase, yielding a data-driven (parameterized) distance measure, $d^{\Lambda}(\cdot,\cdot)$, [40]. After training, the elements of Λ, also called relevance vector, reflect the importance of the different features for classification: the weights of non-informative or noisy dimensions are reduced, while discriminative features gain high weight values [41].

Generalized Matrix LVQ (GMLVQ) is an approach that focus on a matrix adaptation scheme for GLVQ by using a full matrix of adaptive relevances in the similarity measure among prototypes and features vectors, which accounts for pairwise correlations of features [29].

Let a training set of FDG-PET data be

$$\mathbb{D} = \left\{\mathbf{x}^{\mu}, y^{\mu}\right\}_{\mu=1}^{P},$$

where P refers to the number of N-dimensional feature vectors $x^{\mu} \in \mathbb{R}^N$ and $y^{\mu} \in \{1, 2, ...C\}$ are the labels that represent their class membership. Although in general, N and the number of samples are independent, in this case the dimension of the number of subject scores, extracted from the FDG-PET brain images, depends on the number of samples due to the SSM/PCA analysis.

As previously mentioned, an NPC needs to define a set of prototypes $\mathbf{w}^j \in \mathbb{R}^N$ carrying class labels c^j, $j = 1, 2, ...M$, together with an appropriate distance metric $d(\mathbf{w}, \mathbf{x})$.

Any feature vector $\mathbf{x}$ will be assigned to the class $y(\mathbf{x}) = c^L$ of the closest prototype c^L if

$$d\left(w^L, \mathbf{x}\right) < d\left(w^j, \mathbf{x}\right) \quad \text{for all} \quad j \neq L.$$

The use of the squared Euclidian distance $d_{Eu}(\mathbf{w}, \mathbf{x}) = (\mathbf{w} - \mathbf{x})^2 = \sum_{n=1}^{N} (w_n - x_n)^2$ is based on the intrinsic presupposition that features display similar magnitude in the data set that are of comparable importance for the classification. In GMLVQ, a adaptive distance measure is used, instead:

$$d_{\Lambda}(\mathbf{w}, \mathbf{x}) = (\mathbf{w} - \mathbf{x})^{\top} \Lambda (\mathbf{w} - \mathbf{x}) = \sum_{m,n=1}^{N} (w_m - x_m) \Lambda_{mn} (w_n - x_n),$$

where $\Lambda \in \mathbb{R}^{N \times N}$, called relevance matrix, is a full $N \times N$ matrix which accounts for correlations between the features and their importance for the classification.

In order to obtain valid distances (non-negative), Λ is forced to be symmetric and positive semi-definite by the parameterization $\Lambda = \Omega^{\top}\Omega$, where $\Omega \in \mathbb{R}^{N \times N}$ realizes a linear transformation of all data and prototypes to a space in which d_{Λ} corresponds to the squared Euclidean distance:

$$d_{\Lambda}(\mathbf{w}, \mathbf{x}) = (\mathbf{w} - \mathbf{x})^{\top} \Omega^{\top} \Omega (\mathbf{w} - \mathbf{x}) = [\Omega (\mathbf{w} - \mathbf{x})]^2.$$

In the following, we will use just one quadratic matrix Ω responsible of the global transformation of the data. Extensions have been suggested which employ individual local relevance matrices to each prototype or each class to measure the (dis-) similarity. Also, the use of rectangular matrices $\Omega \in R^{N \times M}$ with $M < N$ has been studied which corresponds to an intrinsic low-dimensional representation of the data.

During training, which consists on the minimization of a cost function (defined below) with respect to the model parameters, these are updated (learned) such that the distance from the closest correct prototype becomes smaller whereas the distance from the closest wrong prototype is increased.

For a given feature vector x^{μ} with class label y^{μ}, we denote by $\mathbf{w}_{\mu}^{J}$ the closest correct prototype, while $\mathbf{w}_{\mu}^{K}$ is the closest incorrect prototype:

$$\mathbf{w}_{\mu}^{J} = \underset{\mathbf{w}^{\mathbf{J}}}{\operatorname{argmin}} \left\{ d\left(\mathbf{w}^j, \mathbf{x}^{\mu}\right) | c^j = y^{\mu} \right\}_{j=1}^{M}.$$

$$\mathbf{w}_{\mu}^{K} = \underset{\mathbf{w^J}}{\operatorname{argmin}} \left\{ d\left(\mathbf{w}^{j}, \mathbf{x}^{\mu}\right) | c^{j} \neq y^{\mu} \right\}_{j=1}^{M}.$$

GMLVQ aims to optimize the prototypes and relevance matrix in the same process with respect to the cost function:

$$E_{GMLVQ}\left(\left\{\mathbf{w}^{j}\right\}_{j=1}^{M}, \Omega\right) = \sum_{\mu=1}^{P} \phi(\gamma_{\mu}) = \sum_{\mu=1}^{P} \phi\left[\frac{d(\mathbf{w}_{\mu}^{J}, \mathbf{x}^{\mu}) - d(\mathbf{w}_{\mu}^{K}, \mathbf{x}^{\mu})}{d(\mathbf{w}_{\mu}^{J}, \mathbf{x}^{\mu}) + d(\mathbf{w}_{\mu}^{K}, \mathbf{x}^{\mu})}\right],$$

where the term γ is called the relative difference distance, interpreted as a measure of confidence for prototype-based classification. By minimizing E_{GMLVQ} a solution is found with a small number of misclassifications in the training set. Then $d(\mathbf{w}_{\mu}^{J}, \mathbf{x}^{\mu})$ and $d(\mathbf{w}_{\mu}^{K}, \mathbf{x}^{\mu})$ represent the distances of the samples $\mathbf{x}$ to the respective closest correct and wrong prototypes. The monotonically increasing function ϕ determines the active region of the algorithm. It is usually a sigmoidal function $\phi(x) = \frac{1}{(1+e^{x})}$ or the identity $\phi(x) = x$. Being the latter the one used in this project.

Frequently, the relevance matrix, Λ, is normalized to prevent the learning algorithm from degeneration. In our case it was normalized such that

$$Tr(\Lambda) = \sum_{m=1}^{N} \Lambda_{mn} = \sum_{m,n=1}^{N} \Omega_{mn}^{2} = 1.$$

E. Classifier validation

We obtain SVM and GMLVQ classifiers over subject scores of HC and subjects with known types of neurodegenerative disease as PD and AD.

In the case of SVM, we use the Matlab R2016a "fitcsvm" and "predict" functions for training and testing, respectively, with default parameters and a linear kernel. The "fitcsvm" returns an SVM classifier which can be used for classification of new data samples by "predict" function. For GMLVQ, we use the gmlvq-toolbox, see [42] for more details, to train and test GMLVQ classifier through "run_single" and "classify_gmlvq" functions, respectively. As inputs parameters, we specify the performance of z-score transformation based on the training set and the realization of the batch gradient descent minimization with 60 steps. The rest of parameters have been set to their default values as suggested in the toolbox.

The classifiers' performance is determined by 10 repetitions of 10-fold cross validation, where in each round the set of samples is randomly partitioned into 10 equal sized subsets. Among these subsets, just one is retained as the validation data for testing the model of the classifier, whereas the remaining 9 subsets are used as training data. This process is repeated, for each run, 10 times (number of folds), using each of the 10 subsets once as the validation data. By performing multiple runs of k-fold cross validation we achieve a more accurate estimation of the model quality by reducing the influence of lucky or unlucky set composition. The validation results are obtained as the average over the runs and folds.

IV. EXPERIMENTS

In this chapter, we will establish the experimental framework of the project by explaining the different experiments that have been carried out.

A. Binary classification of diseases by center

Our objective is to distinguish among all the different types of subjects by centers taking into account pairs of classes. For that, in the phase of feature extraction we have applied three different masks in order to see how that can affect the classification. First, the subject scores are obtained just by using the mask of PDRP. Then, we use the mask of ADRP and finally, the subject scores are defined as a combination of applying both masks and concatenating the corresponding features. We have used SVM and GMLVQ as classifiers.

Next, we specify the experiments and settings done in this section, see Table II.

- **E1**: Classify HC and PD subjects from UGOSM.
- **E2**: Classify HC and AD subjects from UGOSM.
- **E3**: Classify PD and AD subjects from UGOSM.
- **E4**: Classify HC and PD subjects from UMCG.

- **E5**: Classify HC and AD subjects from UMCG.
- **E6**: Classify PD and AD subjects from UMCG.
- **E7**: Classify HC and PD subjects from CUN.

Experiment	Source	Subjects	Class 1	Class 2
E1	UGOSM	44 HC 58 PD	HC	PD
E2		44 HC 55 AD	HC	AD
E3		58 PD 55 AD	PD	AD
E4	UMCG	19 HC 20 PD	HC	PD
E5		19 HC 21 AD	HC	AD
E6		20 PD 21 AD	PD	AD
E7	CUN	19 HC 68 PD	HC	PD

Table II: Binary classification of diseases by center.

B. Ternary classification of diseases by center

Our objective is to distinguish among all the different types of subjects by centers. Unlike the previous experiments, in this occasion we use the three different classes, available in some centers, instead of just pairs of them. As it was previously mentioned in section IV-A, in the phase of feature extraction we have applied three different masks in order to see how it can affect the classification. The experiments have been done with SVM and GMLVQ classifiers.

Next, we specify the experiments and settings done in this section, see Table III .

- **E8**: Classify HC, PD and AD subjects from UGOSM.
- **E9**: Classify HC, PD and AD subjects from UMCG.

Experiment	Source	Subjects	Class 1	Class 2	Class 3
E8	UGOSM	44 HC 58 PD 55 AD	HC	PD	AD
E9	UMCG	19 HC 20 PD 21 AD			

Table III: Ternary classification of diseases by center.

C. Classification of centers by diseases

We aim to distinguish the different centers for a specific type of subjects taking into account pairs of centers. As it was explained in section IV-A, in the phase of feature extraction we have applied three different masks in order to see how this can affect to the performance of the classifiers. We have used both SVM and GMLVQ as classifiers.

Next, we specify the experiments and settings done in this section, see Table IV .

- **E10**: Classify PD subjects from CUN and UGOSM.
- **E11**: Classify PD subjects from CUN and UMCG.
- **E12**: Classify PD subjects from UMCG and UGOSM.
- **E13**: Classify AD subjects from UMCG and UGOSM.
- **E14**: Classify HC subjects from CUN and UMCG.
- **E15**: Classify HC subjects from CUN and UGOSM.
- **E16**: Classify HC subjects from UMCG and UGOSM.

Experiment	Disease	Sources	Subjects	Class 1	Class 2
E10	PD	CUN and UGOSM	68 CUN 58 UGOSM	CUN	UGOSM
E11		CUN and UMCG	68 CUN 20 UMCG	CUN	UMCG
E12		UMCG and UGOSM	20 UMCG 58 UGOSM	UMCG	UGOSM
E13	AD	UMCG and UGOSM	21 UMCG 55 UGOSM	UMCG	UGOSM
E14	HC	CUN and UMCG	19 CUN 19 UMCG	CUN	UMCG
E15		CUN and UGOSM	19 CUN 44 UGOSM	CUN	UGOSM
E16		UMCG and UGOSM	19 UMCG 44 UGOSM	UMCG	UGOSM

Table IV: Classification of centers by diseases.

D. *Binary classification of diseases without distinction of centers*

Our goal is to differentiate within the types of subjects without distinguishing from which center the subjects come. Again performing a binary classification. For that, in the phase of feature extraction, as explained in section IV-A, we have applied three different masks in order to see how it can affect the classification. We have used both SVM and GMLVQ as classifiers.

Next, we specify the experiments and settings done in this section, see Table V.

- **E17**: Classify HC and PD subjects from CUN and UGOSM.
- **E18**: Classify HC and PD subjects from CUN and UMCG.
- **E19**: Classify HC and PD subjects from UMCG and UGOSM.
- **E20**: Classify HC and AD subjects from UMCG and UGOSM.
- **E21**: Classify PD and AD subjects from UMCG and UGOSM.

Experiment	Sources	Subjects	Class 1	Class 2
E17	CUN and UGOSM	(19 HC 68 PD) CUN (44 HC 58 PD) UGOSM	HC	PD
E18	CUN and UMCG	(19 HC 68 PD) CUN (19 HC 20 PD) UMCG	HC	PD
E19	UMCG and UGOSM	(19 HC 20 PD) UMCG (44 HC 58 PD) UGOSM	HC	PD
E20	UMCG and UGOSM	(19 HC 21 AD) UMCG (44 HC 55 AD) UGOSM	HC	AD
E21	UMCG and UGOSM	(20 PD 21 AD) UMCG (58 PD 55 AD) UMCG	PD	AD

Table V: Binary classification of diseases without distinction of centers.

E. *Ternary classification of diseases without distinction of centers*

Our goal is to differentiate HC and PD subjects from the three centers. In the phase of feature extraction we have applied three different masks, see section IV-A, in order to see how it can affect the classification. We have used both SVM and GMLVQ as classifiers.

- **E22**: Classify HC and PD subjects from CUN, UGOSM and UMCG.

Next, we specify the experiments and settings done in this section, see Table **??**.

Experiment	Sources	Subjects	Class 1	Class 2
E22	CUN, UGOSM and UMCG	(19 HC 68 PD) CUN (44 HC 58 PD) UGOSM (19 HC 20 PD) UMCG	HC	PD

Table VI: Ternary classification of diseases without distinction of centers.

F. *Classifier transfer of binary classification of diseases among centers*

We want to know the ability to classify subjects of a specific center (testing center) when the classifier used has been obtained from data from another center (training center). It can be useful because no matters where the subjects origin the affected regions of a particular disease should be similar. In the phase of feature extraction we have applied three different masks in order to see how it can affect to the classification. First, the subject scores are obtained just by using the mask of PDRP. Then, we get them through the use, in this case, of the mask of ADRP and finally, the subject scores are defined as a combination of applying both masks. We have used SVM and GMLVQ as classifiers.

Next, we specify the experiments and settings done in this section, see Table VII.

- **E23**: Classify HC and PD subjects from UGOSM after training the classifier with HC and PD CUN subjects.
- **E24**: Classify HC and PD subjects from CUN after training the classifier with HC and PD UGOSM subjects.
- **E25**: Classify HC and PD subjects from UMCG after training the classifier with HC and PD subjects from CUN.
- **E26**: Classify HC and PD CUN subjects after training the classifier with HC and PD subjects from UMCG.
- **E27**: Classify HC and PD subjects from UMCG after training the classifier with HC and PD UGOSM subjects.
- **E28**: Classify HC and PD subjects from UGOSM after training the classifier with HC and PD subjects from UMCG.
- **E29**: Classify HC and AD subjects from UMCG after training the classifier with HC and AD UGOSM subjects.
- **E30**: Classify HC and AD subjects from UGOSM after training the classifier with HC and AD subjects from UMCG.
- **E31**: Classify PD and AD subjects from UMCG after training the classifier with PD and AD UGOSM subjects.
- **E32**: Classify PD and AD subjects from UGOSM after training the classifier with PD and AD subjects from UMCG.

Experiment	Sources	Subjects	Class 1	Class 2	Train	Test
E23	CUN and UGOSM	(19 HC 68 PD) CUN (44 HC 58 PD) UGOSM	HC	PD	CUN	UGOSM
E24					UGOSM	CUN
E25	CUN and UMCG	(19 HC 68 PD) CUN (19 HC 20 PD) UMCG			CUN	UMCG
E26					UMCG	CUN
E27	UGOSM and UMCG	(19 HC 20 PD) UMCG (44 HC 58 PD) UGOSM			UGOSM	UMCG
E28					UMCG	UGOSM
E29		(44 HC 55 AD) UGOSM (19 HC 21 AD) UMCG	HC	AD	UGOSM	UMCG
E30					UMCG	UGOSM
E31		(58 PD 55AD) UGOSM (20 PD 21 AD) UMCG	PD	AD	UGOSM	UMCG
E32					UMCG	UGOSM

Table VII: Classifier transfer of binary classification of diseases among centers.

V. RESULTS

In this chapter, we will explain the performance evaluation procedure and analyze the results. For that, first we will describe the efficiency measures that have been used. Then, we expose the results for all the experiments explained in section IV, in both a numerical and a graphical way. Finally, we will discuss about the obtained results from the performance of the SVM and GMLVQ classifiers.

A. Objective performance measures

The performance measures used in this project to evaluate the classification methods that have been used, are objective measures based on the idea of a good classification of the different classes. For this it is necessary to compare the results obtained by each classifier with the corresponding label.

When we compare the label of a classifier with its corresponding true label, we can divide the classification into four possible types: True Positive, True Negative, False Positive and True Negative. These terms are explained in Table VIII.

TP	Samples belonging to class 1 and that have been classified as class 1.
TN	Samples belonging to class 2 and that have been classified as class 2.
FP	Samples belonging to class 1 and that have been classified as class 2.
FN	Samples belonging to class 2 and that have been classified as class 1.

Table VIII: Definition of the four possible types of samples classification.

Most of the evaluation measures are based on combinations of these terms. Next, we describe the indicators used in this project for the evaluation of the results.

- ***Sensitivity***: measures the proportion of positives that are correctly identified as such.

$$Sensitivity = \frac{TP}{TP + FN}.$$

- ***Specificity***: measures the proportion of negatives that are correctly identified as such.

$$Specificity = \frac{TN}{FP + TN}.$$

- ***The AUC (area under the curve) of the ROC***: provides a useful measure of quality which is independent of an individual choice of the working point, see [43] also for an intuitive statistical interpretation of the AUC.

B. Quantitative results

In this section we will use the objective measures of performance already introduced to present, in a numerical way, the results obtained by the two classifiers.

- **Binary classification of subjects by center**

	Classifier	Sens. (%)	Spec. (%)	AUC
PD pattern	GMLVQ	68.59	81.07	0.851
	SVM	68.89	78.11	0.801
AD pattern	GMLVQ	73.45	83.58	0.805
	SVM	69.26	76.54	0.786
Combination AD and PD pattern	GMLVQ	72.05	81.83	0.835
	SVM	70.62	78.56	0.784

(a) E1: Classify HC and PD subjects.

	Classifier	Sens. (%)	Spec. (%)	AUC
PD pattern	GMLVQ	88.28	95.58	0.975
	SVM	94.84	95.14	0.983
AD pattern	GMLVQ	91.41	94.65	0.981
	SVM	96.30	97.02	0.999
Combination AD and PD pattern	GMLVQ	93.50	95.17	0.995
	SVM	94.36	97.85	0.998

(b) E2: Classify HC and AD subjects.

	Classifier	Sens. (%)	Spec. (%)	AUC
PD pattern	GMLVQ	82.86	83.05	0.924
	SVM	84.35	84.69	0.919
AD pattern	GMLVQ	94.75	94.27	0.972
	SVM	93.38	90.57	0.973
Combination AD and PD pattern	GMLVQ	87.79	86.79	0.935
	SVM	92.87	90.25	0.971

(c) E3: Classify PD and AD subjects.

Table IX: Classification among all the different types of subjects from UGOSM, taking into account pairs of classes. The results are obtained for both SVM and GMLVQ classifiers, using three different types of patterns. For more details see Table II.

	Classifier	Sens. (%)	Spec. (%)	AUC
PD pattern	GMLVQ	70.37	77.95	0.660
	SVM	82.82	92.29	0.890
AD pattern	GMLVQ	64.67	76.16	0.758
	SVM	79.04	84.63	0.853
Combination AD and PD pattern	GMLVQ	66.33	77.16	0.778
	SVM	86.25	92.02	0.870

(a) E4: Classify HC and PD subjects.

	Classifier	Sens. (%)	Spec. (%)	AUC
PD pattern	GMLVQ	82.50	95.00	0.908
	SVM	87.50	92.11	0.918
AD pattern	GMLVQ	79.33	90.64	0.898
	SVM	78.65	83.33	0.909
Combination AD and PD pattern	GMLVQ	80.33	90.14	0.883
	SVM	85.50	94.44	0.949

(b) E5: Classify HC and AD subjects.

	Classifier	Sens. (%)	Spec. (%)	AUC
PD pattern	GMLVQ	74.33	67.12	0.693
	SVM	73.78	80.22	0.829
AD pattern	GMLVQ	77.83	68.35	0.775
	SVM	79.76	80.65	0.857
Combination AD and PD pattern	GMLVQ	71.83	70.83	0.770
	SVM	74.58	82.42	0.870

(c) E6: Classify PD and AD subjects.

Table X: Classification among all the different types of subjects from UMCG, taking into account pairs of classes. The results are obtained for both SVM and GMLVQ classifiers, using three different types of patterns. For more details see Table II.

	Classifier	Sens. (%)	Spec. (%)	AUC
PD pattern	GMLVQ	85.50	95.71	0.910
	SVM	81.70	93.88	0.933
AD pattern	GMLVQ	80.83	98.42	0.990
	SVM	79.06	93.54	0.942
Combination AD and PD pattern	GMLVQ	75,00	95,86	0.962
	SVM	83.92	94.93	0.961

Table XI: E7: Classification among all the different types of subjects from CUN, taking into account pairs of classes. In this case, we have just HC and PD subjects. The results are obtained for both SVM and GMLVQ classifiers, using three different types of patterns. For more details see Table II.

- **Ternary classification of diseases by center**

In this case, the sensitivity, specificity and the auc represents the ability to classify HC vs the PD and AD diseases.

	Sens. (%)	Spec. (%)	AUC
PD pattern	65.67	70.41	0.768
AD pattern	71.60	85.31	0.802
Combination AD and PD pattern	69.89	81.71	0.792

(a) E8: Classify HC, PD and AD subjects from UGOSM.

	Sens. (%)	Spec. (%)	AUC
PD pattern	77.64	68.78	0.608
AD pattern	74.81	88.64	0.708
Combination AD and PD pattern	78.70	80.59	0.707

(b) E9: Classify HC, PD and AD subjects from UMCG.

Table XII: Ternary classification for the centers that have more than two different types of subjects. The results are obtained just for GMLVQ classifier, using three different types of patterns.For more details see Table III.

- **Classification of centers by diseases**

	Classifier	Sens. (%)	Spec. (%)	AUC
PD pattern	GMLVQ	95.80	93.90	0.987
	SVM	91.31	88.99	0.970
AD pattern	GMLVQ	97.15	98.36	0.991
	SVM	95.99	95.64	0.978
Combination AD and PD pattern	GMLVQ	97.50	98.43	0.990
	SVM	96.63	95.06	0.984

(a) E10: Classify PD subjects from CUN and UGOSM.

	Classifier	Sens. (%)	Spec. (%)	AUC
PD pattern	GMLVQ	97.63	100	1
	SVM	95.74	96.30	0.990
AD pattern	GMLVQ	97.50	96.67	0.999
	SVM	98.34	96.33	0.999
Combination AD and PD pattern	GMLVQ	100	100	1
	SVM	97.98	96.60	0.998

(b) E11: Classify PD subjects from CUN and UMCG.

	Classifier	Sens. (%)	Spec. (%)	AUC
PD pattern	GMLVQ	87.88	97.66	0.989
	SVM	95.75	95.91	0.975
AD pattern	GMLVQ	97.00	97.29	1
	SVM	99.17	96.50	0.989
Combination AD and PD pattern	GMLVQ	99.63	94.73	0.984
	SVM	99.16	98.44	0.994

(c) E12: Classify PD subjects from UMCG and UGOSM.

Table XIII: Classification of centers for a specific type of subjects. In this case, we focus on PD subjects of two different centers. The results are obtained for both SVM and GMLVQ classifiers, using three different types of patterns. For more details see Table IV.

	Classifier	Sens. (%)	Spec. (%)	AUC
PD pattern	GMLVQ	84.98	91.43	0.887
	SVM	83.94	91.85	0.917
AD pattern	GMLVQ	88.83	96.86	0.960
	SVM	91.92	96.16	0.968
Combination AD and PD pattern	GMLVQ	92.67	96.69	0.987
	SVM	88.10	95.38	0.974

Table XIV: E13: Classify AD subjects from UMCG and UGOSM. The results are obtained for both SVM and GMLVQ classifiers, using three different types of patterns. For more details see Table IV.

	Classifier	Sens. (%)	Spec. (%)	AUC
PD pattern	GMLVQ	100	100	1
	SVM	97.67	100	1
AD pattern	GMLVQ	100	97.00	1
	SVM	100	100	1
Combination AD and PD pattern	GMLVQ	100	100	1
	SVM	100	100	1

(a) E14: Classify HC subjects from CUN and UMCG.

	Classifier	Sens. (%)	Spec. (%)	AUC
PD pattern	GMLVQ	92.42	96.21	0.982
	SVM	99.66	97.81	0.999
AD pattern	GMLVQ	100	100	1
	SVM	100	100	1
Combination AD and PD pattern	GMLVQ	100	99.83	1
	SVM	100	99.83	1

(b) E15: Classify HC subjects from CUN and UGOSM.

	Classifier	Sens. (%)	Spec. (%)	AUC
PD pattern	GMLVQ	79.33	91.33	0.939
	SVM	95.58	95.90	0.972
AD pattern	GMLVQ	94.67	94.10	0.959
	SVM	93.37	95.46	0.996
Combination AD and PD pattern	GMLVQ	96.67	96.13	0.987
	SVM	95.02	93.43	0.988

(c) E16: Classify HC subjects from UMCG and UGOSM.

Table XV: Classification of centers for a specific type of subjects. In this case, we focus on HC subjects of two different centers. The results are obtained for both SVM and GMLVQ classifiers, using three different types of patterns.For more details see Table IV.

- **Binary classification of diseases without distinction of centers**

	Classifier	Sens. (%)	Spec. (%)	AUC
PD pattern	GMLVQ	71.96	87.93	0.865
	SVM	71.08	85.89	0.882
AD pattern	GMLVQ	72.43	89.53	0.873
	SVM	68.12	84.77	0.849
Combination AD and PD pattern	GMLVQ	68.07	87.20	0.876
	SVM	71.14	86.58	0.874

(a) E17: Classify HC and PD subjects from CUN and UGOSM.

	Classifier	Sens. (%)	Spec. (%)	AUC
PD pattern	GMLVQ	79.02	87.23	0.904
	SVM	79.92	89.07	0.893
AD pattern	GMLVQ	81.02	92.97	0.945
	SVM	75.10	88.95	0.903
Combination AD and PD pattern	GMLVQ	72.27	91.24	0.927
	SVM	75.90	89.37	0.898

(b) E18: Classify HC and PD subjects from CUN and UMCG.

	Classifier	Sens. (%)	Spec. (%)	AUC
PD pattern	GMLVQ	76.32	82.81	0.879
	SVM	79.47	86.33	0.882
AD pattern	GMLVQ	77.14	82.72	0.862
	SVM	76.32	81.63	0.854
Combination AD and PD pattern	GMLVQ	72.48	82.50	0.842
	SVM	72.88	77.90	0.806

(c) E19: Classify HC and PD subjects from UMCG and UGOSM.

Table XVI: Classification of HC and PD subjects without distinguishing from the center of origin. The results are obtained for both SVM and GMLVQ classifiers, using three different types of patterns. For more details see Table V.

	Classifier	Sens. (%)	Spec. (%)	AUC
PD pattern	GMLVQ	87.48	96.04	0.983
	SVM	90.35	93.78	0.975
AD pattern	GMLVQ	85.48	92.96	0.959
	SVM	84.77	91.79	0.950
Combination AD and PD pattern	GMLVQ	82.89	92.02	0.936
	SVM	90.75	95.05	0.968

Table XVII: E20: Classify HC and AD subjects without distinguishing that they come from UMCG and UGOSM. The results are obtained for both SVM and GMLVQ classifiers, using three different types of patterns. For more details see Table V.

	Classifier	Sens. (%)	Spec. (%)	AUC
PD pattern	GMLVQ	83.31	79.68	0.870
	SVM	77.79	78.20	0.857
AD pattern	GMLVQ	81.95	85.42	0.913
	SVM	86.23	87.81	0.927
Combination AD and PD pattern	GMLVQ	83.72	85.23	0.895
	SVM	78.87	80.18	0.855

Table XVIII: E21: Classify PD and AD subjects without distinguishing that they come from UMCG and UGOSM. The results are obtained for both SVM and GMLVQ classifiers, using three different types of patterns. For more details see Table V.

- **Ternary classification of diseases without distinction of centers**

	Classifier	Sens. (%)	Spec. (%)	AUC
PD pattern	GMLVQ	79.93	87.43	0.908
	SVM	76.18	86.64	0.898
AD pattern	GMLVQ	69.37	85.55	0.868
	SVM	73.39	85.05	0.877
Combination AD and PD pattern	GMLVQ	72.12	86.78	0.884
	SVM	73.40	84.17	0.867

Table XIX: E22: Classify HC and PD subjects without distinguishing that they come from CUN, UMCG or UGOSM. The results are obtained for both SVM and GMLVQ classifiers, using three different types of patterns. For more details see Table VI.

- **Classifier transfer of binary classification of diseases among centers**

Here we study the ability to classify subjects of a specific center (testing center) when the classifier used has been obtained from data from another center (training center). The results are obtained for both SVM and GMLVQ classifiers, using three different types of patterns. For more details see Table VII.

	Classifier	Sens. (%)	Spec. (%)	AUC
PD pattern	GMLVQ	37.50	56.38	0.556
	SVM	0	53.19	0.534
AD pattern	GMLVQ	0	54.64	0.448
	SVM	0	53.19	0.410
Combination AD and PD pattern	GMLVQ	0	54.64	0.476
	SVM	0	55.10	0.443

Table XX: E23: Classify HC and PD subjects from UGOSM after training the classifier with HC and PD CUN subjects.

	Classifier	Sens. (%)	Spec. (%)	AUC
PD pattern	GMLVQ	24.05	100	0.709
	SVM	25.00	100	0.618
AD pattern	GMLVQ	24.05	100	0.673
	SVM	25	93.33	0.511
Combination AD and PD pattern	GMLVQ	26.03	100	0.644
	SVM	22.22	80.00	0.524

Table XXI: E24: Classify HC and PD subjects from CUN after training the classifier with HC and PD UGOSM subjects.

	Classifier	Sens. (%)	Spec. (%)	AUC
PD pattern	GMLVQ	60.87	68.75	0.703
	SVM	58.33	66.67	0.718
AD pattern	GMLVQ	60.87	68.75	0.703
	SVM	64.00	78.57	0.761
Combination AD and PD pattern	GMLVQ	59.09	64.71	0.684
	SVM	41.18	45.45	0.450

Table XXII: E25: Classify HC and PD subjects from UMCG after training the classifier with HC and PD subjects from CUN.

	Classifier	Sens. (%)	Spec. (%)	AUC
PD pattern	GMLVQ	26.47	94.74	0.824
	SVM	27.94	100	0.860
AD pattern	GMLVQ	27.14	100	0.915
	SVM	27.94	100	0.897
Combination AD and PD pattern	GMLVQ	26.39	100	0.922
	SVM	27.54	100	0.906

Table XXIII: E26: Classify HC and PD CUN subjects after training the classifier with HC and PD subjects from UMCG.

	Classifier	Sens. (%)	Spec. (%)	AUC
PD pattern	GMLVQ	43.75	28.57	0.361
	SVM	45.45	33.33	0.432
AD pattern	GMLVQ	53.13	71.43	0.532
	SVM	54.55	58.82	0.576
Combination AD and PD pattern	GMLVQ	51.52	66.67	0.443
	SVM	51.61	62.50	0.484

Table XXIV: E27: Classify HC and PD subjects from UMCG after training the classifier with HC and PD UGOSM subjects.

	Classifier	Sens. (%)	Spec. (%)	AUC
PD pattern	GMLVQ	38.98	51.16	0.477
	SVM	36.49	39.29	0.383
AD pattern	GMLVQ	51.35	78.57	0.786
	SVM	44.62	59.46	0.518
Combination AD and PD pattern	GMLVQ	50.88	66.67	0.639
	SVM	41.43	53.13	0.515

Table XXV: E28: Classify HC and PD subjects from UGOSM after training the classifier with HC and PD subjects from UMCG.

	Classifier	Sens. (%)	Spec. (%)	AUC
PD pattern	GMLVQ	45.45	51.72	0.516
	SVM	47.06	52.17	0.434
AD pattern	GMLVQ	50	57.14	0.459
	SVM	30.77	44.44	0.376
Combination AD and PD pattern	GMLVQ	38.89	45.45	0.414
	SVM	50.00	54.55	0.441

Table XXVI: E29: Classify HC and AD subjects from UMCG after training the classifier with HC and AD UGOSM subjects.

	Classifier	Sens. (%)	Spec. (%)	AUC
PD pattern	GMLVQ	49.38	77.78	0.639
	SVM	45.88	64.29	0.537
AD pattern	GMLVQ	46.07	70	0.636
	SVM	46.51	69.23	0.683
Combination AD and PD pattern	GMLVQ	47.19	80.00	0.655
	SVM	46.67	77.78	0.654

Table XXVII: E30: Classify HC and AD subjects from UGOSM after training the classifier with HC and AD subjects from UMCG.

	Classifier	Sens. (%)	Spec. (%)	AUC
PD pattern	GMLVQ	51.35	75.00	0.693
	SVM	48.48	50.00	0.563
AD pattern	GMLVQ	47.22	40.00	0.605
	SVM	47.22	40.00	0.494
Combination AD and PD pattern	GMLVQ	51.35	75.00	0.635
	SVM	48.72	50.00	0.514

Table XXVIII: E31: Classify PD and AD subjects from UMCG after training the classifier with PD and AD UGOSM subjects.

	Classifier	Sens. (%)	Spec. (%)	AUC
PD pattern	GMLVQ	51.11	47.83	0.507
	SVM	40.00	44.00	0.467
AD pattern	GMLVQ	52.38	62.50	0.608
	SVM	51.52	50.00	0.586
Combination AD and PD pattern	GMLVQ	52.88	66.67	0.589
	SVM	52.58	56.25	0.556

Table XXIX: E32: Classify PD and AD subjects from UGOSM after training the classifier with PD and AD subjects from UMCG

C. Graphical results

In this section we illustrate, from a graphical point of view, the behavior of the classifiers for the set of experiments that have been performed.

- **Binary classification of diseases by center**

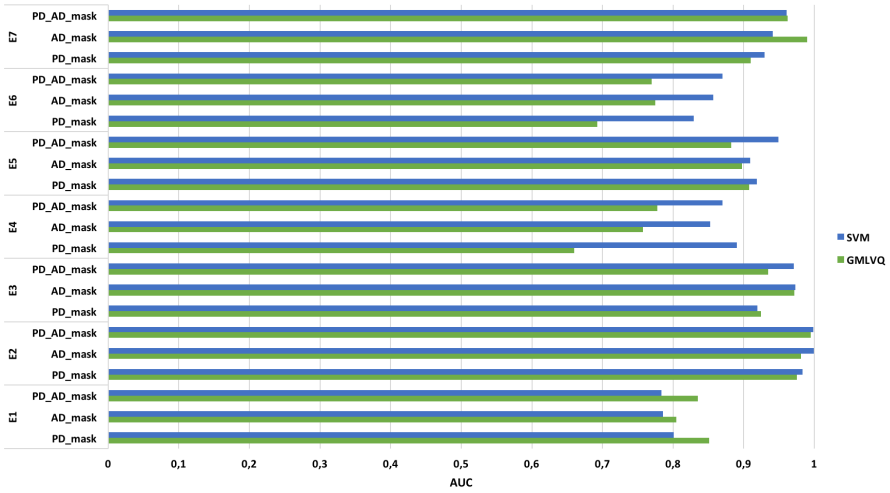
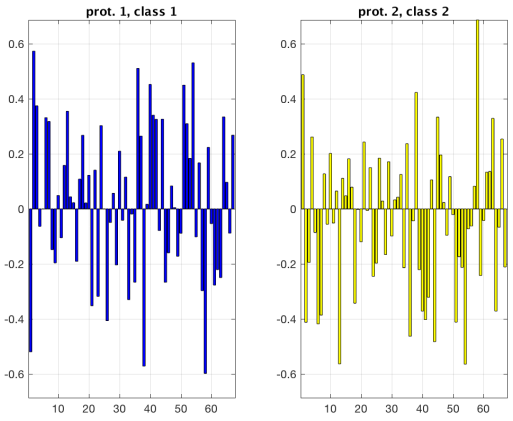
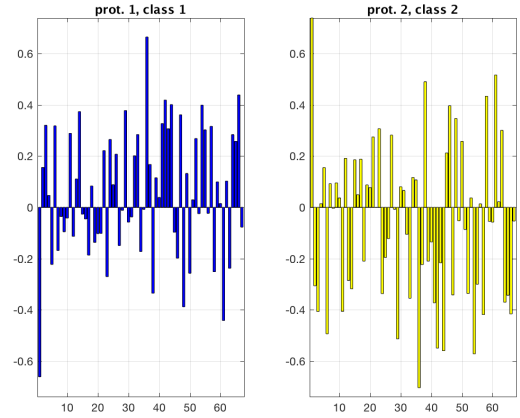


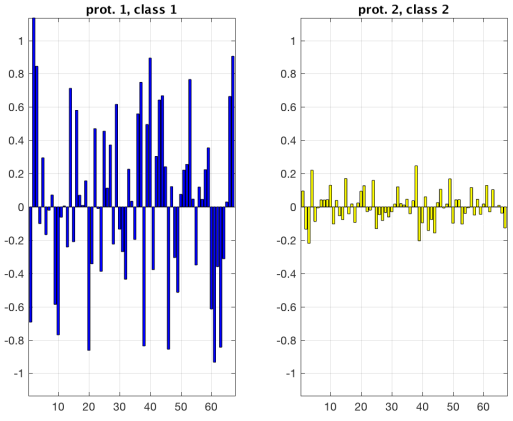
Figure 11: This graph was plotted according to the Tables IX-XI. It shows the AUC from the experiments, E1-E7, described in the Table II.



(a) E1: Classify HC (class 1) and PD (class 2) from UGOSM.

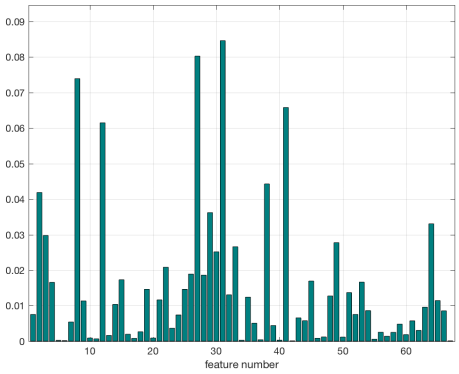


(b) E4: Classify HC (class 1) and PD (class 2) from UMCG.

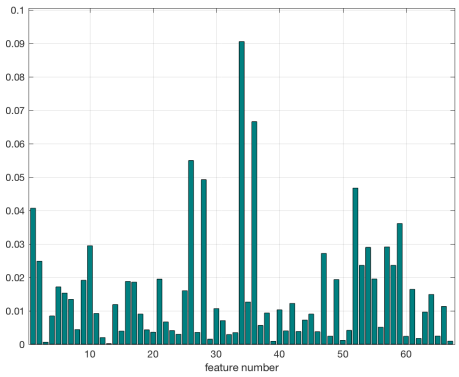


(c) E7: Classify HC (class 1) and PD (class 2) from CUN.

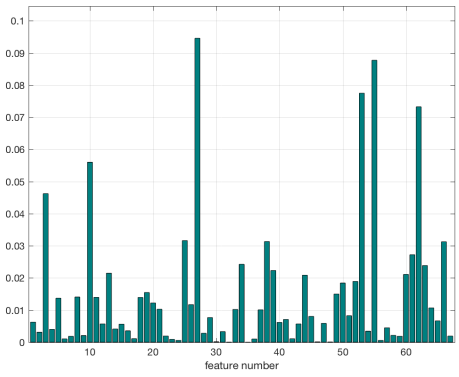
Figure 12: Example of comparison among the prototypes obtained for the experiments E1, E4 and E7, using GMLVQ and applying the combination of both PD and AD pattern.



(a) E1: Classify HC (class 1) and PD (class 2) from UGOSM.



(b) E4: Classify HC (class 1) and PD (class 2) from UMCG.



(c) E7: Classify HC (class 1) and PD (class 2) from CUN.

Figure 13: Example of comparison among the diagonal of relevance matrices obtained for the experiments E1, E4 and E7, using GMLVQ and applying the combination of both PD and AD pattern.

- **Ternary classification of diseases by center**

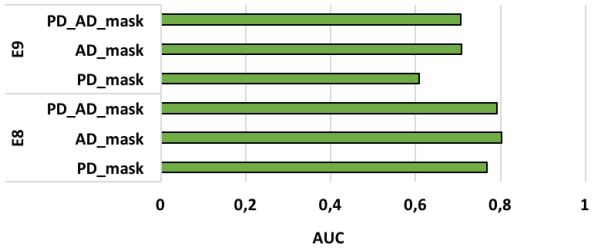


Figure 14: This graph was plotted according to the Table XII. It shows the AUC from the experiments, E8-E9, described in the Table III.

- **Classification of centers by diseases**

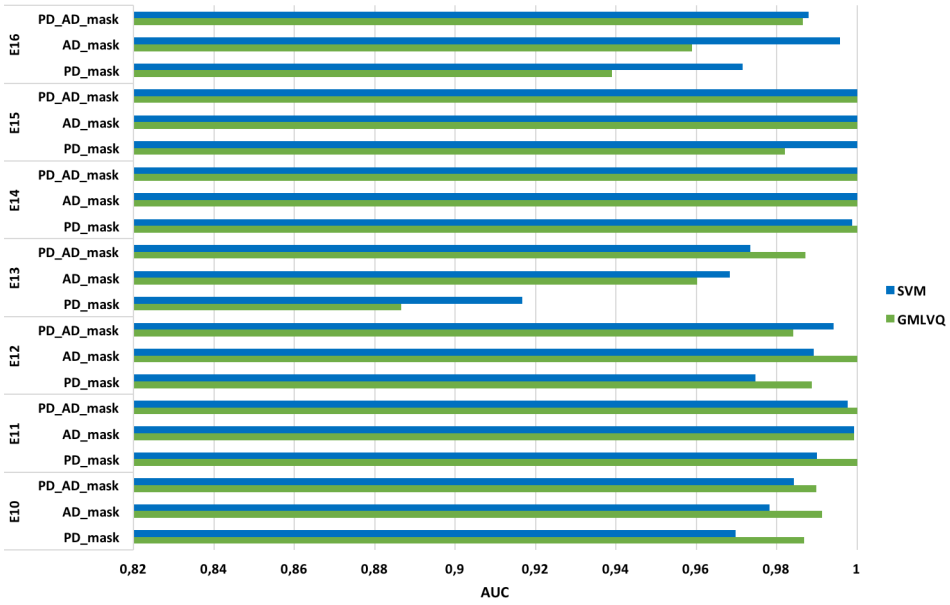


Figure 15: This graph was plotted according to the Tables XIII-XV. It shows the AUC from the experiments, E10-E16, described in the Table IV.

- **Binary classification of diseases without distinction of centers**

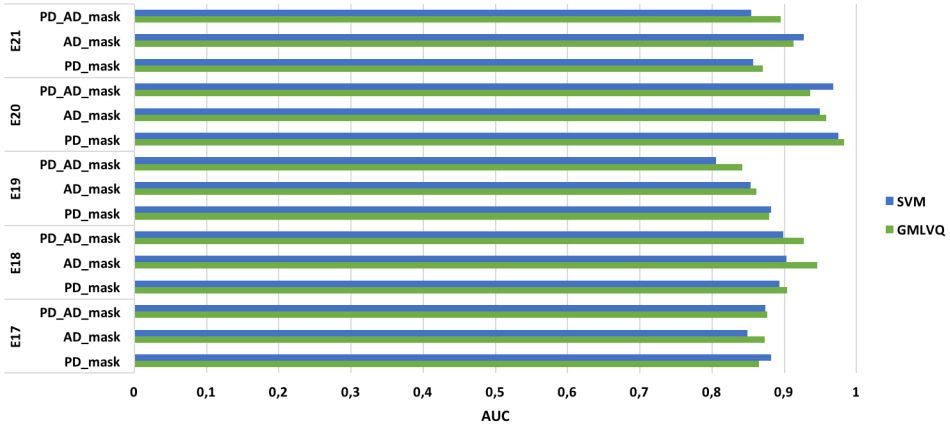


Figure 16: This graph was plotted according to the Tables XVI-XVIII. It shows the AUC from the experiments, E17-E21, described in the Table V.

- **Ternary classification of diseases without distinction of centers**

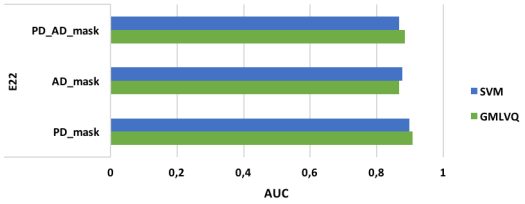


Figure 17: This graph was plotted according to the Table XIX. It shows the AUC from the experiment E22, described in the Table VI.

- **Classifier transfer of binary classification of diseases among centers**

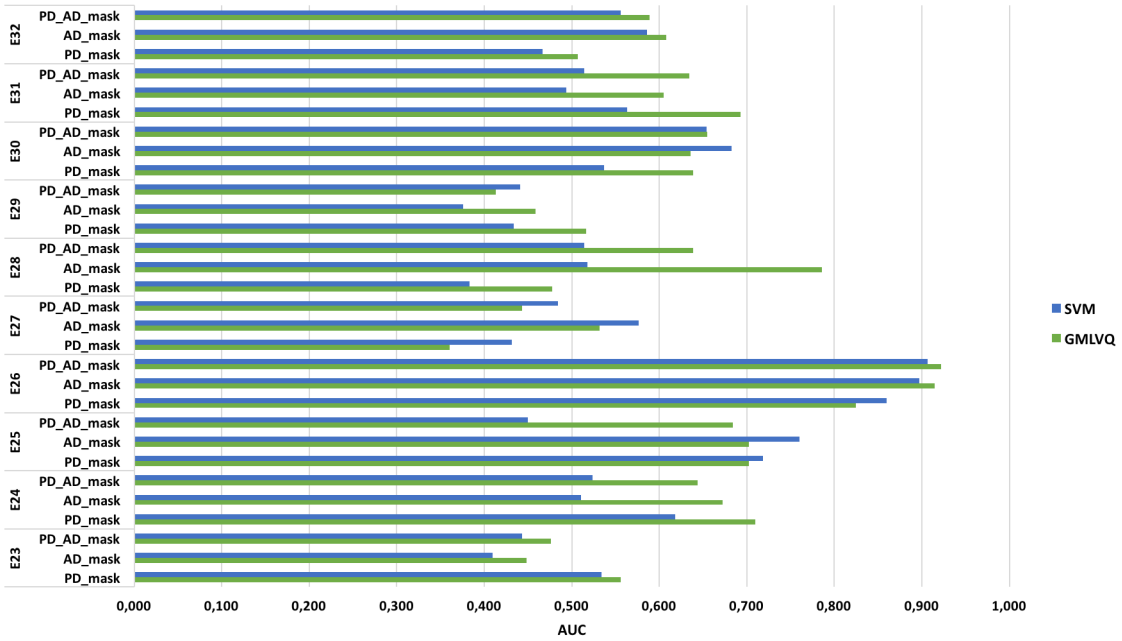


Figure 18: This graph was plotted according to the Tables XX-XXIX. It shows the AUC from the experiments E23-E32, described in the Table VII.

D. Discussion

From the experiments described in the Tables II-V, the performance of which is given in Tables IX-XVIII, we can conclude that the different subjects of the centers with whom we have worked are clearly distinguishable, both when it comes to distinguish diseases by centers and to discriminate the origin center of the subjects. Also we can observe from the results that the measures of performance do not point to a single best classifier.

Regarding the performance of the classifiers, from the experiments of binary classification of diseases by center, see Table II and Figure 11, we can see that the SVM classifier seems to provide a better performance than GMLVQ. While in the case of the experiments where we compare the centers by diseases and the ones where we classify the diseases without distinction of the different centers, see Figures 15-16, we have found that it is the GMLVQ classifier which slightly overcomes, or equals the performance of the SVM classifier. So, we find that there is not a significant difference among the performance of the GMLVQ and SVM classifiers. Both classifiers have an average standard deviation of 0.08 when computing the AUC.

We observe from the results in Tables IX-XII,XVI-XIX that our initial objective, which was based on distinguishing the subjects condition (HC, PD or AD) has been achieved, both when it comes to a specific center and when the data from the three centers (CUN, UMCG and UGOSM) were mixed together.

These results can be analyzed through the information provided by the performance of GMLVQ classifier. In particular, a comparison of the resulting prototypes and diagonal of the relevance matrices obtained from the experiments E1, E4 and E7,

using GMLVQ and applying the combination of both PD and AD pattern, will be consider by way of example. Remember that the prototypes give us information about how each class can be described in terms of features and the diagonal of the relevance matrix represent how important each feature is when it comes to the classification.

Thereby, first we can see from Figure 12 how when trying to distinguish between HC (class 1) and PD (class 2), the classifier uses different features to build a profile of the classes (prototype) depending on if the samples comes from CUN, UMCG or UGOSM. For instance, the last feature is quite important for defining the HC prototype coming from CUN data, when this prototype for the same class has to be defined for the UGOSM samples we realize that the importance is reduced and simply not used in the case of UMCG. This is a sign that the data of the centers are different and could indirectly explain why we can differentiate so easily the origin center where a sample comes from, as each center has different features to characterize the same disease. Consequently, we do not have the ideal conditions that we intended in our objectives to build a universal classifier capable of classifying any subject from any center. This is reflected in the results shown in Tables XX-XXIX from the experiments described in Table VII, where we have tried to test the classifier in data slightly different from the data in which it was trained.

Besides which are the most important features for each center when distinguishing the subjects condition, we can also observe from Figure 12 that the variation among the relevance of the features in the case of the prototype of PD from CUN is very different in comparison with the same prototypes of the others centers.

Finally, from Figure 13, we can see how the relevance profile, that reflect the importance of the different features for classification, varies for each center. Thus, for each center there are different discriminative features during classification, which may be related to the fact that the centers end up being so distinguishable.

VI. Conclusions and Outlook

In this project we have seen, in section II, the need to develop algorithms that process large amounts of data and provide an automatic diagnosis of neurodegenerative diseases, in our case of PD and AD. Then, in section III, the implementation of the algorithm to classify these diseases from FDG-PET brain images has been presented. Finally, in section IV and V, the experiments that have been carried out and the corresponding numerical and graphical results have been described, respectively.

As discussed above, from the experiments described in the Tables II-V, the performance of which is given in Tables IX-XVIII, we can conclude that the different subjects of the centers with whom we have worked are clearly distinguishable, both when it comes to distinguishing diseases within a particular center, and to discriminate the origin center of the subjects. Also we can observe from the results that the measures of performance do not point to a single best classifier.

Regarding the performance of the classifiers, from the experiments of binary classification of diseases by center, see Table II and Figure 11, we can see that the SVM classifier seems to provide a better performance than GMLVQ. While in the case of the experiments where we compare the centers by diseases and the ones where we classify the diseases without distinction of the different centers centers, see Figures 15-16, we have found that it is the GMLVQ classifier which slightly overcomes, or equals the performance of the SVM classifier. So, we find that there is not a significant difference among the performance of the GMLVQ and SVM classifiers. Both classifiers have an average standard deviation of 0.08 when computing the AUC.

We observe from the results in Tables IX-XII,XVI-XIX that our initial objective, which was based on distinguishing the subjects condition (HC, PD or AD) has been achieved, both when it comes to a specific center and when the data from the three centers (CUN, UMCG and UGOSM) were mixed together.

These results can be analyzed through the information provided by the performance of GMLVQ classifier. In particular, a comparison of the resulting prototypes and diagonal of the relevance matrices obtained from the experiments E1, E4 and E7, using GMLVQ and applying the combination of both PD and AD pattern, will be considered by way of example. Remember that the prototypes give us information about how each class can be described in terms of features and the diagonal of the relevance matrix represent how important each feature is when it comes to the classification.

Thereby, first we can see from Figure 12 how when trying to distinguish between HC (class 1) and PD (class 2), the classifier uses different features to build a profile of the classes (prototype) depending on if the samples comes from CUN, UMCG or UGOSM. For instance, the last feature is quite important for defining the HC prototype coming from CUN data, when this prototype for the same class has to be defined for the Italian samples we realize that the importance is reduced and simply not used in the case of UMCG. This is a sign that the data of the centers are different and could indirectly explain why we

can differentiate so easily the origin center where a sample comes from, as each center has different features to characterize the same disease. Consequently, we do not have the ideal conditions that we intended in our objectives to build a universal classifier capable of classifying any subject from any center. This is reflected in the results shown in Tables XX-XXIX from the experiments described in Table VII, where we have tried to test the classifier in data slightly different from the data in which it was trained.

Besides being able to know which are the most important features for each center when distinguishing the subjects condition, we can also observe from Figure 12 that the importance of the features when it comes to build the prototypes varies from center to center.

Then, from Figure 13, we can see how the relevance profile, that reflect the importance of the different features for classification, varies for each center. Thus, for each center there are different discriminative features during classification, which may be related to the fact that the centers end up being so distinguishable.

From the results we can derive that:

- For a single center we can distinguish diseases very well, see Tables IX-XII.
- Data from the centers are too different, see Tables XIII-XV, to get a universal classifier, when it comes to test it with data that differs on the origin of the training set.
- If we want a universal classifier the data has to be calibrated in a more homogenous and uniform way, i.e. all the centers performing neuroimaging should adhere to the same calibration, settings and protocols.

We conclude that although we have been able to achieve one of our main objectives, i.e. the discrimination of the different diseases from the FDG-PET scans, we have not been able to obtain a universal classifier, which has been trained on data of one center and then applied to data from another center (data which is slightly different). Our results show that the data sets display center specific properties which hinders the development of a generally applicable classifier. To do this, the centers should have a more universal approach when preprocessing the data and normalizing them in a different way to get rid of the specific properties of the center, becoming more homogeneous.

From the aforementioned points we can summarize that even though for binary classification SVM and GMLVQ classification performances were comparable, GMLVQ enables multi-class classification and it is interpretable unlike SVM. From GMLVQ we could approach our problem (of distinguishing between HC, AD, and PD) in a more straightforward way which SVM could not deal with. Due to the interpretable nature of GMLVQ we could gain further insight into the diseases themselves (which features are more important for one disease and which are for another, and how the relevance profile of these features are in a healthy subject). Also we got an idea of what hindered the development of a universal AD/PD/HC classifier.

We propose as an interesting future line of work to project back the relevant characteristics from the feature space (subject scores) to the voxel space (actual image) in order to get insight into the relevant patterns that distinguish the diseases. The clinicians might be able to gain a better understanding of the disease state themselves, their deviation from the HC, and their progression. It would also be interesting to study if the centers used the exact same calibration settings and protocols, the knowledge transfers of classifier trained on subjects of one center would be possible on the subjects of another center (test center).

ACKNOWLEDGMENT

Firstly, I would like to express my gratitude to my supervisor Prof. Michael Biehl, who provided me an opportunity to join their team as intern, for the continuous support of my Master thesis, his guidance helped me in all the time of research and writing of this thesis.

I would like to thank Prof. Xavier Giro who provided me with his supervision the opportunity to do my internship abroad.

My sincere thanks goes to Deborah Mudali for providing her code as a starting point for this project. As well as R.Remco, S.Meles, N.Leenders and R.Kogan for providing me the data.

Also I am grateful to Sreejita Ghosh, for her patience, support and valuable guidance.

Last but not the least, I would like to express my very profound gratitude to my parents, Pedro and Luciana, my boyfriend Ismael and mainly, to my sister Estefanía, for providing me with unfailing support and continuous encouragement throughout

my years of study. This accomplishment would not have been possible without them. Thank you.

REFERENCES

[1] Society for neuroscience. https://www.sfn.org/. Accessed: 2017-06-23.
[2] European commission. http://ec.europa.eu/health/. Accessed: 2017-05-29.
[3] S. Przedborski, M. Vila, and V. Jackson-Lewis, "Series introduction: Neurodegeneration: What is it and where are we?" *The Journal of clinical investigation*, vol. 111, no. 1, pp. 3–10, 2003.
[4] Web site of EU joint programme - neurodegenerative disease research. https://http://www.neurodegenerationresearch.eu/. Accessed: 2017-05-29.
[5] M. Firbank, A. Yarnall, R. Lawson, G. Duncan, T. K. Khoo, G. Petrides, J. O'brien, R. Barker, R. Maxwell, D. Brooks *et al.*, "Cerebral glucose metabolism and cognition in newly diagnosed parkinson's disease: Icicle-pd study," *J Neurol Neurosurg Psychiatry*, vol. 88, no. 4, pp. 310–316, 2017.
[6] C. C. Tang, K. L. Poston, T. Eckert, A. Feigin, S. Frucht, M. Gudesblatt, V. Dhawan, M. Lesser, J.-P. Vonsattel, S. Fahn *et al.*, "Differential diagnosis of parkinsonism: a metabolic imaging study using pattern analysis," *The Lancet Neurology*, vol. 9, no. 2, pp. 149–158, 2010.
[7] Mayfield - brain & spine. https://https://www.mayfieldclinic.com/PDF/PE-PD.PDF. Accessed: 2017-05-29.
[8] L. K. Teune, "Glucose metabolic patterns in neurodegenerative brain diseases," 2013.
[9] G. Garraux, C. Phillips, J. Schrouff, A. Kreisler, C. Lemaire, C. Degueldre, C. Delcour, R. Hustinx, A. Luxen, A. Destée *et al.*, "Multiclass classification of fdg pet scans for the distinction between parkinson's disease and atypical parkinsonian syndromes," *NeuroImage: Clinical*, vol. 2, pp. 883–893, 2013.
[10] European parkinsons disease association. https://http://www.epda.eu.com/. Accessed: 2017-05-29.
[11] S. K. Meles, L. K. Teune, B. M. de Jong, R. A. Dierckx, and K. L. Leenders, "Metabolic imaging in parkinson disease," *Journal of Nuclear Medicine*, vol. 58, no. 1, pp. 23–28, 2017.
[12] F. Niccolini, P. Su, and M. Politis, "Dopamine receptor mapping with pet imaging in parkinsons disease," *Journal of neurology*, vol. 261, no. 12, pp. 2251–2263, 2014.
[13] C. Loane and M. Politis, "Positron emission tomography neuroimaging in parkinson's disease," *American journal of translational research*, vol. 3, no. 4, p. 323, 2011.
[14] D. Aarsland, B. Creese, M. Politis, K. R. Chaudhuri, D. Weintraub, C. Ballard *et al.*, "Cognitive decline in parkinson disease," *Nature Reviews Neurology*, 2017.
[15] S. M. Broski, C. H. Hunt, G. B. Johnson, R. F. Morreale, V. J. Lowe, and P. J. Peller, "Structural and functional imaging in parkinsonian syndromes," *Radiographics*, vol. 34, no. 5, pp. 1273–1292, 2014.
[16] D. J. Bernal. Parkinson's is not Alzheimer's. https://drbernalneurologo.net/2015/12/09/parkinson-no-es-alzheimer/. Accessed: 2017-05-29.
[17] Nih senior health. https://nihseniorhealth.gov. Accessed: 2017-05-29.
[18] C. P. Weingarten, M. H. Sundman, P. Hickey, and N.-k. Chen, "Neuroimaging of parkinson's disease: Expanding views," *Neuroscience & Biobehavioral Reviews*, vol. 59, pp. 16–52, 2015.
[19] J. G. Goldman and R. Postuma, "Premotor and non-motor features of parkinsons disease," *Current opinion in neurology*, vol. 27, no. 4, p. 434, 2014.
[20] R. K. Brown, N. I. Bohnen, K. K. Wong, S. Minoshima, and K. A. Frey, "Brain pet in suspected dementia: patterns of altered fdg metabolism," *Radiographics*, vol. 34, no. 3, pp. 684–701, 2014.
[21] Alzheimer association. http://www.alz.org/. Accessed: 2017-05-29.
[22] World health organization. http://www.who.int. Accessed: 2017-05-29.
[23] D. Mudali, L. Teune, R. J. Renken, K. L. Leenders, and J. Roerdink, "Classification of parkinsonian syndromes from fdg-pet brain data using decision trees with ssm/pca features," *Computational and mathematical methods in medicine*, vol. 2015, 2015.
[24] M. Politis, "Neuroimaging in parkinson disease: from research setting to clinical practice," *Nature Reviews Neurology*, vol. 10, no. 12, pp. 708–722, 2014.
[25] J. Moeller and S. Strother, "A regional covariance approach to the analysis of functional patterns in positron emission tomographic data," *Journal of Cerebral Blood Flow & Metabolism*, vol. 11, no. 1_suppl, pp. A121–A135, 1991.
[26] G. E. Alexander and 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," *Human Brain Mapping*, vol. 2, no. 1-2, pp. 79–94, 1994.
[27] P. G. Spetsieris and D. Eidelberg, "Scaled subprofile modeling of resting state imaging data in parkinson's disease: methodological issues," *Neuroimage*, vol. 54, no. 4, pp. 2899–2914, 2011.
[28] J. R. Quinlan, "Learning decision tree classifiers," *ACM Computing Surveys (CSUR)*, vol. 28, no. 1, pp. 71–72, 1996.
[29] M. Biehl, D. Mudali, K. Leenders, and J. Roerdink, "Classification of fdg-pet brain data by generalized matrix relevance lvq," in *International Workshop on Brain-Inspired Computing*. Springer, 2015, pp. 131–141.
[30] T. Kohonen, "Learning vector quantization for pattern recognition," 1986.
[31] D. Mudali, M. Biehl, S. Meles, R. Renken, D. G. Garcia, P. Clavero, J. Arbizu, J. A. Obeso, M. R. Oroz, K. L. Leenders *et al.*, "Differentiating early and late stage parkinson's disease patients from healthy controls," *Journal of Biomedical Engineering and Medical Imaging*, vol. 3, no. 6, p. 33, 2016.
[32] J. R. Quinlan, "C4. 5: Programming for machine learning," *Morgan Kauffmann*, vol. 38, 1993.
[33] D. Garcia-Garcia, P. Clavero, C. G. Salas, I. Lamet, J. Arbizu, R. Gonzalez-Redondo, J. A. Obeso, and M. C. Rodriguez-Oroz, "Posterior parietooccipital hypometabolism may differentiate mild cognitive impairment from dementia in parkinsons disease," *European journal of nuclear medicine and molecular imaging*, vol. 39, no. 11, pp. 1767–1777, 2012.
[34] L. K. Teune, A. L. Bartels, B. M. de Jong, A. Willemsen, S. A. Eshuis, J. J. de Vries, J. C. van Oostrom, and K. L. Leenders, "Typical cerebral metabolic patterns in neurodegenerative brain diseases," *Movement Disorders*, vol. 25, no. 14, pp. 2395–2404, 2010.
[35] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.
[36] B. Schölkopf and A. Smola, "Support vector machines and kernel algorithms," *Alex Smola. Accessed March*, vol. 20, p. 2016, 2002.
[37] T. Kohnen, "Learning vector quantization for pattern recognition," Technical report no. TKK-F-A601, Helsinki, Tech. Rep., 1986.
[38] A. Sato and K. Yamada, "Generalized learning vector quantization," in *Advances in neural information processing systems*, 1996, pp. 423–429.
[39] T. Bojer, B. Hammer, D. Schunk, and K. Tluk von Toschanowitz, "Relevance determination in learning vector quantization," in *ESANN'2001*, M. Verleysen, Ed. D-facto publications, 2001, pp. 271–276.
[40] P. Schneider, M. Biehl, and B. Hammer, "Adaptive relevance matrices in learning vector quantization," *Neural Computation*, vol. 21, no. 12, pp. 3532–3561, 2009.
[41] P. Schneider, *Advanced methods for prototype-based classification*. PhD thesis, University of Groningen. Atto Producties Europe, 2010.
[42] M. Biehl. A no-nonsense matlab (tm) toolbox for gmlvq. software available at. http://www.cs.rug.nl/biehl/gmlvq.html. Accessed: 2017-05-29.
[43] T. Fawcett, "An introduction to roc analysis," *Pattern recognition letters*, vol. 27, no. 8, pp. 861–874, 2006.