Classification of Mild Cognitive Impairment and Alzheimer's Disease Patients: a Multiscale and Multivariate SVM Approach

Eduard Vilaplana Martínez

LTS 5, EPFL, June 20, 2012

Classification of Mild Cognitive Impariment and Alzheimer's Disease Patients: a Multiscale and Multivariate SVM Approach

presented on July 11th 2012
as Erasmus student
School of Engineering
Institute of Electrical Engineering
Signal Processing Laboratory 5
École Polytechnique Fédérale de Lausanne

for the Degree of Enginyeria de Telecomunicació
by

Eduard Vilaplana Martínez

supervised by:

Prof Jean- Philippe Thiran, LTS 5, EPFL, director
Dr Meritxell Bach Cuadra, MIAL, CHUV, UNIL, supervisor
Dr Jonas Richiardi, MIP, STI, EPFL, supervisor

Lausanne, EPFL, 2012

Everybody is a genius.
But if you judge a fish by its ability to climb a tree,
it will live its whole life believing that it is stupid.
— Albert Einstein

To all that have made this dream come true.

Acknowledgements

It is pleasure for me to present this work and I would like to give thanks to all people who gave me the oportunity to make it possible.

First of all, I would like to express my special gratitude to Meritxell Bach Cuadra for his availability, constant support and all the invaluable lessons learned.

My gratitude also goes to Jonas Richiardi for the precious scientific and methodological guidance in SVM and classifiers theory.

I also would like to express my thanks to Prof. Jean-Philippe Thiran, Associate Professor in Signal Processing at EFPL, for his kindness and availability for responding emails more than one year ago.

Last but not least, we would like to thank Prof. Armin Von Gunten and Alessia Donati for providing us with Lausanne Data set and for their fruitful discussions on MCI.

Lausanne, 19 de julio de 2012 E. V. M.

Abstract

Alzheimer's Disease (AD) is the most common form of dementia and a growing health and socioeconomic problem. Moreover, the impact of the disease is expected to increase even more as the life expectancy is going to grow over the years. Consequently, a lot of research is focused on computer-aided diagnosis techniques that aim at quantitatively study Magnetic Resonance brain images of early stage patients. Early diagnosis could help in better future cure or disease- modifying treatments. An example of AD early stage is Mild Cognitive Impariment (MCI), as the 50 % of the individuals who suffer from this pathology develop AD in three of four years.

In this work, we use Support Vector Machines to classify subjects from AD, MCI and healthy control (CTL) groups. Our main objective is to study whether combining different anatomical scale brain regions and different image modalities could improve the classification accuracy. Thus, regional and global Grey Matter (GM) volumes (multiscale approach), Withe Matter (WM) regional volumes, Regional Asymmetry coefficients and T_1- quantitative MRI data (multivariate) are combined. Our accuracies when comparing CTL vs AD and CTL vs MCI with large public databases (ADNI) are comparable to the results in the literature: 88.3 % and 81.8 % respectively. In this master thesis we study also smaller databases of MCI patients from Lausanne University Hospital. We pay special attention to the study of pre-processing steps: Intra Craneal Volume normalization and age correction. Our results show that for our small group of patients, better accuracies can be obtained when combining different types of features (multiscale and multivariate) than when only using classical GM region volumes. Moreover, the new region-based age-correction method proposed here presents encouraging results when applied prior to both CTL vs MCI and CTL vs AD classification.

Resumen

La Enfermedad de Alzheimer (EA) es la forma más común de demencia y se ha convertido en un problema socioeconómico creciente. Además, se prevé que el impacto de la enfermedad será aún mayor dentro de unos años debido al progresivo envejecimiento de la población mundial y al crecimiento de la esperanza de vida. Es por estas razones que en los últimos años se ha centrado la atención en técnicas computarizadas para la diagnosis que están dirigidas al estudio cuantitativo de imágenes de resonancia magnética (MRI) de cerebro de pacientes que se encuentran en una etapa temprana de la enfermedad. Un diagnostico precoz podría mejorar la efectividad de los futuros tratamientos de curación o modificación del curso natural de la enfermedad. Un ejemplo de etapa temprana de EA es el Deterioro Cognitivo Ligero (Mild Cognitive Impariment o MCI), puesto que el 50 % de los pacientes que padecen esta patología desarrollan EA en tres o cuatro años.

En este estudio, usamos Support Vector Machines para clasificar sujetos de tres grupos diferentes: EA, MCI i sujetos sanos de control (CTL). Nuestro objetivo es estudiar si combinando información a diversas escalas anatómicas del cerebro y diferentes modalidades de imágen se puede mejorar la precisión de la clasificación. De este modo, se han utilizado volúmenes regionales y globales (multiscale) de Materia Gris (GM), volúmenes regionales de Materia Blanca (WM), Coeficientes de asimetría e información de MRI T_1 cuantitativa (multivariate). Nuestras precisiones cuando comparamos CTL vs EA y CTL vs MCI usando bases de datos públicas (ADNI) son comparables a los resultados de la literatura: 88.3 % i 81.8 % respectivamente. En este proyecto también estudiamos una base de datos más pequeña de pacientes con MCI del Lausanne University Hospital. Prestamos especial atención al estudio de los pasos de pre-procesado: normalización por Volumen InterCraneal y corrección de edad. Los resultados obtenidos muestran que, para nuestro grupo reducido de pacientes, se obtienen precisiones mejores cuando se combinan diferentes tipos de datos (multiscale y multivariate) que cuando solamente se usan los clásicos volúmenes regionales de GM. Además, el nuevo método propuesto de corrección de edad basado en regiones presenta resultados esperanzadores cuando se aplica previo a ambas clasificaciones CTL vs MCI y CTL vs EA.

Resum

La Malaltia d'Alzheimer (MA) és la forma més comú de demència i ha esdevingut un problema socioeconòmic creixent. A més, es preveu que l'impacte de la malaltia serà encara més gran d'aquí a uns anys a causa del progressiu envelliment de la població mundial i al creixement de l'esperança de vida. És per aquestes raons que en els últims anys s'ha centrat l'atenció en tècniques computarizadas per la diagnosi que estan dirigides a l'estudi quantitatiu d'imatges de ressonància magnètica (MRI) de cervell de pacients que es troben en una etapa primerenca de la malaltia. Un diagnòstic precoç podria millorar l'efectivitat dels futurs tractaments de curació o modificació del curs natural de la malaltia. Un exemple d'etapa primerenca de MA és el Deteriorament Cognitiu Lleuger (Mild Cognitive Impariment o MCI), ja que el 50 % dels pacients que pateixen aquesta patologia desenvolupen MA en tres o quatre anys.

En aquest estudi, fem servir Support Vector Machines per classificar subjectes de tres grups diferents: MA, MCI i subjectes sans de control (CTL). El nostre objectiu és estudiar si combinant informació a diverses escales anatòmiques del cervell i diferents modalitats d'imatge es pot millorar la precisió de la classificació. D'aquesta manera, s'han utilitzat volums regionals i globals (multiscale) de Matèria Gris (GM), volums regionals de Matèria Blanca (WM), Coeficients d'asimetria i informació de MRI T textsubscript 1 quantitativa (Multivariate). Les nostres precisions quan comparem CTL vs MA i CTL vs MCI amb bases de dades públiques (ADNI) són comparables als resultats de la literatura: 88.3 % i 81.8 % respectivament. En aquest projecte també estudiem una base de dades més petita de pacients amb MCI l'Lausanne University Hospital. Prestem especial atenció a l'estudi dels passos de pre-processat: normalització per Volum intercranial i correcció d'edat. Els resultats obtinguts mostren que, pel nostre grup reduït de pacients, s'obtenen millors precisions quan es combinen diferents tipus de dades (multiscale i Multivariate) que quan només s'usen els clàssics volums regionals de GM. A més, el nou mètode proposat de correcció d'edat basat en regions presenta resultats esperançadors quan s'aplica previ a la classificació CTL vs MCI i CTL vs MA.

Índice general

Índice de figuras

Índice de cuadros

1 Introduction

1.1. Clinical Motivation

Alzheimer's disease (AD) is the most common form of dementia, an overall term for a decline in mental state. Dementia does not refer to an specific disease, it's a general term that describes a wide range of symptoms related to memory loss or other type of decline severe enough to reduce the patient's ability to carry out with daily normal tasks. AD is related to pathological amyloid depositions and hyperphosphorylation of structural proteins in the brain which progressively lead to brain disorders, such as loss of function, metabolic alterations and structural changes in the brain.

AD is nowadays, besides the major cause of dementia, a growing health and socioeconomic problem, due to the progressive ageing of the world population. In 2012, the direct costs of caring for AD patients to American society is estimated in $200 billion (comparable to the $500 billion per year in education in elementary and secondary school[63]), and 1,1$ trillion in 2050[41]. So the impact of the disease will increase as the life expectancy is going to grow even more over the years, according to the UN (Figure 1.1)

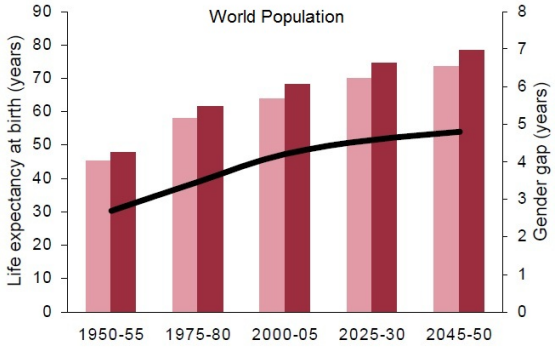


Figura 1.1: Male and female life expectancy at birth and gender gap, 1950-2050, got from http://www.un.org/esa/population/publications/worldageing19502050/pdf/8chapteri.pdf).

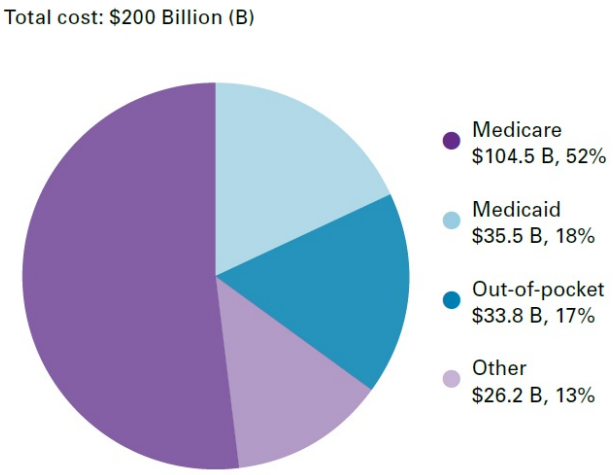


Figura 1.2: Aggregate Costs of Care by Payer for Americans Age 65 and Older with Alzheimer's Disease and Other Dementias, got from [41]

AD is not just the disease but it is also an important cause of death, specifically the 6th leading cause of death in the United States, and the only one in the top 10 in America without a way to prevent, cure or slow its progression[41].

The progression depends on each individual but three stages are usually considered[41]: Pre-clinical Alzheimer's Disease, Mild Cognitive Impairment (MCI) due to AD and Dementia due to AD. Many studies had concluded that the future treatments to slow or stop the progression of Alzheimer's disease and preserve brain functionality will be more effetive when applied during the initial stages of the disease, such as preclinical AD or Mild Cognitive Impairment. Therefore, it is very important to have useful tools that can identify at an early stage which individuals will progress with the disease. But this issue takes even greater importance for those individuals with MCI: 15 % of them will develope AD every year and the half of them will develop AD in three or four years [44]. It is estimated that between 10 % and 20 % of people older than 65 have MCI [42, 43], so the problem is already a reality.

Due to this socioeconomic costs, a lot of effort is focused on finding out which brain regions are affected at an early stage of the disease and could be used as biomarkers to diagnose and monitorize the disease. Nowadays, many techniques and tests contribute to AD and MCI diagnosis, but there is still interest in developing new techniques. One of the most popular tests is, for instance, the Mini-Mental State Examination (MMSE), which is a brief 30-point questionnaire used to screen for cognitive impariment. But there is more, the usal diagnosis process also includes physical and neurological exams, laboratory tests, neuropsycological testing and brain imaging. In the last 20 years there have been special interest in the last one, as it appears to be one of the most promising tools for diagnosis and monitoring the disease. The Magnetic Resonance Imaging (MRI) is a medical imaging technique used in radiology to visualize internal structures of the human body. When applied to the brain the resultant

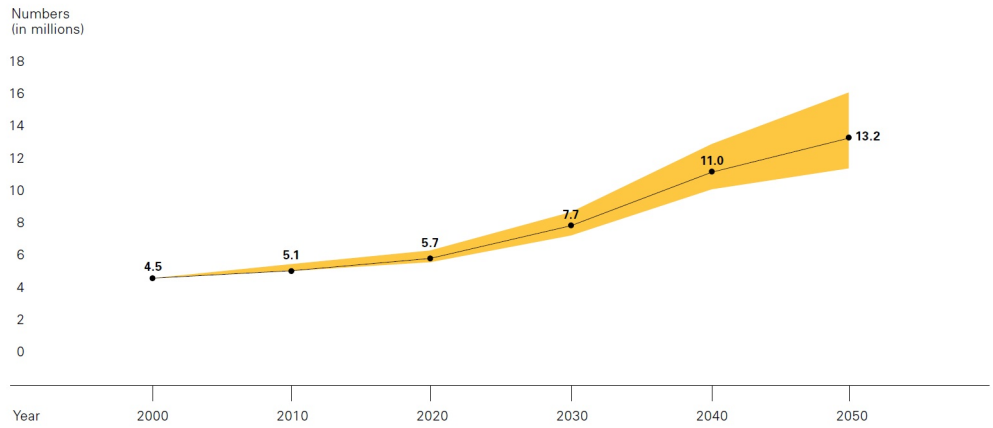


Figura 1.3: Projected numbers of people age 65 and over in the U.S. population with AD using the U.S. Census Bureau Estimates of Population Growth, got from [41]

images provide distinction between the different brain tissues, such as Gray Matter(GM), White Matter (WM) and CerebroSpinal Fluid (CSF). Moreover, MRI is a non-invasive method that let physicians explore the brain before the dead of the patient (More details about MRI are given in the Section 1.2).

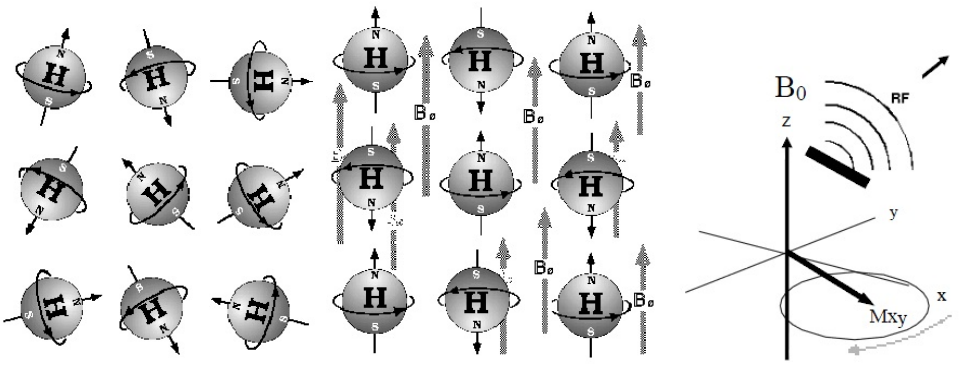
On account of the growing interest in this techniques, a lot of organizations have begun to find their way in neuroimaging recruiting. One of the most popular is the Alzheimer's Disease Neuroimaging Initiative (ADNI)[1], which is a multisite longitudinal clinical/imaging/genetic/biospecimen/biomarker study. Its objective is to determine the characteristics of AD progression, starting from normal aging, evolving to mild symptoms or MCI and finally ending in dementia. ADNI provides a huge amount of useful data from several different tests, including genetics, clinical and, of course, imaging.

Many studies have been performed in the last decade about AD diagnosis based on neuroimaging, as it has proven to be an adequate tool to play a major role in this scenario. The discrovery of robust image biomarkers will not only let future the treatments focus on the early affected regions but make a diagnosis before the disease symptomatology appears.

1.2. Medical Imaging and Pattern Recognition for Diagnosis

MRI is a noninvasive medical test that helps physicians diagnose and treat medical conditions. One of the advantages of the MRI is that it is harmless to the patients. It uses strong magnetic fields and non-ionizing radiaton, unlike X-ray Computed Tomography and traditional X-rays.

[1] Part of the data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://www.loni.ucla.edu/ADNI). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. ADNI investigators include (complete listing available at http://www.loni.ucla.edu/ADNI/Collaboration/ADNIAuthorshiplist.pdf)



(a) Absence of magnetic field (b) Applying magnetic field (c) Protons release the absorbed energy

Figura 1.4: Proton behaviour when (a) absence of magnetic field, (b) applying magnetic field, (c) releasing the absorbed energy. Got from http://www-ee.uta.edu/Online/alavi/ee4328-5339Spring12/MRI_Physics.pdf

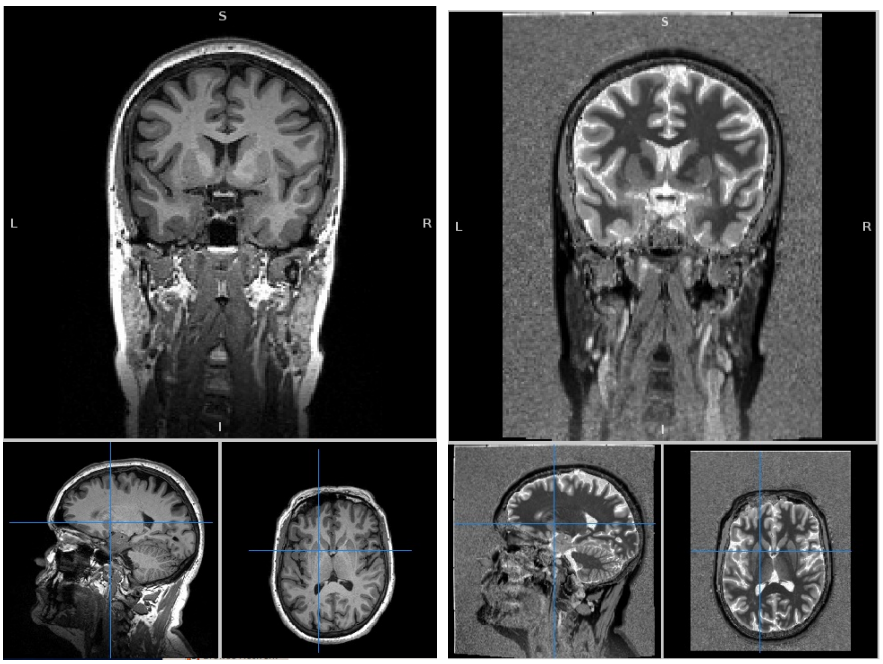
Moreover, MRI provides comparable spatial resolution and better contrast resolution.

MRI involves imaging of the proton, the positively charged spinning nucleus of hydrogen atoms that are common to be found in tissues containing water, proteins, lipids and other macromolecules. Due to the spin and charge, the protons act like a compass needle when placed in a magnetic field, assuming an alignment with respect to the field. But unlike a compass needle, a proton can align in two directions, either with or against the field (Figure 1.4(b)). When radiofrequency energy at the appropriate frequency is applied, protons aligned with the magnetic field absorb the energy and changes the orientation. The protons subsequently release the absorbed energy (Figure 1.4(c)) and go back to the original position at a rate determined by the T_1 and T_2 relaxation times. Those times depend in a complex way on the physical and chemical characteristics of the tissue. In this process of relaxation, the protons produce a voltage known as magnetic resonance signal which is captured by an antenna that surrounds the patient. A magnetic resonance image represents a display of spatially localized signal intensities, drawn on the final image as points of relative brightness or darkness[55].

Although there are several basic types of MRI, such as Diffusion MRI, Magnetization transfer MRI, Functional MRI and others, here we focus only on the ones useful for our work:

1. T_1*-weighted*. Refers to a set of standard scans that represents differences in the T_1 relaxation time of various tissues in the body. In a determinate instant, the voltage produced by the protons is capted by the antenna. Then, it is proportionally plotted in the final image. In this type of scan, water appears darker than fat, whereas in the brain, T_1-weighted scans provide appreciable contrast between gray and white matter. See Figure 1.5 (a)

2. T_1*-quantitative*. This scan is a measure of the promptness of a tissue to return to its



(a) T_1-weighted MRI (b) T_1-quantitative MRI

Figura 1.5: (a) T_1-weighted MRI, (b) T_1-quantitative MRI, got from Lausanne Database

longitudinal state of magnetic equilibrium, after removal from this state with an RF pulse[65]. The equilibration of the longitudinal magnetization is a exponential recovery process, that can be expressed as a function of time:

$$f(t) = 1 - e^{\frac{-t}{T1}} \tag{1.1}$$

In this scan, the relaxation time T_1 of each voxel is estimated by capturing the emitted signal by the protons at, at least, two different time instants and isolating T_1 from the equation (1.1). Finally, this T_1 value is plotted in the resultant image. See Figure 1.5 (b)

MRI it is a very common used technique as it has demonstrated to be a good solution for the study, diagnosis and monitoring of the disease, and had let the physicians explore the brain in a way that we had not ever imagined 20 years ago. Today it is widely accepted that changes measured in MRI are appropiate biomarkers for AD and MCI[3]. Actually, it had been proved through MRI studies that many structures are affected in AD, like Hippocampus, Amygdala and Entorhinal Cortex [1, 2, 3, 5, 13, 16]. However, less studies about MCI have been done and is not yet clear which structures are affected in this early stage.

Thus, to study the large databases, many tools dedicated to neuroimage processing have begun to find their way in the last ten years, providing a lot of information of the brain regions. For example, the Statistical Parametric Mapping (SPM: http://www.fil.ion.ucl.ac.uk/spm/),

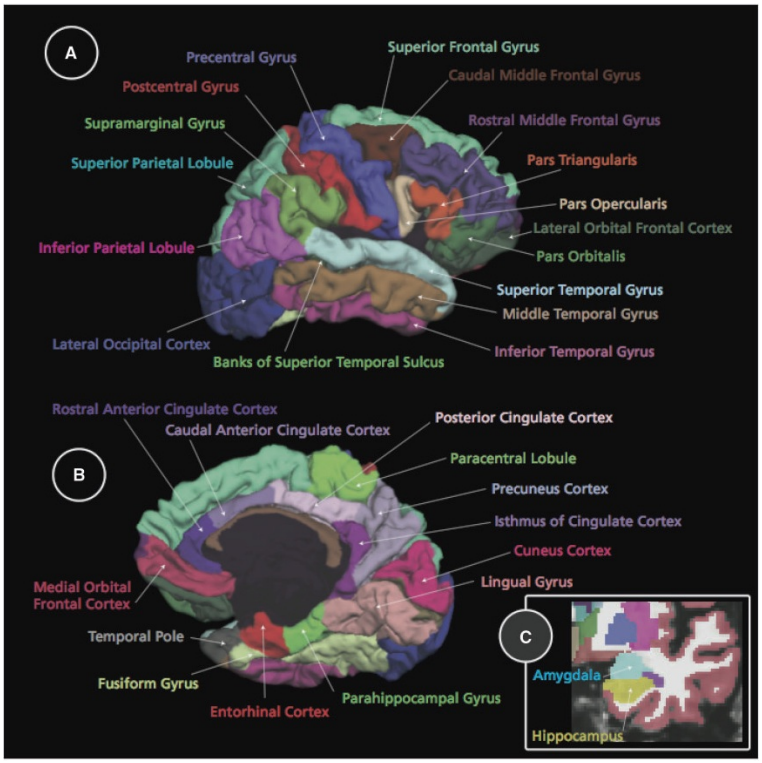


Figura 1.6: 3D representation of all 34 Regions of Interest (ROI) provided by the cortical parcellation from FreeSurfer, (A) lateral view, (B) medial view, (C) the two non-neocortical regions visible in the coronal view of a T_1-weighted MRI, got from [13]

that is a software designed for the analysis of brain imaging data sequences. The sequences can be a series of images from different cohorts, or time-series from the same subject. The SPM voxel based approach permits realign images, spatially normalize into a standard space and smooth them. Another tool that has generated interest in this research field is FreeSurfer (Massachusetts General Hospital, Boston, MA: http://surfer.nmr.mgh.harvard.edu/), that is a set of automated tools for reconstruction of the brain's cortical surface from structural MRI data, and overlay of functional MRI data onto the reconstructed surface. This software package performs a complete analysis of a brain MRI and provide many statistics, including cortical thickness, surface area or volume from several regions. For instance, the cortical parcellation provides volumetric information from 68 regions, 34 from each hemisphere (Figure 1.6).

Until the last decade, the image study has been normally done manually, but nowadays it can no longer be done in this way if one want to take advantage of the huge amount of available data. Is for that reason that there is growing interest in automatized methods, specially in machine learning techniques [4, 5, 6, 9], as they are less time consuming, and not observer dependant. Such techniques are able to use information from the whole brain at the same time, considering the relationship between regions and structures, which make them able to better distinguish among groups. Moreover, the results obtained by those techniques are at least comparable to a radiologist diagnosis accuracy[8]. There exist a wide number of types of

classifiers, such as Generative Models, Discriminative Models, Nearest Neighbour Models. The one chosen for this work is Support Vector Machines (SVM). It is not difficult to understand the overall working of this kind of techniques. Consider a data set, consisting of examples from two different classes. Some examples from the whole data set are given to the machine learning algorithm to identify the differences between groups in order to generalize and be able to predict the class of any input from the original dataset. More detailed information about machine learning classifiers will be given in Section 3.1.

1.3. Goals of this work

In the previous sections it has been explained the importance of an early diagnosis of subjects suffering from AD, but especially of those individuals suffering from MCI. To make the future treatments more effective it is very important to identify which individuals are at most risk of developing the disease and also which regions are early affected.

Many studies[1, 4, 5, 6, 9] can be found in the literature, which have attempted to provide robust approaches to solve this diagnosis problem, either using MRI voxels (Voxel-Based or VB[2]) or brain region volumes (Regions Of Interest or ROI[3]) for the classification, as explained in Chapter 2. The first ones used manual segmentations and pattern analysis, but the recent advances in signal processing and pattern recognition have changed the scenario. First of all, automated methods are able to extract very large amount of data from the brain regions such as volume, thickness or surface. Then machine learning techniques perform a complete analysis taking into account all the variables at the same time, considering the relationships between regions and capting in a better way the full patern of atrophy. More details about the methods used in the literature will be given in Chapter 2. In this work an SVM classifier has been used, as it has been proved to be a powerful tool for this kind of study[4, 5, 6, 9, 14, 17, 19].

Altough many studies about which brain regions are involved in AD degeneration have been performed, there is still a lot to discover about MCI. The main goal of the work is to find out which brain regions are early affected in MCI and provide best class separation. The approach of this study is region-based as we believe that anatomically grouped voxels in brain regions are more likely to show differences between classes than only isolated voxels. Region volumes have been widely used in the literature to feed classifiers, but in this work we hypothesized that combining more input variables can provide best classification accuracies and overall performance. Apart from the classical region volumes (Figure 1.6), different brain scales have been combined: the SVM will be fed with brain lobe volumes (Figure 1.7). This is known as multiscale approach. But there is still more, also other different variables have been combined: an asymmetry coefficient and data from T1 quantitative MRI have been added to our analyisis, what is known as multivariate approach.

[2] VB: the input features of the classifier are based directly on the voxels of the MRI, without grouping them into anatomical regions

[3] ROI-based: the input features of the classifier are region characteristics, such as volume, thickness or surface. A more accurate description about ROI, especially those used in our study, will be given in Section 4.2.

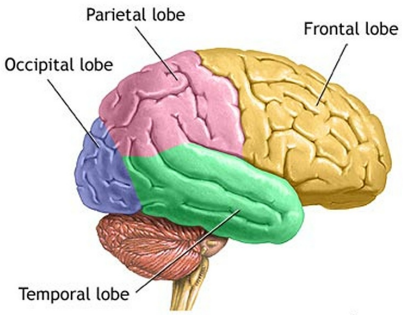


Figura 1.7: Brain divided in the four main lobes: Temporal, Occipital, Parietal and Frontal, got from [66]

It is not the first time that different variables have been used to perform classifications. For example, in 2011 different studies have reported better results when combining different types of information, either multiscale[1] or multivariate[19], but none of them had combined both techniques. In this work, both different anatomical brain scales and variables have been combined to test whether adding them provide more robustness to the classifier[4].

Moreover, a recent publication by *Dukart et al.,[17]* presented a new technique to control the brain normal age-related effect and suggested it should be treated. In this study the effectivity of this method have been checked, as well as if it should be usually applied.

In this study we are aware that many considerations must be taken into account when working with classifiers. As explained in Chapter 2, one must be very prudent with some results given in the literature; we suspect that some studies are probably giving better classification results than the average performance you would outcome from a real clinical environment. Is for that reason that we have been very clear, rigorous and methodic through all the work, in order not to give overestimated results that would not be useful in practice in real applications.

1.4. Structure

The structure of this study will be as follows. In Chapter 2 state of the art on neuroimage processing and pattern recognition for CTL, MCI and AD diagnosis will be presented. Then, mathematical formulation lying behind pattern recognition and classification problem will be summarized in Chapter 3. In Chapter 4.1 and 4.2 databases and MR imaging features that have been used for the study will be presented. In Chapter 4.3 the methodology used in this work will be explained. Then, the results will be analized in Chapter 6. Finally the conclusion and the future research lines will be presented in Chapter 7.

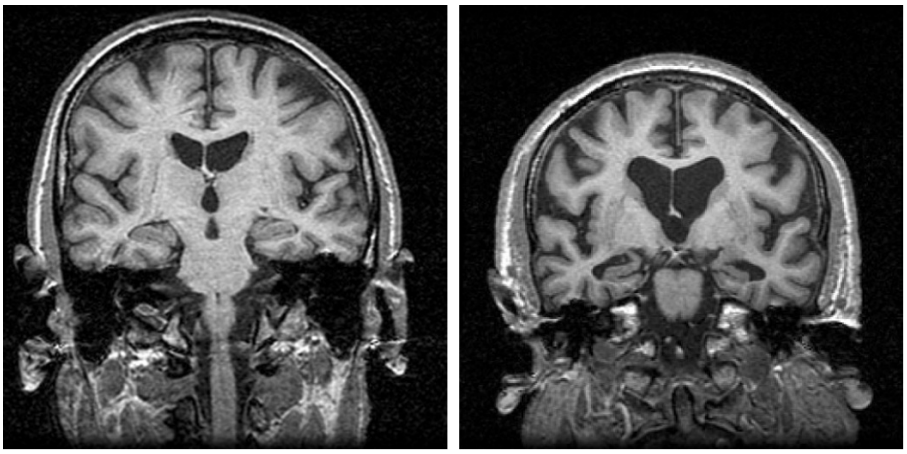
[4] A detailed description of the features used in this work is given in the Section 4.2

2 State of the art

MRI has been widely used for detection and diagnosis of AD as it appears to be one of the most immediate promises in terms of providing information on which patients are at risk of progress with dementia[49]. Although an autopsy is required for a definite diagnosis, the growing interest over the last 30 years in early detection had made imaging techniques progress and enhance the accuracy of ante-mortem diagnosis[48].

In 1991 *Braak et al., [47]* described the stages of AD. The course of the disease was divided into a determinate number of stages, although the speed of mental deterioration was subjected to interindividual variation. The study concluded that the accumulation of amyloid started before the appearence of clinical symptoms, what is known as the preclinical phase. Early changes have been demonstrated on the entorhinal cortex and the hippocampus[49] (see Figure 2.1) with the help of MRI and these changes are consistent with the underlying pathology of AD, but it is not yet clear which structures are most useful for early diagnosis of the disease, especially for MCI[49]. Nowadays, volumetric MRI is still too variable to be used as a reliable and valid clinical measure for MCI, and further investigation is required to more accurately determine which are the areas involved in an early stage.

The first studies in this area used manual segmentation of the hippocampus[51, 52](see Figure 4.2), reporting very accurate results, up to 92% of correct classifications. However, some results suggest that this kind of study has limitations and inaccuracies. For example, [52] conlcuded that the age had no effect on the hippocampal volumes of the CTL subjects at the age span of this study (21-79). Although the author had cited many references to support the assertion, more recent studies, for instance [17], contradict this statement. Moreover, the missclassified CTL subjects in [52] tended to be the older ones, what suggests that the age-related effect should be treated (this matter will be discussed later in Section 3.2). Even so, the study concluded with a interesting statement: this kind of missclassification would be due to the fact that hippocampal atrophy may precede the symptoms of dementia, so these subjects may possibly represent preclinical dementia. Entorhinal cortex measures had also been used to compare AD to CTL subjects [53, 54].



(a) Normal Control subject (b) Subject with AD

Figura 2.1: Coronal T_1 weighted MR scans showing (a) CTL subject and (b) AD subject. Note the visible marked atrophy of the hippocampus and the temporal cortex in the AD subject compared with control. Got from [48]

Despite the good results in AD versus CTL subjects differentiation, manual measurements of these structures on MR images are extremely time consuming, observer dependant and probably will not capture the full pattern of the atrophy. Actually, the main reason for using this manual delineation was that image processing techiques were not as advanced as they are nowadays. The recent advances have allowed the researchers to have more accurate and more amount of information of the brain regions, which would be useful for early detection of AD and MCI. Thus, multivariate tools are needed to analyze the huge amount of data that is nowadays available. Is for that reason that there has been growing interest in machine-learning techniques, such as Support Vector Machines, as they are able to perform analysis taking account multiple variables at the same time and extract the complex pattern of atrophy obtained from different brain regions.

Hitherto, many studies about automatic AD and MCI image classification have been done, either region (ROI) ([5]) or voxel-based (VB) ([4, 5, 8, 9, 10, 12, 16]). The methods and results got by each one are summarized in the Table 2.1, and are more explained in the paragraphs below. The most discriminant features in the literature can be found in the Table 2.2.

In 2007 *Fan et al., [12]* had done two different approaches to solve the AD, MCI and CTL classification problem, using the ADNI database. First of all, a ROI analysis was performed, using volumetric information of the hippocampus and the entorhinal cortex, both normalized by the total intracraneal volume. Then, the volumetric information was used to feed an SVM, and the cross-validation accuracies were 82.0 %, 76.0 % and 58.3 %, for AD versus CTL, MCI versus CTL and AD versus MCI, respectively. For the second experiment *Fan et al.* proposed a voxel based (VB) approach, the same method described in [6]. The basic idea is to perfom

Figura 2.2: Manual delineation of the hippocampus of an AD subject. The most posterior slice is on the lower right panel, got from [52]

a feature pre-selection of the image voxels and then apply an Recursive Feature Elimination (RFE[1]) feature selection. Those final selected voxels were applied to the SVM, which finally lead to cross-validated accuracies of 94.3 %, 81.8 %, and 74.3 %, respectively, for AD versus CTL, MCI versus CTL and AD versus MCI. As seen in the Figure 2.3, entorhinal and hippocampus volumes are not discriminative enough to well separate early stages of AD, such as MCI. In the voxel-based analysis, the regions that resulted to be more discriminative were the temporal lobe, especially the hippocampus, the superior, inferior temporal gyrus and the uncus, as well as medial Grey Matter atrophy, especially in the posterior cingulate and adjancent precuneus, and the medial aspect of the uncus. Additional Grey Matter atrophy was also found between AD and MCI patients: hippocampus, entorhinal cortex and middle and inferior temporal gyrus, also including the White Matter surrounding the hippocampus and the ventricles. The better results of the VB analysis suggest that more sophisticated methods for measuring structural brain differences between groups should be used for diagnosis and prognosis. The authors also made two last remarks. The first one is that the finding of reduced White Matter volumes between MCI and CTL merited further research. The second one is that the right hemisphere displayed higher magnitude and more widespread extent of atrophy of both GM and WM. The interpretation of such asymmetries is known to be problematic and also requires further studies.

In 2008 *Klöppel et al., [4]* did a VB approach using linear SVM to classify AD versus CTL. Three databases were used all along the paper[2]; the first one consisted of only 20 subjects from each

[1] RFE: Recursive Feature Elimination, is a feature selection technique. More details about RFE will be given in Section 3.2.

[2] In this study the best results have been got when using different databases for training and testing. Thus, in

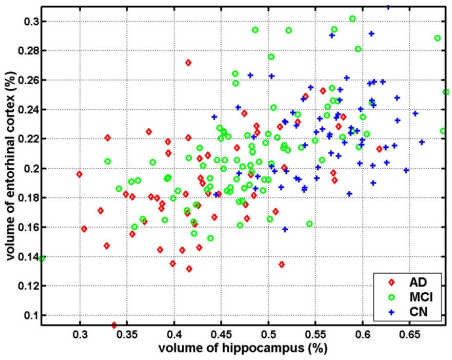


Figura 2.3: Scatter plot of the hippocampus and entorhinal cortex, both normalized by ICV, got from [12]

class, the second one of 14 subjects from each class and, finally, the third one, which was larger, consisted of 33 AD and 57 CTL subjects. The most discriminative voxels were clustered around the parahippocampal gyrus and parietal cortex. From different experiments, the best results were a sensitivity[3] of 100 % and a specificity[4] of 92 %. These are great results but some considerations should be taken into account. First of all, the fact that really small databases are used for test and training makes very difficult the interpretability and generalization of the results. The best results had been obtained when using the first and the second databases, getting accuracies up to 96.4 %. When using the larger database, the results dramatically go down to 81.1 %, significantly lower than [12]. According to our remark, the author published an article[59] one year after explaining that the accuracy becomes very variable when using around 20 subjects per group. Second, the MMSE scores of the AD subjects from the first and second databases are extremely low, what means that they are more severly diseased than the subjects used in other studies. Moreover, we have not been able to interpretate some procedure aspects. For example, a new interesting concept of combined kernel is presented, but it is not widely explained, so it is not easy for the reader to figure out what exactly is really being performed. Thus, one must be extremely prudent with this results.

Another voxel-based approach was made by [9] *Vemuri et al., 2008*, that used SVM and the databases from Mayo Clinic Alzheimer's Disease Research Center and Alzheimer's disease Patient Registry to perform a classification between AD and CTL. The data used for each patient include a structural MR scan, ApolopoproteinE (APOE) genotype information, and demographic details: age and gender. The APOE information was added because there is a well established positive risk for AD associated with the presence ok the $\epsilon 4$ allelle while $\epsilon 2$ is protective. The method used in this study was the a-STAND score, wich is a voxel based approach with demographic and APOE corrections of the original input features. In this case, the method used was clear and well explained. First of all, the database was split in train

the Summary Table 2.1, age and MMSE are expressed as A-B, being A the value for the first database and B for the second

[3] Sensitivity: Rate of correctly classified AD, or true positives. This issue will be explained in Chapter 5.1

[4] Specificity: Rate of correctly classified CTL, or true negatives. This issue will also be explained in Chapter 5.1

group and test group. All the training steps (feature selection, model selection and model optimization) were carried out only in the training set as the test set would be completely new and unseen to the model. This way, a minimally biased estimate ot the true diagnostic performance of the classifier is got. The best results were a sensitivity of 86 % and a specificity of 92 % when using combined information from the MRI, demographic variables, and APOE, although that results were not far from that the results for the combination of MRI information and demographic variables. Finally, the brain structural changes reported were in the medial temporal lobe, particularly the entorhinal cortex and the hippocampus, the posterior cingulate gyrus, the precuneus and the insula. Also White Matter losses were reported in the entorhinal parahippocampal gyrus and parietal lobe, what merits further research.

As seen, many different approaches have been proposed in the literature, but there is still some controversy in which strategy performs better. Moreover, one must be very careful when comparing results between studies, as many variables should be taken into account, like the database demography or the way the images are acquired. In 2010, *Cuingnet et al., [5]* did a comparision of ten methods using the ADNI database. The main idea was to simplify the task of comparing results accross experiments. Five voxel-based methods, three methods based on cortical thickness and two methods based on the hippocampus were tested. In order to obtain the most unbiased estimate as possible, the set of participants was randomly split in training set and testing set, both of the same size. The optimal parameters of the SVM classifier were found by cross-validation on the training set, and then the perfomance of the trained classified was evaluated on the testing set. Three group comparisions had been done: CTL versus AD, CTL versus MCIc and MCIc versus MCInc. MCIc is a subgroup that had converted to AD within 18 months, while MCInc had not converted at that time. For the last comparison all methods were unable to got valid results. For the CTL versus AD experiment the best results were a sensitivity of 81 % and a specificity of 95 % for the voxel based approach and linear SVM. For the comparision between CTL versus MCIc the best results were a sensitivity of 68 % and a specificity of 95 % with a linear SVM and the voxel atlas method, which is actually a region based method. The accuracy obtained is really high, but a consideration should be taken into account. The class separation is expected to be higher when classifiying MCIc subjects than when classifying MCI subjects as MCIc are more AD like patients than MCI subjects. Thus, the results are also expected to be more accurate. The authors reported that the oldest controls and the youngest patients were more often misclassified, even though no age effect was corrected. Finally, a combination of three approaches was tested. A convenient approach to combine different SVM-based methods is to consider that the resulting classifier is a SVM which kernel is a linear convex combination of the kernels of each method, known as the multiple kernel learning (MKL) solution. None of them improved the accuracies of the comparation AD versus CTL and only one slightly improved the MCIc versus CTL results, up to 76 % sensitivity and 85 % specificity.

Recent studies had suggested that carefully combining MRI information with clinical assessment and other variables and biomarkers would be useful for a better prognosis value for those patients suffering from MCI. For example, Cerebrospinal fluid (CSF) biomarkers have

also been studied in the diagnosis of AD and higher levels of τ and lower β-amyloid have been described as good predictors of the progression to AD in those patients suffering from MCI. As result of the growing interest in combining variables some studies had been carried out. A recent example is, for instance, *Westman et al., [1]* that made a different approach of the problem. First of all, the analysis was region-based instead of voxel-based. Moreover, different scales were combined to feed the classifier, i.e., not only global measurements but also small measurements; for instance, the hippocampus manual measurements. The author used Orthogonal Partial Least Squares (OPLS[5]) for the classification, which also differs from the others. The obtained results were as follows: for AD versus CTL a sensitivity of 90 % and a specificity of 94 % for the cross-validation but 81 % and 82 % for an external test set. The cross-validation results for CTL versus MCI were a sensitivity of 69 % and a specificity of 73 %, whereas for MCI versus AD were a sensitivity of 75 % and a specificity of 79 %. In this article the power of combining manual measures of the hippocampus and automated volume measures together was tested and showed better results when comparing AD versus CTL and AD versus MCI, but not when comparing MCI versus CTL. In this last case, the hippocampus measures alone showed the best predictive values.

In 2011, *Heckemann et al., [2]* presented a study about statistical analyses of automatically generated segmentations. The report showed many measure comparisions across groups, including single brain region volumes, lobe volumes and an asymmetry coefficient. Significant statistical differences have been found on the temporal lobe, as well as in the classical studied single volumes; for instance, in the hippocampus, the amygdala, the fusiform and the parahippocampal. Also left/right asymmetry have been found in posterior cortical regions, suggesting that merits further research. The technique used for the asymmetry will be discussed in Section 4.2.

Another example of combining different measures is the study carried out by Dukart et al., [19]. The proposed approach was to do a combined evaluation of FDG-PET and MRI to detect and differentiate between types of dementia. FDG-PET and MRI data were processed to get a precise overlap of all regions in both modalities. A new algorithm was designed to enable an accurate anatomical registration of both modalities. All processing steps were performed as far as possible simultaneously by applying the same deformations and preprocessing parameters to both modalities of the same subject. This procedure resulted in an accurate anatomical overlap of both imaging modalities and in an accurate between-subject registration, with both images having the same voxel size and approximately the same effective smoothness. Then, once the ROIs were extracted, SVM classification was applied with varying parameters separately for both modalities and to combined information obtained from MR and FDG-PET images. The best results were got when combining information from MRI and FDG-PET, yielding to an accuracy of 100 % for the CTL versus AD comparision. Those results suggest that the integration and combination of results from different imaging modalities may provide a

[5] OPLS is a statistical method related to principal components regression; instead of finding hyperplanes of minimum variance between the response and independent variables like SVM does, OPLS finds a linear regression model by projecting the predicted variables and the observable variables to a new space.

new way to improve the diagnostic accuracy. However, one must be cautious with the results as the used database consisted of only 13 CTL and 21 AD.

As seen above, many studies and several approaches have been done in medical image classification for diagnosis of AD, but there is still controversy when studying MCI. Recently, some studies had concluded that multiscale[1] or multivariate[19] approaches would yield to better results than single approaches, but none of them had combined both techniques. The main purpose of this work is to find out which brain regions are early affected in MCI and can be used as reliable biomarkers. Our hypothesis is that combining more measures than only the classical volumetric or voxel data, higher perfomances can be achieved. It will be tested wether combining different variables and scales, i.e., small brain regions, lobes, etc, adds robustness and accuracy to the prediction, making it more useful for future clinical applications. As seen in the literature, machine learning techniques had proven to be a good solution to address this issue and have become an standard for image classification for diagnosis. The classification technique chosen for this work is Support Vector Machines, that is also a very popular method in the literature and appears to yield to the best results. Moreover, continuing with the study presented in 2011 by *Dukart et al., [17]*, the age-related effect have been studied to determinate whether the correction techniques are really useful.

Cuadro 2.1: Summary of the State of the Art on automated classification for AD, MCI and CTL. When demographics characteristics are expressed as A-B, A refers to the train and B to the test database. Further details will about classifiers will be given in Section 3.1.

Author		Subjects			Methods	Accuracy CTL vs MCI	Accuracy CTL vs AD
		CTL	MCI	AD			
	Number	66	88	56			
[12]Fan et al., 2007	Age	75.2	76.4	77.4	Compare ROI & VB, SVM	ROI: 76.0 %, VB:81.8 %	ROI:82.0 %, VB:94.3 %
	MMSE	29.08	26.78	23.07			
	Number	20-14	-	20-14			
[4]Klöppel et al., 2008[4]	Age	79.5-63.0	-	81.0-65.0	VB, SVM	-	96.0 %
	MMSE	29.0-29.2	-	16.7-16.1			
	Number	140-50	-	140-50			
[9]Vemuri et al., 2008	Age	77.0-79.0	-	78-78.5	a-STAND score, SVM	-	89.0 %
	MMSE	29-29	-	22-20			
	Number	81-81	39-37	69-68			
[5]Cuignet et al., 2010	Age	76.1-76.5	74.7-74.9	75.8-76.2	Comparision 10 Methods	81.5 %	88.0 % (VB)
	MMSE	29.2-29.2	26.0-26.9	23.3-23.2			
	Number	100	100	100			
[1]Westman et al., 2011	Age	73	75	76	Multiscale, OPLS	71.0 % (only hippocampus)	92.0 %
	MMSE	29	27	21			
	Number	79	-	80			
[17]Dukart et al., 2011	Age	75.8	-	75.7	VB, SVM	-	85.0 %
	MMSE	28.7	-	23.6			
	Number	13	-	21			
[19]Dukart et al., 2011	Age	53.9	-	61.1	Multivariate, SVM	-	100.0 %
	MMSE	N/A	-	23.2			

Cuadro 2.2: Summary of the most discriminative features in the literature. The complete description of the features used in this work can be found in Section 4.2

Author	**Most discriminant features**
[12]Fan et al., 2007	*AD vs CTL*: Entorhinal Cortex and Hippocampus *MCI vs CTL*: Temporal Lobe (Hippocampus, temporal gyrus); Posterior cingulate, Precuneus, Insula, Medial orbitofrontal
[4]Klöppel et al., 2008	*CTL vs AD*: Parahippocampal gyrus and Parietal Cortex
[9]Vemuri et al., 2008	*CTL vs AD (GM)*: Medial temporal lobe, Temporal-parietal association cortex, Posterior cingulate and Precuneus, Insula *CTL vs AD (WM)*: Temporal lobe (entorhinal-parahippocampal gyrus) i Parietal Lobe
[5]Cuignet et al., 2010	*CTL vs AD and CTL vs MCI*: Hippocampus, Amygdala, Parahippocampal gyrus, Inferior and Middle temporal gyri, Posterior cingulate gyrus, Rosterior midlefrontal gyrus, Entorhinal cortex, Lateral Temporal Lobe, Inferior Parietal Lobe
[1]Westman et al., 2011	*AD vs CTL*: Hippocampus, Temporal GM, Total GM, Fornix *CTL vs MCI*: Temporal GM, Hippocampus, Total GM, Frontal GM, Fornix
[2]Heckemann et al., 2011	*AD vs CTL*: Hippocampus, Amygdala, Anterior Temporal Lobe, Parahippocampal *CTL vs MCI*: same as AD, and Insula, Occipital Lobe, Parietal Lobe, Lateral Ventricle

3 Technical Background

In this Chapter, the most important technical aspects are explained. In the first section, the fundamentals of SVM are introduced. Then, in the second section all the pre-processing steps are presented.

3.1. Support Vector Machines

3.1.1. Introduction

Machine learning techniques have become very popular in the last 20 years in this research field as have been proven to be a robust approach for the problem of MR image classification [4, 5, 6, 9, 17]. The recent advances in quantitative medical image analysis have provided scientists a lot of novel measures. Specialized softwares like FreeSurfer are able to perform a complete analyisis of an MR image and extract a huge amount of data, such as surface, thickness or volumes from many regions. Thus, techniques able to handle this huge amount of information are extremely required. Machine learning techniques have demonstrated to be a suitable tool, as they can learn differences between groups using all the data togheter at the same time, considering complex relationships between features. In addition, recent studies[8] have reported equal or better accuracies when comparing to an expert radiologist. The use of computer-asisted methods for diagnosis would not only provide another diagnosis tool, but also improve the speed of diagnosis without compromising accuracy.

Support Vector Machine has been used in this study as it is the most widely chosen option when dealing with machine learning techniques in neuroimage classification. Although the readers do not need to understand the underlying theory behind SVM, we briely introduce the basics necessary for explaining our method in the next subsection.

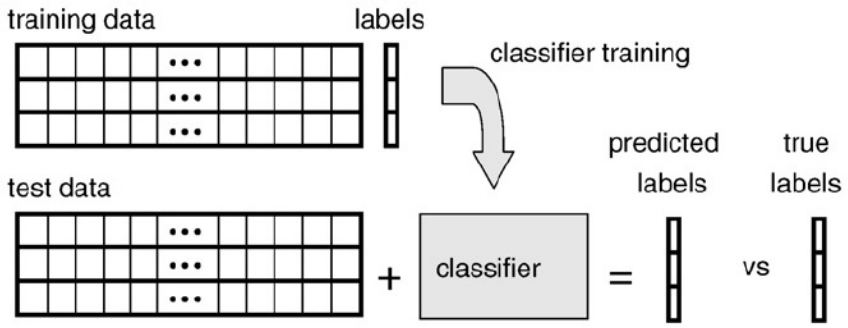


Figura 3.1: General overview of an SVM, got from [14]

3.1.2. Fundamentals

An SVM can be considered like a "black box"that determines the belonging of group of a new instance, based on a training procedure . A classification often involves separating data in training and testing set. Each example on the training set contains a label, depending on which group does it belong, and several attributes, called features. In our study the possible class labels are CTL, MCI or AD. The main objective of the SVM is to produce a model able to predict the labels of the test data given only the test data features (Figure 3.1). Then, the accuracy is reported comparing the predicted labels with the true labels.

Given a training set of instance-label pairs $(x_i, y_i), i = 1, ..., l$ where $x_i \in R^n$ and $y \in \{1, -1\}^l$, the SVM has to find the hyperplane that maximize the margin between classes:

$$\langle w, \Phi(x) \rangle + b = 0 \tag{3.1}$$

Corresponding to the decision function

$$\text{f(x)} = \text{sign}(\langle w, \Phi(x) \rangle + b) \tag{3.2}$$

In order to find this maximum margin hyperplane, SVM require the solution of the following optimization problem:

$$\begin{aligned} \min_{w,b,\xi} \quad & \frac{1}{2} w^T w + C \sum_{i=1}^{l} \xi_i \\ \text{subject to} \quad & y_i (w^T \Phi(x_i) + b) \geq 1 - \xi_i, \\ & \xi_i \geq 0 \end{aligned} \tag{3.3}$$

where $\Phi(x_i)$ maps x_i into a higher-dimensional space and C>0 is the regularization parameter.

SVM finds a linear separating hyperplane with the maximal margin in this higher dimensional space. But due to the possible high dimensionality of the vector variable w, usually the dual problem is solved:

$$\begin{aligned} \min_{\alpha} \quad & \frac{1}{2}\alpha^T Q\alpha - e^T\alpha \\ \text{subject to} \quad & y^T = 0, \\ & 0 \le \alpha_i \le C, \qquad i = 1,...,l \end{aligned} \tag{3.4}$$

where $e = [1,...,1]^T$ is the vector of all ones, Q is an l by l positive semidefinite matrix, $Q_{ij} \equiv y_i y_j K(x_i, x_j)$ and $K(x_i, x_j) \equiv \Phi(x_i)^T \Phi(x_j)$ is the kernel function. After the equation (3.4), using the primal-dual relationship, the optimal w satisfies:

$$w = \sum_{i=1}^{l} y_i \alpha_i \Phi(x_i) \tag{3.5}$$

and the final decision function is

$$\text{sign}\ (w^T\Phi(x) + b) = \text{sign}\ (\sum_{i=1}^{l} y_i \alpha_i K(x_i, x) + b) \tag{3.6}$$

The SVM technique uses information from all the features at the same time. Looking at the Figure 3.2 it is easy to see that neither the feature y_1 nor y_2 are able to separe groups, and it's only when combining information from both that a correct separation is achieved. Ideally, an SVM analysis would yield to an hyperplane that completely separates the feature vectors in two non-overlapping groups. However, in real world problem it is not likely to get an exactly separate line dividing the data within the space. It also may produce a model with high-dimensional feature vector that is overadapted to the training data and does not generalize well; this is known as overfitting. To allow some flexibility in separating the classes, SVM uses the hyperparameter C (see second term in the Equation (3.3)), that controls the trade off between allowing training errors and forcing rigid margins. This way, it allows a point to be on a determinate distance on the wrong side of the hyperplane without violating the constraint[62] (See Figure 3.2). A high cost value C will force the SVM to create a more complex model to missclassify as few training examples as possible, while a lower cost parameter will lead to a simpler prediction function, which will probably generalize better.

The C value should be optimized, as the accuracy of an SVM highly depends on the selection of the model parameters. The common way to do the optimization is to separe the testing set at the begining, and split the training data set in different parts, for example in two. Then, the

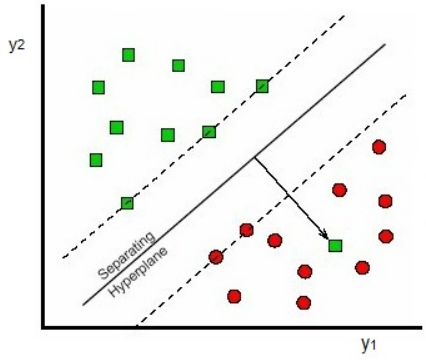


Figura 3.2: Linear separation used, but admiting training errors. Penalty of the error: distance to the hyperplane multiplied by C, got from [61]

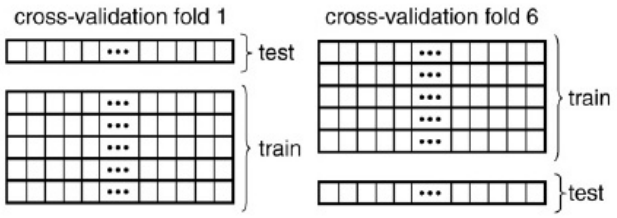


Figura 3.3: Typical v-fold Cross-Validation procedure, got from [14]

model is trained changing the C value in one of this parts of the training set and tested in the other. Finally, the C value that yield to the best accuracy is selected to train the final model and predict the unseen data. An improved version of this procedure is known as cross-validation, in which the method described above is done once for each partition of the training data set. In v-fold cross-validation (See Figure 3.3), the training set is divided into v subsets of equal size. Sequentially one subset is tested using the classifer trained on the remaining v-1 subsets. This way, each instance of the whole training set is predicted once so the cross-validation accuracy is the percentage of data which are correctly classified. If v=n, being n the total number of instances in the training set, the procedure is known as Leave-One-Out Cross Validation (LOOCV). This is a common option to deal with small databases, although it is computationally expensive. Cross-validation prevents the overfitting problem since the C value is optimized without using the testing data.

In the Equation (3.3), the features are mapped in a higher dimensional space. The reason to do this is that sometimes the examples are not separable in the original space, but may be classificable in the new high-dimensional space. For example, as seen in the Figure 3.4, the original data was completely overlapped, but the mapping in the higher dimensionality space make it possible to distinguish among groups. The function $K(x_i, x_j) \equiv \Phi(x_i)^T \Phi(x_j)$ is called the kernel function. There exist several kernels, but the following four are considered the basic ones:

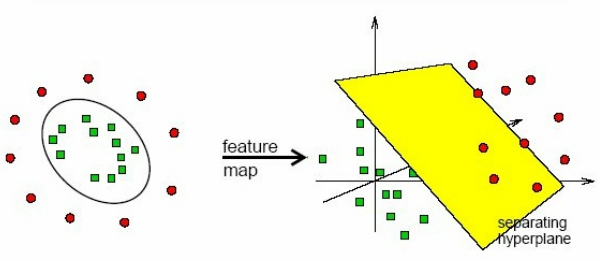


Figura 3.4: The kernel function may map the data into a higher dimensional space, what would make it possible to perform the separation, got from [61]

Kernel 1. *Linear.* $K(x_i, x_j) = x_i^T x_j$

Kernel 2. *Polynomial.* $K(x_i, x_j) = \gamma x_i^T x_j + r)^d$

Kernel 3. *Radial Basis Function (RBF).* $K(x_i, x_j) = exp(-\gamma \left\| x_i - x_j \right\|^2, \gamma > 0$

Kernel 4. *Sigmoid.* $K(x_i, x_j) = \tanh(\gamma x_i^T x_j + r)$

The chosen of a deterimate kernel depends on the application. This issue is discussed in Section 4.3

3.2. Data pre-processing

As the SVM accuracy is very sensitive to the input data, some pre-processing steps must be done before feeding the classifer. The theory underlying these steps is explained in the following sub-sections.

3.2.1. IntraCranial Volume Normalization

The Intracranial Volume (ICV) normalization consists in dividing each volumetric feature by the total Intracranial volume of the subject. This way the differences between subjects in the same region due to the size of the head are reduced. This also helps to reduce the variability between male and femal differences. Although the ICV normalization is a very commonly used in the literature [1, 2, 7, 12], in this work both with and without ICV normalization results are presented. As no correlation between head size and disease has been reported, it is normal to think that the global volumetric mesures should be ICV normalized, otherwise SVM will classify mostly by the head size. In our opinion, with smaller structures, where the variation between subjets is even smaller, the ICV normalization is perhaps removing subtile changes. Is for that reason that the results are reported with ICV and without ICV normalization.

3.2.2. Age Correction

It is widely known that there are age-related changes within healthy population in many different brain structures measured by MRI. The growing interest in group classification have given importance to the development of techniques to control the age-related effects, that normally lead to hide the disease-related effects and, therefore, higher rates of missclassification.

As age-related and disease-related effects deteriorates the brain in similar way, the classification algorithms are not able to differentiate between both and tends to misclassify the old control and the young demented subjects[5]. Is for that reason that it is very important to control those effects, remove the age-related one and let the classifier focuse only on the disease-related changes.

The most popular resort in the literature[1, 4, 5, 9, 10, 12] is to select, as much as possible, age matched groups. But to avoid the kind of missclassifications described above, groups used for training the classifier should ideally be matched at least once to every single subject in the test set. At practice, it is almost impossible to find groups of subjects large enough which match each subject in age and other confounding variables.

A recent study by *Dukart et al., 2011*[17] had presented a method to correct the age effect, a linear detrending method in terms of the general linear model (GLM). The age-related effects were estimated only with the CTL group beacuse, if the AD group was also considered, some disease-effect would have taken into account. A linear model was chosen in this study as it yielded to significantly higher perfomance respect the quadratic model. The procedure is described below.

First of all, a GLM was calculated for each feature separately; in this study, as the input features were directly the MRI voxels, the correction was done for each voxel. In our study, the regression is done per region. The X_c matrix is composed by two columns, which are a constant and age, and only the CTL subjects are used to compute the regression coefficients β, wich is composed by a constant β_0 and the first order term β_c. The following regression model has to be solved for β by the minimization of the sum of squared residuals, $\sum \epsilon_c^2 \rightarrow min$:

$$y_c = X_c\beta + \epsilon_c \tag{3.7}$$

Solving (3.7) for least squares (LS) estimates of β satisfies the following normal equations:

$$X_c^T X_c \beta = X_c^T y_c \tag{3.8}$$

Then, solving the linear equations system (3.8) for β results in:

$$\begin{pmatrix}\beta_0\\ \beta_c\end{pmatrix} = (X_c^T X_c)^{-1} X_c^T y_c \tag{3.9}$$

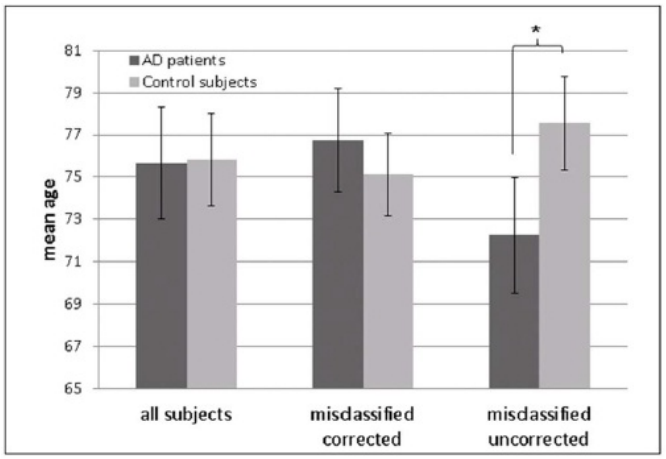


Figura 3.5: Age characteristics of missclassified subjects using SVM before and after aplying age correction, got from [17]

The Equation (3.9) returns a pair of values β_0 and β_c which are, respectively, the constant and the slope of the regression line. To obtain the age-corrected feature value, the equation (3.10) has to be applied:

$$y_{corrected} = y_{uncorrected} - \beta_c X_{age} \tag{3.10}$$

Finally, the equation (3.10) should be applied to each voxel and each subject, including both CTL and diseased subjects.

The accuracy reported in [17] was slightly higher if applying age correction, 85 % compared to 83 %, when classifying CTL versus AD. But more imporant, the groups of missclassified CTL and missclassified diseased subjects did not further show a difference in mean age[17], as seen in the Figure 3.5. This means that the classifier is working more independtly of the age, focusing more on the disease-related differences between groups and no longer sistematically misclassifying younger AD and older CTL subjects.

In this work another method has also been been studied. The used database is expected to have variability, as it is real data, there are not infinite examples, and probably there will be some outliers. The classical way to deal with regression outliers is to use LS and try to find the influent observations. The influence of one observation $z_i = (x_i, y_i)$ depends on being y_i too large or too small compared to other y's from similar x's. After the outliers are identified, some decision must be taken such as modifying or deleting them and applying LS again to the modified data[46]. This procedure is call Robust Regression, and there is a large and complex theory behind this concept that overcome the purpose of this work. For further details, see *Maronna et al., 2006 [46]* and *Davies et al., 1993 [45]*. A comparision of both can be seen in the Figure 3.6. The robust regression gives weights to the possible outliers and the slope changes respect the LS regression. Due to this modifications, the final Root Mean Square (RMS) error is likely to be lower than when using the classical LS Regression.

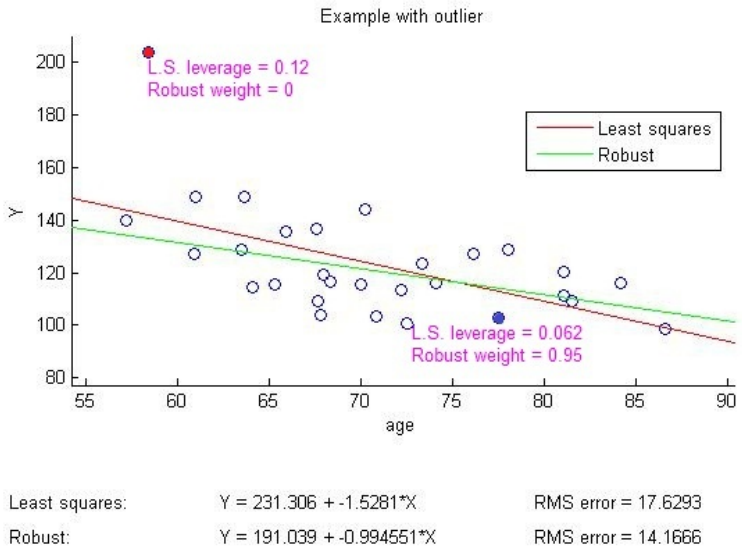


Figura 3.6: Comparision of Robust and Least squares Regressions. The robust regression gives weights to the possible outliers (in red)

Both ordinary linear and robust linear regression will be tested in this work.

3.2.3. Feature Scaling

It is very important to scale the input data before applying SVM [37, 21]. The purpose of scaling is to avoid higher range features having more importance than others with lower ranges. Another aspect that should be taken into account is the fact that SVM kernel values normally depend on the inner products of feature vectors, and large numbers might cause numerical problems [37]. At last but not least, high range attributes also make all the calculations more time consuming. So it is very important to scale the input data and, of course, scale both training and testing set with the same scale factor.

As seen in [11], there are many methods for scaling the input data but three different methods had been considered:

Method 1. *Norm-1*. Each subject's$_i$ feature value is divided by the sum of the values of each subject's value for this feature$_f$. The 1-normalized value for the subject$_i$ and the feature$_f$ is:

$$x_{i,f}^{norm_1} = \frac{x_{i,f}}{\sum_i x_{i,f}} \tag{3.11}$$

Method 2. *Norm-2*. Each subject's$_i$ feature value is divided by the root of the sum of the

square values of each subject's value for this feature$_f$. The 2-normalized value for the subject$_i$ and the feature$_f$ is:

$$x_{i,f}^{norm_2} = \frac{x_{i,f}}{\sqrt{\sum_i x_{i,f}^2}} \tag{3.12}$$

Method 3. *Norm-z*. The standard deviation σ_f and the mean μ_f value of the feature$_f$ vector are computed. The mean μ_f is substracted and then, this value is divided by the standard deviation σ_f. With this method, the new values have zero mean and standard deviation 1.The z-normalized value for the subject$_i$ and the feature$_f$ is:

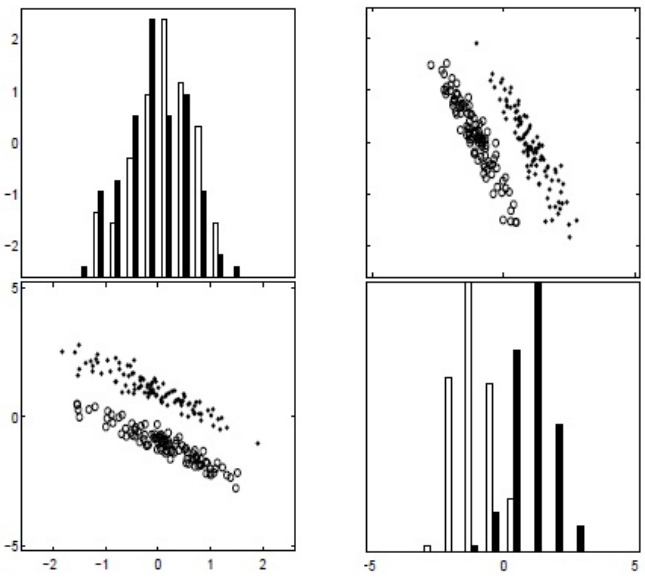
$$x_{i,f}^{norm_z} = \frac{x_{i,f} - \mu_f}{\sigma_f} \tag{3.13}$$

The method selected for the whole study from now on is the z normalization for two reasons. The first one is because it yield to better results than the other two options. The second is that it is the normalization typically used in the literature, as seen in [1], [9] or [21].

3.2.4. Feature Selection

A popular problem in machine learning and classification is to find ways to reduce the dimensionality of the feature space to overcome the risk of overly adapt the trained model to the training data, creating a model that do not generalize well. Data overfitting arises when the relationship between the number of features is higher than the number of instances. Normally, it is allowed to work with a difference of one magnitude order: $F \leq 10N$, being F the dimensionality of the feature space and N the dimensionality of the examples space. However, during the simulations we have noticed that, normally, the classifier works better when the relation is $F \approx N$. Given that there are usually more features than examples, it is normally worthy to reduce the number of the input features in order to let the classifier focuse in the important ones. There are many potential advantages on the feature selection: facilitating data undestanding, reducing the measurement and storage requirements, reducing the training and testing times and finally improving the accuracy of the final classifier[30].

As seen in [28, 29, 30, 31, 32, 33] there exist many techniques of feature selection, for instance Correlation-based Feature Selection, F-score or RFE. Another strategy which is not a feature selection method but could be useful to reduce the feature space in a fast and easy way is a t-test. N t-tests are performed in the feature space to find statistically significant differences across groups in each feature$_n$ of the feature space. The theoretically irrelevant features for classification, those that the null hypotesis is not rejected, are removed. Although it is a very easy and fast method, it has disadvantages. For example, that it does not take into account the relationship between features. This is a very critical issue, because as seen in [30], a variable that is completely useless by itself can provide a significant performance improvement when taken with others (see Figure 3.7). Also two variables that are useless when taken separately



(a) Useless variable by itself

(b) Useful variable in combination with others

Figura 3.7: A variable usless by itself (a) can provide improvement in class separation when combining with another variable (b). Got from [30]

can be useful together. Moreover, this method presented lower results in the analysis, so it was no longer used.

In 2002, a new technique named Recursive Feature Elimination (RFE) had been presented by *Guyon et al., [21]*. RFE is a method which performs backward feature elimination, i.e, it starts with all features F and sequentially removes the more irrelevant features until a subset S of a determinate size is left, according to the stop criteria. This is done by iterativilly perfoming this procedure:

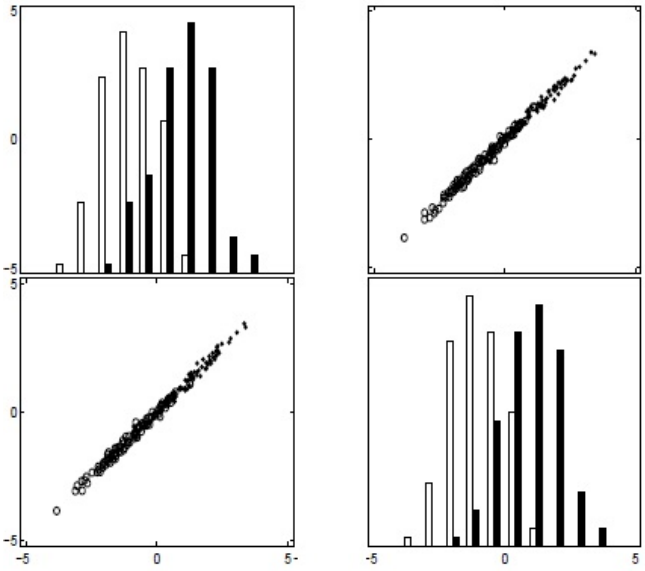
Step 1. Train an SVM with the actual feature space. In the first iteration it will be F, in the next iterations it will be S'.

Step 2. Sort the features by the values of w^2, being w the weight vector.

Step 3. Remove the feature with the smallest value of w^2. The new feature space S' has one less feature. Also M features can be removed in each iteration. In our case, it is done one by one.

Step 4. If the stop criteria is satisfied, go to *Step 5*. If not, return to *Step 1*.

Step 5. The resultant feature space $S = S'$ will be used to feed the classifier.



(a) Using two perfectly correlated variables (b) Using only one variable

Figura 3.8: There is no significant gain when using two perfectly correlated variables (a) instead of one (b). Got from [30]

Another aspect to take into account is the feature correlation. The procedure described above removes attributes with little weights, but probably some redundant features would be kept. If the correlation between two variables tends to be perfect, it means that no information is gained by combining them[30]. This can be seen in the Figure 4.1: combining perfect correlated variables does not provide an increase of the class separation power comparing when taking only one of them, what means that are truly redundant. Is for that reason that feature correlation must be controled before feeding the SVM. Nevertheless, according to *Guyon et al.,[30]* high variable correlation (or anti-correlation) dones not mean absence of complementarity. Thus, in order not to delete important information, the minumum threshold to be able to remove a variable has to be very high, at least 0.9, according to [64]. In our method, a classifier is trained before the decorrelation step in order to know which are the most discriminative features. Then, the feature correlation is checked, starting from the most discriminative, i.e. $F_1 - F_2$, $F_1 - F_3$, ..., $F_1 - F_F$, and removing the second term if necessary, in order to preserve the most discriminative features when correlated with some others. This way, if the most discriminative feature was correlated with the F_5, this one will be removed instead of F_1.

The feature selection method used in this work is the RFE algorithm, as it is one of the most popular methods in the literature [6, 12] and has proven good performances. Moreover, a decorrelation step has been introduced before the RFE procedure in order to remove the truly

redundant information and make things easier to the feature selection algorithm. More details about the complete procedure are given in the Section 4.3.

3.3. Potential Common Mistakes

Dealing with classifiers is not a simple task and many potential pitfalls must be avoided. Basically, the most important thing is that the testing set on which the results are reported must be completely independent of the training procedure; once a pure set is initially separed it must not be *touched* again but to report a final pure accuracy. A study performed in 2009 by *Pereira et al., [14]* presents the machine learning classifiers giving a tutorial overview and reviewing the crucial issues that must be beared in mind. The most important aspects that must be taken into account are summarized below:

- Classifier Parameter Tuning
  - *Feature selection must be independent of the testing set.* It is not allowed to select features that appear to distinguish one class from another in the whole dataset. The reason for this is that, actually, it permits information from the test set to affect the learning of the classifier in the training set, leading to overoptimistic accuracy estimates. Looking at the labels for the entire dataset is sometimes called *peeking*. However, this does not mean that the class labels cannot be used at all in feature selection. They can be used only once the data have been split into training and test sets, considering solely the training set[14].
  - *The parameter tuning must be independent of the testing set.* Exactly for the same reason of the first point, the C hyperparameter must be optimized only in the training set. Otherwise, the classifier would be overadapted to the whole dataset, and provide a completely bieased estimation of the true classification performance.
- Group Balance
  - *The groups must be balanced in terms of examples per group.* If this is not the case, the machine learning algorithm may tend to focus on the larger group, classifying the most numerous class per by default.
  - *If there is a variable that is likely to present differences between classes, the groups must be balanced in terms of this variable.* For example, it happens with the subject gender or when adding MRI acquired in multiple sites. In this last case, as explained in Chapter 4, balanced number of examples from each site must be taken to present reliable results. Otherwise, the classifier will focuse more in those variations than in the diseases-related differences.
- Data used
  - *Only one MRI scan should be selected from each subject.* When working whith MRI, two different scans from the same subject, even when acquired at different time

points, have a strong correlation degree. Is for that reason that the algorithm would classify in an easier way scans from one person if another scan of the same subject have been included in the train set. This also will outcome overestimated results.

4 Materials and Methods

In this Chapter, the Databases, features and methodology of our work are explained in the first, second and third sections respectively.

4.1. Database

All the experiments in this study were performed on three databases containing CTL, MCI and AD subjects. In this section the databases and features used in this study are presented below and summarized in the Table 4.1.

4.1.1. ADNI 1 Database

The first database contains indidividuals from the Alzheimer's disease Neuroimaging Initiative 1 (ADNI 1) database (http://www.loni.ucla.edu/ADNI), and is used to compare our method and results to the approaches in the literature (see Chapter 2). Only T1-weighted MPRAGE 1.5T images have been used. The MRI scan from the baseline visit has been used for each subject when available and from the screening visit otherwise. Only the amnestic MCI patients whose impariment was due to AD were selected. Finally, the database is composed by 185 individuals from each class, randomly selected to match age and gender as much as possible. The ages are for the CTL subjects (mean age ± SD) 76,59 ± 5,07 , 75,39 ± 7,37 for the MCI and 75,87 ± 7,52 for the AD subjects. The Mini Mental State Examination (MMSE) scores (mean ± SD) of the CTL, MCI and AD are respectively 29,08 ± 1,00, 27,06 ± 1,76 and 23,32 ± 2,00. Finally, the gender distribution is, M/F, 93/92, 92/93 and 96/89 for CTL, MCI and AD.

4.1.2. Lausanne Database

The Lausanne Database comes from: FNS project Number 122263, 2009-2013, "The impact of personality characterisitcs on the clinical expression of MCI", PI: Prof. Von Gunten, and it is composed by T1-weighted MPRAGE 3T MR images. Despite there were Amnestic, Non-

Cuadro 4.1: Summary of Databases Demography

	ADNI 1			Lausanne		Expanded	
Group(n)	CTL(185)	MCI(185)	AD(185)	CTL(29)	MCI(29)	CTL(58)	MCI(58)
Sex (M/F)	93/92	92/93	96/89	9/20	9/20	24/34	18/40
Age at MRI scan	76.6	75.4	75.9	70.5	70.4	71.7	71.5
MMSE score	29.1	27.1	23.3	28.9	27.9	29.3	27.5

Amnestic and Mixte MCI, only the Amnestic and Mixte were selected. From the Lausanne Database, T1-quantitative MRI were also added to our analysis. It contains 29 CTL and 29 MCI subjects. The age for the CTL is $70{,}47 \pm 7{,}39$ and $70{,}36 \pm 9{,}67$ for the MCI subjects. The MMSE scores are $28{,}90 \pm 1{,}23$ and $27{,}86 \pm 1{,}22$ for CTL and MCI respectively. The gender distribution is 9/20 for both groups.

4.1.3. Expanded Database

Finally, in order to increase the number of subjects per class, the Lausanne Database is exapanded adding ADNI 1 subjects. In order to create a well balanced model, only 29 subjects from each class can be added. Otherwise there will be more subjects from one database and there would be the potential risk to classify by site instead of by disease; the model would overadapt to site specific differences. Is for that reason that is very important to keep the same number of subjects per site per class. Thus, 29 subjects from each class were randomly added to Lausanne Database, matching age and gender as much as possible, to configure the Expanded Database. The age for the CTL is $71{,}66 \pm 7{,}86$ and $71{,}51 \pm 8{,}28$ for the MCI subjects. The MMSE scores are $29{,}28 \pm 1{,}12$ and $27{,}51 \pm 1{,}50$ for CTL and MCI respectively. The gender distribution is 24/34 for the CTL group and 18/40.

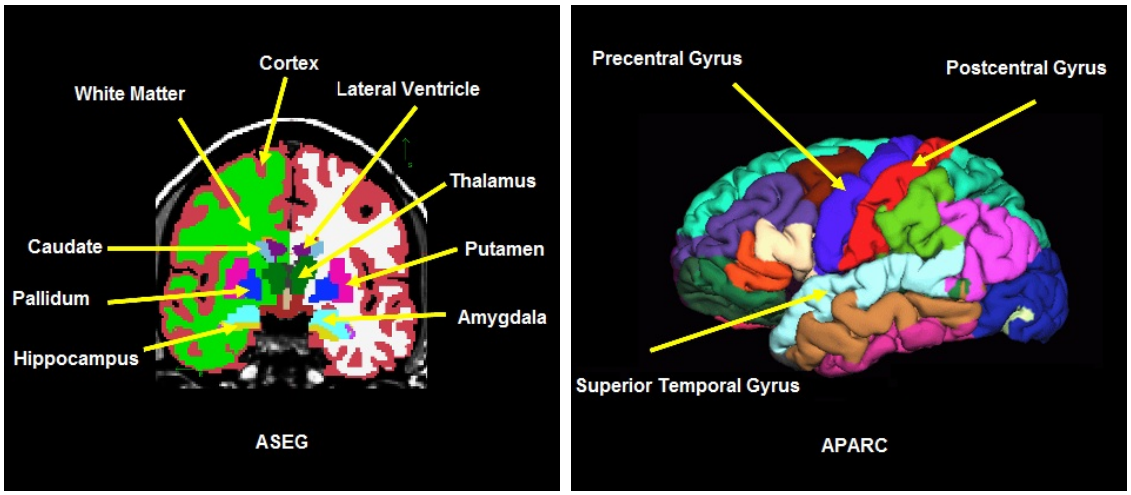
4.2. SVM Input Data

In this study five types of data have been used to feed the classifier, including Grey Matter Region Volumes, Lobe Volumes, Asymmetry coefficients, WM parcellation and T1-quantitative data. All of them are presented in the subsections that follow. The complete list of brain regions provided by FreeSurfer is given in the Appendix A.

4.2.1. Grey Matter Volumes

Specialized softwares like FreeSurfer perform a complete anaylisis of brain MRI and extract a lot of data about brain regions, such as thickness or volume. For instance, two commonly used output files are:

- *aseg*: is the statistical output from the subcortical segmentation. This file provide volume-



(a) Subcortical Segmentation, *aseg* from FreeSurfer

(b) Cortical Parcelation, *aparc* from FreeSurfer

Figura 4.1: Two different output files from FreeSurfer (a)*aseg* file (b)*aparc* file. Got from [67].

tric information of 55 regions, including, for example, the left and the right hippocampus. See Figure 4.1a.

- *aparc*: is the statistical output from the cortical parcellation. This file can provide, for instance, volumetric or thickness information of 68 regions, 34 from each hemisphere. See Figure 4.1b. In this study, only volumetric information have been considered.

Region Grey Matter Volumes

In this study, classical brain region volumes extracted with FreeSurfer have been used. However, the classifier has not been fed with all volumetric features, but only with a pre-clinical selection. The regions involved in AD have been widely studied, as seen in Chapter 2, and considering all the volumetric data provided by FreeSurfer would probably add unnecessary noise instead of valuable information. Thus, those features that commonly appears in the AD and MCI classification literature have been selected. Also features that are normally considered as *usual supects* in AD and MCI classification problems have been added. Finally, 47 brain regions have been used for this work:

- From the *aseg* file:
  - Right and Left Hippocampus
  - Right and Left Amygdala
  - Right and Left Inferior Lateral Ventricles
  - Right and Left Lateral Ventricles

- Right and Left Accumbens
- Right and Left vessel
- White Matter Hypointensities
- Right and Left Cortical White Matter.

- From the *aparc* file:
  - Right and Left Entorhinal
  - Right and Left Temporal pole
  - Right and Left Superior and Inferior Parietal
  - Right and Left Parahippocampal
  - Right and Left Lateral and Medial Orbitofrontal
  - Right and Left Middletemporal
  - Right and Left Insula
  - Right and Left Rostral and Caudal Anterior Cingulate
  - Right and Left Fusiform
  - Right and Left Precuneus
  - Right and Left Caudal Middlefrontal
  - Right and Left Isthmus Cingulate
  - Right and Left Lateral Occipital

The MRI files have been processed with FreeSurfer 4.4 for the ADNI 1 Database[1], while with Freesurfer 5.1 for the Lausanne and the Expanded Database. The complete list of the regions provided by the *aseg* and the *aparc* file can be found in the Appendix A.1 and Appendix A.2 respectively. For further details about the FreeSurfer outputfiles see *https://surfer.nmr.mgh.harvard.edu/fswiki*.

Lobe Volumes

As explained in Section 1.3, one of the most important goals of this work is to find out whether adding information on multiple anatomical brain scales (multiscale approach) improve the overall perfomance of the classifier. Our hypothesis is that, although multiple single isolated regions may not show differences, they probably do when grouped in higher scale structures. To carry out this task, the different brain lobe volumes have been calculated and added to

[1] The FreeSurfer 4.4 processed data used from ADNI do not provide exactly the same regions that can be found in FreeSurfer 5.1. In this case, the features have been selected to fit as much as possible the Lausanne Database clinical feature selection. The complete list of features selected from the ADNI 1 Database can be found in the Appendix A.5.

the classifier, using FreeSurfer volumes extracted from the *aseg* and the *aparc* files. The brain have been divided in 12 lobes, 6 per hemisphere, which are Frontal, Limbic, Parietal, Occipital, Temporal and Noyaux. The brain regions composing each lobe are those that follows:

- Frontal Lobe:
  - Lateral Orbitofrontal
  - Pars Orbitalis
  - Frontal Pole
  - Medial Orbitofrontal
  - Pars Triangularis
  - Pars Opercularis
  - Rostral Middlefronta
  - Superiorfrontal
  - Caudal Middlefrontal
  - Precentral
  - Paracentral
- Limbic Lobe:
  - Rostral Anterior Cingulate
  - Caudal Anterior Cingulate
  - Posterior Cingulate
  - Isthmus Cingulate
  - Parahippocampal
  - Entorhinal
  - Temporal Pole
  - Thalamus
  - Hippocampus
  - Amygdala
- Parietal Lobe:
  - Postcentral
  - Supramarginal
  - Superiorparietal
  - Inferiorparietal
  - Precuneus
- Occipital Lobe:
  - Cuneus
  - Pericalcarine
  - Lateraloccipital
  - Lingual
- Temporal Lobe:
  - Inferiortemporal
  - Middletemporal
  - Bankssts
  - Superiortemporal
  - Transversetemporal
- Noyaux Lobe:

- Caudate
- Putamen
- Pallidum
- Accumbens Area

As well as did for the classical volumetric data, the lobe volumes have been also calculated for both right and left hemisphere.

4.2.2. Asymmetry

As far as we know, very few studies refering about brain regions asymmetry in AD or MCI subjects can be found in the literature. For example, in 2006 *Fan et al.,[12]* reported right asymmetry pattern of atrophy in MCI. However, the interpretation of the brain asymmetry studies is known to be problematic. The author suggested that further research was required. *Heckemann et al., [2]* also studied the asymmetry in AD and MCI for some brain areas. The author reported statistically significant results in the Hippocampus for both AD and MCI versus CTL comparision, and in large regions when comparing AD versus CTL. Finally, according to *Fan et al.,[12]*, the study conlcuded that this was an area for future exploration[2]. The asymmetry coefficient for each region A_r used in *Heckemann et al., [2]* was calculated as follows:

$$A_r = \frac{2\,|V_R - V_L|}{V_R + V_L} \tag{4.1}$$

However, we suspect that computing the absolute value some useful information is being killed. Imagine the case that a region tends to be left predominant asymmetric in CTL but right predominant asymmetric in MCI. If the absolute value is computed, both CTL and MCI values will be classified in the same way, so the discriminative value is lost. Is for that reason that in our study the asymmetry coefficient used have been computed as follows:

$$A_{r-Lausanne} = \frac{V_R - V_L}{V_R + V_L} \tag{4.2}$$

As seen above, few studies have been performed about asymmetry in AD, and even less about MCI. Moreover, those that have studied asymmetry, normally have not looked beyond the hippocampus the amygdala[2]. In this study, 19 regions have been considered for the Asymmetry Study, including:

- Hippocampus
- Accumbens
- Inferior Lateral Ventricle
- Lateral Ventricle
- Vessel
- Amygdala
- Caudal Anterior Cingulate
- Rostral Anterior Cingulate
- Entorhinal
- Fusiform

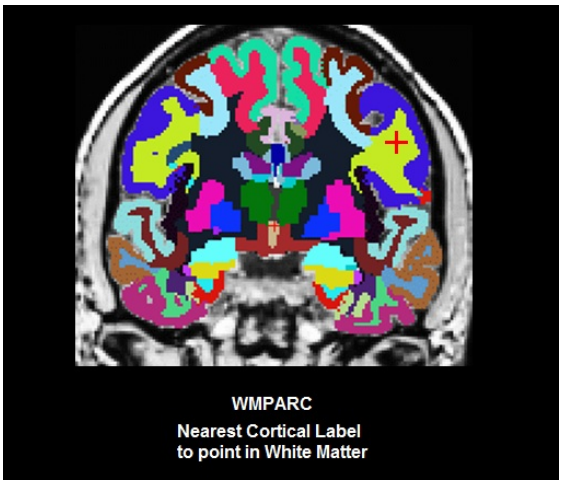


Figura 4.2: *wmparc* file from FreeSurfer, got from [67]

- Lateral Orbitofrontal
- Medial Orbitofrontal
- Middletemporal
- Parahippocampal
- Precuneus
- Insula
- Caudal Middlefrontal
- Temporal Pole
- Inferior Parietal

4.2.3. White Matter

Another FreeSurfer file has been studied in this work, the *wmparc* stats file, which is a table of the white matter brain region parcellation volumes. This file is based on the cortical parcellation technique and extends this labeling to the subcortical WM directly underlying the cortical parcellation[68][2]. As far as we know, really few studies have used this file to perfom their analysis, so no feature pre-selection based on the literature can be done. The input features will be the whole set of regions provided by this file. The complete list of the regions considered in the *wmparc* is given in the Appendix A.3.

4.2.4. T_1-quantitative MRI data

As explained in Section 1.3, our hypothesis is that combining measures from different image modalities the overall perfomance is likely to increase; this is known as multivariate approach.

[2]The WMparcellation method is an extension of the cortical parcellation procedure that utilized spherical spatial normalization to label gyral and sulcal areas throughout the brain. Cortical parcellations were subsequently used to assign a label to the underlying white matter by the construction of a Voronoi diagram in the WM voxels of the MR volume based on distance to the nearest cortical parcellation label. Each Voronoi polygon then inherited the label of the parcellation unit, yielding a complete labeling of the cerebral WM. For further details, see [68].

Thus, apart from the multiscale information, T_1-quantitative MRI data have been used to feed the classifier. Both single regions and lobe T_1-quantitative measures have been used. In both cases, the T_1 value is obtained computing the mean value of the voxels of a region. The complete list of the single regions cosidered in the T_1-quantitative analysis can be found in the Appendix A.4. Refering to the lobe measurements, the list is presented below:

- Frontal White Matter
- Frontal Grey Matter
- Parietal White Matter
- Parietal Grey Matter
- Occipital White Matter
- Occipital Grey Matter
- Temporal White Matter
- Temporal Grey Matter

4.3. Methodology

In this Section, the most important aspects of our methodology are explained. The summary of the complete procedure can be seen in the Figure 4.5 and is explained in the sections that follow.

The MRI measures described in Section 4.1 have been analysed with Support Vector Machines. The SVM implementation have relied on the LIBLINEAR Library[69], which is freely available at *http://www.csie.ntu.edu.tw/~cjlin/liblinear*. All the processes have been implemented in MATLAB R2011b[3](MathWorks Inc., Sherborn, MA). The classifier chosen for our simulations have been a linear C-SVM for many reasons. The first one is because, since the feature space dimensionality F (which is the number of features per subject) is normally higher than the examples space N (total number of subjects) in all the analysis, mapping the features in a space of higher dimensionality would not provide advantages. Moreover, during the analysis we have realized that the accuracy using linear SVM tended to be higher than when using more sophisticated kernels. Finally, the computational cost was also reduced comparing with the others.

4.3.1. Pre-processing steps

- **ICV normalization**. An ICV normalization pre-processing step is introduced as explained in Section 3.2. This step is skipped in case that no normalized volumes are required for the analysis.
- **Age Correction**. As explained in Chapter 2, several problems associated with the age-related effect arise when trying to perform AD, MCI and CTL classifications. A recent study performed by *Dukart et al., [17]* presented a method to control this age-related

[3] © 2012 The MathWorks, Inc. MATLAB and Simulink are registered trademarks of The MathWorks, Inc. See www.mathworks.com/trademarks for a list of additional trademarks. Other product or brand names may be trademarks or registered trademarks of their respective holders.

effect, showing promising results in CTL versus AD classification. Thus, in our method an Age Correction procedure is perfomed, as explained in Section 3.2, but with one important difference: instead of applying the regression voxel by voxel, we compute the correction per brain regions. To estimate the regression coefficients the whole database, on which the analysis is being performed, is used[4] in order to introduce as less bias as possible[5]. Therefore, this step is placed before splitting into *Training* and *Testing Set*[6].

- **Train Set and Testing Set**. The classifier has a number of parameters that have to be learned from the training data, which is a part from the whole dataset reserved for that purpose. The learned classifier will be a model of the relationship between the features and the class label in the training set. Finally, the perfomance of the created model is tested trying to predict another part of the data set, called test set, which should be, ideally, unseen data for the classifier.

  Thus, the next thing to do is to split the whole data set in:

  - *Training Set*: is used to tune absolutely all the parameters of the classifier, as explained in the next subsections.
  - *Pure Test Set*: is used to test the accuracy of the classifier. Never, and absolutely never, *touched* once separed from the original dataset. Otherwise, the final trained classifier will depend on this Test Set, what would lead to bias in the real overall perfomance. Thus, the reported results are directly what it is expected to obtain when used in a real clinical enviroment.

- **Data scaling**. For this work, a z-normalization have been applied to the examples in the training set. This way, it is assumed that all region attributes have zero mean and standard deviation one throughout this work. Of course it is very important to use the same criteria to scale both training and test set[37]. Is for that reason that the mean and standard deviation estimates used for the training set normalization are considered to be part of the machine, saved and applied to the pure test set to ensure that they are scaled consistently[9].

4.3.2. Parameter Setting

The procedure described below is shown in the Figure 4.3. This part is one of the most important steps in our work, as both features and the C parameter will be selected for the final pure testing. This procedure is based in what is known as nested cross-validation. First of all, one

[4] The age correction coefficients are computed directly with the database on which the experiments are being performed. For example, when working with ADNI 1 Database, this age-correction coefficients will be computed with this database.

[5] It has to be considered that the performance of a regression directly depends on the number of examples, and it do its task better when dealing with databases that tend to be Gaussian. Thus, the more subjects are included in the regression, the better performance is expected

[6] Ideally, the regression coefficients should be universal and well-standarized, but there are no studies about this matter in the recent literature. This issue will be discussed in Section 7.2.

$subject_i$ from the training set is selected (this step is repeated for all the subjects in the training set, what is known as cross-validation). Then, a decorrelation and RFE feature selection steps are performed to select the set of most discriminative features in this *reduced* train set (which is the whole train set without the previous selected individual). The threshold of the decorrelation step is fixed to 0.95 in order not to remove important information. The next step is to find the C that best predicts new data. To do so, a subject is selected from the *reduced* train set, that will be used to test, and the others are used to train a model with a determinate C. This process is repeated till all subjects in the *reduced* train set have been used to test, what is known as a nested[7] cross-validation. Then, the same procedure is repeated giving many values to C hyperparameter, what is known as Grid-Search[8]. Each nested cross-validation has a final accuracy, and the C that yield to the best accuracy is selected to train the model with the *reduced* train set. This *reduced* model will be used to predict the $subject_i$. This process is also repeated once per subject in the training set, and outcoming a cross-validation accuracy. Moreover, the number of features selected is chosen in the same way. Many iterations of the described loop are done, changing the number of selected features in the RFE. Finally, the number of selected features that yield to best cross-validation accuracies is selected to be used in the final model. Summarizing, the model is optimized in the way that follows:

- *Feature Selection*: try different number of selected features in the RFE. The one that yield to best cross-validation accuracies is selected. This behaviour is showed in the Figure 4.4 .
- *Parameter Tuning*: using this selected number of features, the most frequent selected C in the nested cross-validation will be used to train the final model.

4.3.3. Final Model Training and Pure Testing

The final global scheme can be seen in the Figure 4.5. Finally, the Train Set is used to train the final model, using the C hyperparameter and the F number of features, both optimized in the Model Tuning step. The F selected features are those top ranked in the cross-validation. Then, the labels of the pure test set are predicted and the final accuracy of the classifier is the result of the comparision of the predicted labels versus the true labels.

[7] It is considered a nested one beacuse it is included in a main cross-validation loop.

[8] Grid-Search refers to exhaustive searching process, giving values to the hyperparameter that is going to be optimized. Usually, it is done in a logarithmic way, i.e. $C = 1^{-7}, 1^{-6}, ..., 1^{-1}$.

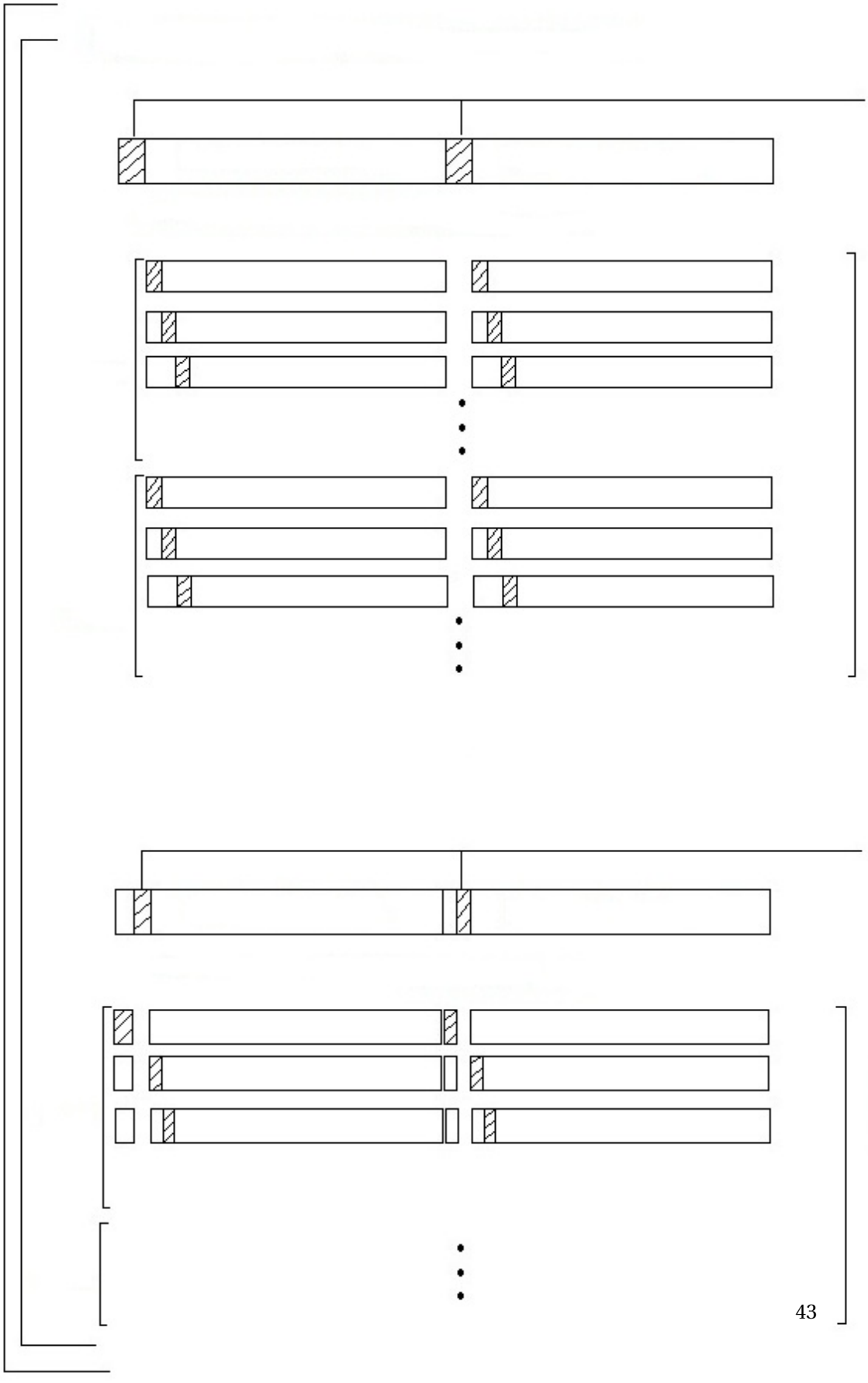


Figura 4.3: Tuning SVM Model Process: the output of this block is the optimum hyperparameter C and features to use to train the final model.

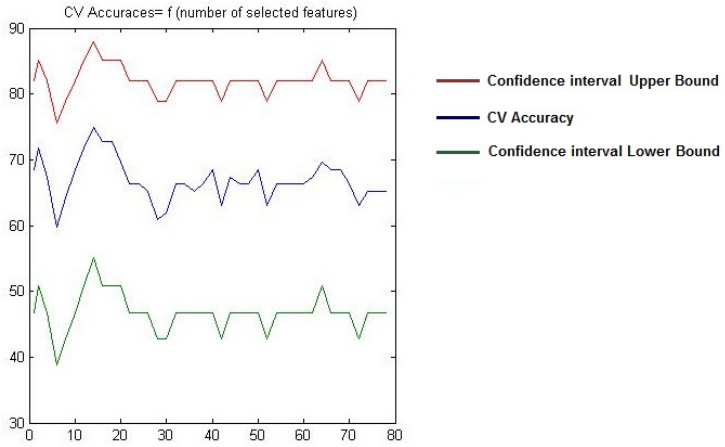


Figura 4.4: CV accuracy when changing the number of features. Expanded Database, GM Region Volumes+ GM Lobe volumes+ Asymmetry Coefficients.

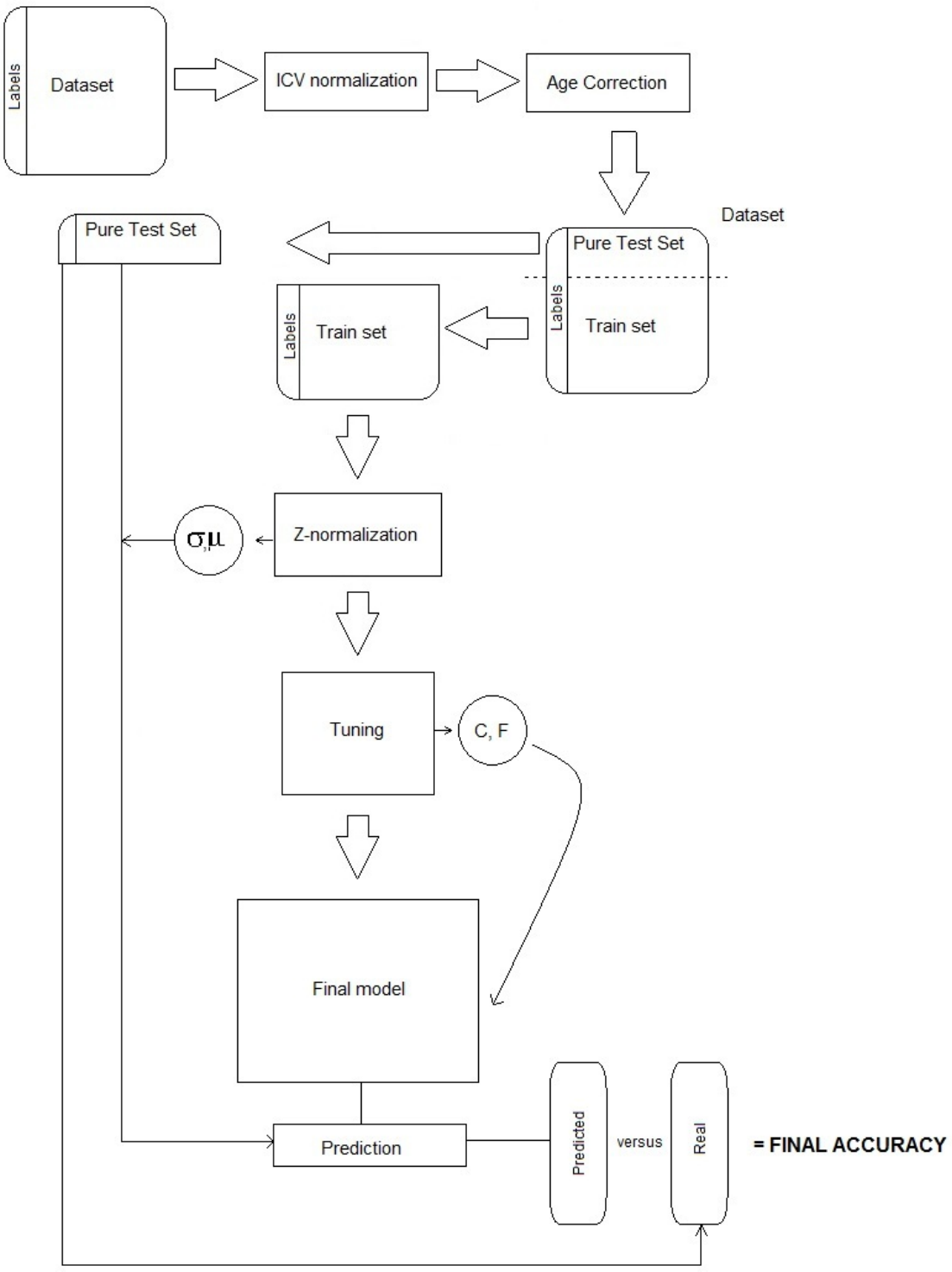


Figura 4.5: Summary of the complete implemented classification process

5 Pre-processing Analysis

In this Chapter, the pre-processing studies are presented. In the first Section, our evaluation method is explained. In the next Section, previous study pre-processing steps have been performed in order to compare our method to the literature and to assess whether some pre-processing steps are worthy or not.

5.1. Quantitative Evaluation

5.1.1. Accuracy, Sensitivity, Specificity

In this work 6 values describing the performance of the classifier are given for each analysis:

- *Sensitivity*: Is the rate of true positives (TP)[1]. For example, when performing a CTL versus MCI analyisis, the sensitivity is the rate of correctly classified MCI subjects[2]:

$$Senstivity = \frac{number\ of\ TP}{number\ of\ TP + number\ of\ FN} \tag{5.1}$$

- *Specificity*: Is the rate of true negatives (TN). For example, when performing a CTL versus MCI analyisis, the sensitivity is the rate of correctly classified CTL subjects[3].

$$Specificity = \frac{number\ of\ TN}{number\ of\ TN + number\ of\ FP} \tag{5.2}$$

- *Accuracy*: Is the total rate of correctly classified subjects, computed in the way that follows:

$$Accuracy = \frac{Sensitivity + Specificity}{2} \tag{5.3}$$

[1] TP: True Positives, FP: False positives, TN: True Negatives, FN: False Negatives.
[2] In a generic analyisis A versus B, the sensitivity is the rate of correctly classified B subjects.
[3] In a generic analyisis A versus B, the specificity is the rate of correctly classified A subjects.

In the present study, the accuracy, sensitivity and specificity are given both for the cross-validation and pure tests.

5.1.2. Results Confidence Intervals

As seen in the previous subsection, the accuracy of the classifier can be considered as the average accuracy of N tests, computed as follows:

$$Accuracy = \frac{1}{N}\sum_{n} C_n \tag{5.4}$$

where $C = \{1,0\}$ is a Bernoulli variable, being $C = 1$ the correct and $C = 0$ the wrong decisions. Thus, the accuracy is like a binomial variable, which can be approximed by a normal distribution.

Estimating the error and the confidence intervals (CI) in an observation is a crucial issue in statistics if one wants to make predictions about what is likely to happen when repeating the experiment any number of times[70]. The CI provides information about what is expected to result from a test, with a certain confidence level $(1-\alpha)$, $0 \leq \alpha \leq 1$. In other words, this interval is the range of values in between the variable is expected to be located, with a probability $1-\alpha$. In this work, the upper and lower confidence interval bounds for this binomial distribution have been computed using Wilson's score interval. The theory underlying this concept totally overcomes the purpose of this work. For further details see [70, 71].

5.2. Study pre-processing steps

5.2.1. Age Correction

It is widely accepted that there are GM alterations in different brain structures due to the normal ageing. We have studied this effect performing many analysis in different regions to verify this effect, for example in the Amygdala, the Hippocampus, the Vessels or the Inferior Lateral Ventricles, and all of them showed age-related effects. As example, the Hippocampus GM losses due to ageing are plotted in the Figure 5.1a. As explained in Section 3.2, the age-related effect has an undesired repercussion in the predictions as the classifier is not always able to distinguish among this and the disease-related effect. Is for that reason that the age-related effect is treated in our work in a similar way than [17], but for each region instead of voxel based. In the Figure 5.1b the effect of the regression is plotted and it can be seen that the slope due to ageing is almost cancelled when applying the correction.

Then, both classical LS and Robust Regressions have been studied in order to test out which performs better. The comparision is seen in the Figure 5.2. Although at theory the robust approach is expected to work better, at practice both Robust and LS regression have yielded to similar results. Thus, as it is not clear that the roubst regression can provide better results, the

Cuadro 5.1: Classification performance before and after the age correction with CTL versus AD analysis, with the ADNI1 Database.

	Without Age Correction	**With Age Correction**
Cross-Validation Accuracy	87.2 %	***88.7 %***
Sensitivity pure	86.2 %	***88.5 %***
Specificity pure	87.9 %	***89.1 %***
Accuracy pure	87.0 %	***88.8 %***

Cuadro 5.2: Classification performance before and after the age correction with CTL versus MCI analysis, with the ADNI1 Database.

	Without Age Correction	**With Age Correction**
Cross-Validation Accuracy	70.6 %	***71.4 %***
Sensitivity pure	67.4 %	***68.4 %***
Specificity pure	73.3 %	***75.1 %***
Accuracy pure	70.3 %	***71.8 %***

LS regression method explained in [17] has been used.

Once the type of regression have been chosen, two tests have been performed, both for AD and MCI, to check whether this technique is useful or not. Two SVM classification using region volumes have been launched, once with and once without correcting the age effect, using 155 subjects to train the classifier and 30 to test. This procedure have been repeated 100 times in order to get a low biased estimation of the true accuracies. The reported results are the average of the 100 iterations and are presented in the Table 5.1. Aplying the age correction resulted in an improvement of the accuracy in all cases, similarly to [17], which also reported a 2 % enhancement of the overall CTL vs AD classification using the same database. But there is more that only an improvement of the accuracy. When comparing the classification errors using a t-test on the misclassified subjects, as expected, there is a significant difference in mean age between missclassified CTL and AD subjects (P-value= 3.82 E-41), being the youngest AD and oldest CTL often misclassified, but there is not when correcting the age-related effects (P-value= 0.7505), with a significance threshold of $p \leq 0{,}05$. The age characteristics of the misclassified subjects is showed in the Figure 5.3.

Exactly the same procedure has been performed classifying CTL versus MCI. As far as we know, there are not other studies in the literature so any comparision can be done. As well as in the CTL versus AD case, there have been an improvement in the pure accuracy, but in this case of 1.5 %. The results are presented in the Table 5.2. Concerning to the missclassified subjects, although the difference in age means are statistically significative in both correcting and not correcting cases, the differences are lower in the corrected case (see Figure 5.4).

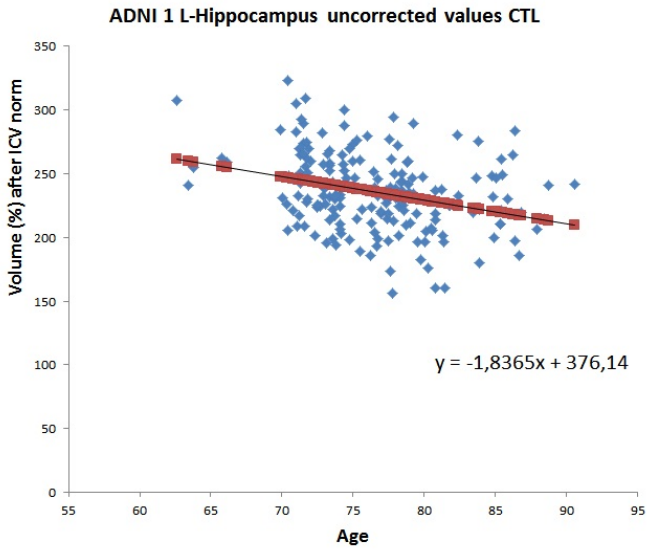
The same test have been performed with the Expanded Database and with Lausanne Database. The results for the first and the second one are a pure accuracy improvement of 3.8 % and 4.5 %

respectively. In both cases, the mean age differences across groups of misclassified subjects have been reduced, but there are still differences, exactly in the same way as reported in the ADNI1 comprovation presented above. In the Expanded Database, the *t*-test p-value is 0.0028 when correcting and 1.2354 E-73 when not correcting, while with the Lausanne Database the *t*-test p-value is 5.0003 E-005 when correcting and 3.5046 E-023 when not.

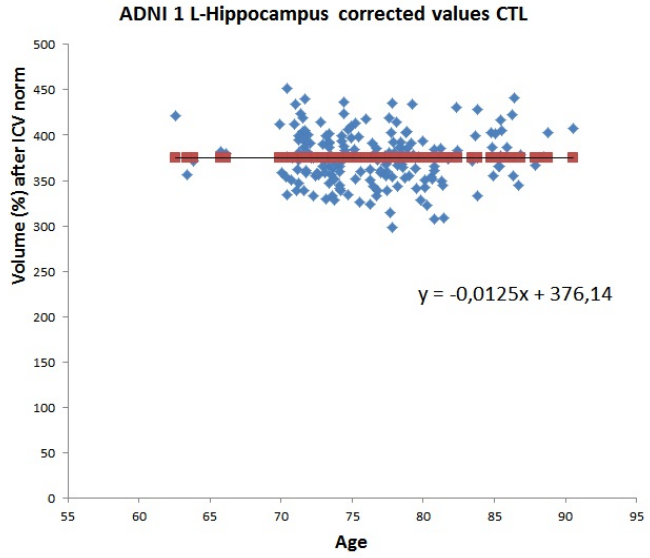
This comprovations show that it is useful to perfom a pre-processing age correction step before feeding the SVM. From now on, this step is applied for all the analysis.

5.2.2. Correlation

In the Section 3.2 it has been explained the importance of reducing the number of input features to the classifier. Perfectly correlated features do not provide more information when taken toghether instead of taking only one of them, as they are truly redundant. Is for that reason that a decorrelation step has been introduced just before feeding the RFE algorithm. In the Figure 5.5 it is shown the absolute value of the whole feature space correlation matrix; appreciate that there exist feature correlation (see red areas). As far as we know, there exisit no studies about this issue in the literature of brain MRI classification, so no previous assumptions can be done. Is for that reason, that a conservative 0.95 threshold has been used to remove correlated variables, in order to prevent deleting valuable information.

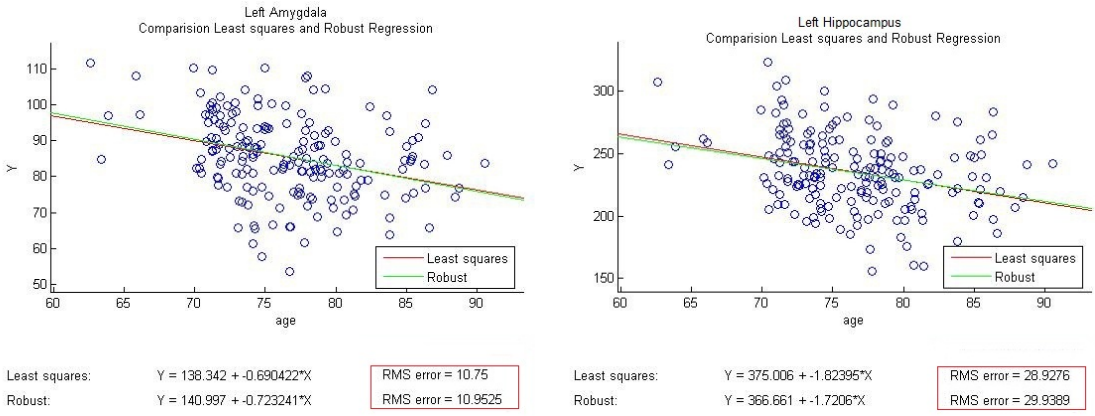


(a) Left Hippocampus volumes plotted against age before the correction.



(b) Left Hippocampus volumes plotted against age after the correction.

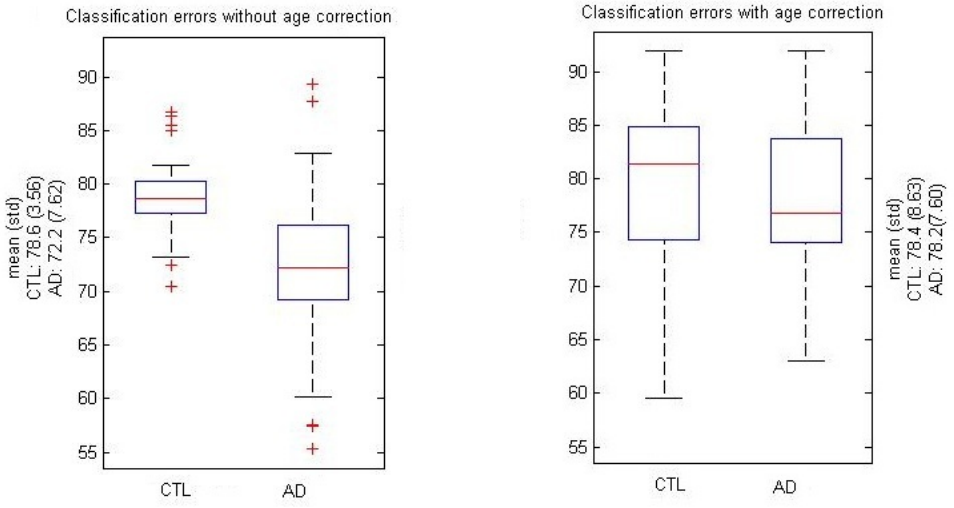
Figura 5.1: Comparision of the Left Hippocampus volumes plotted against age before (a) and after (b) the age correction in the CTL group.



(a) Left Amygdala

(b) Left Hippocampus

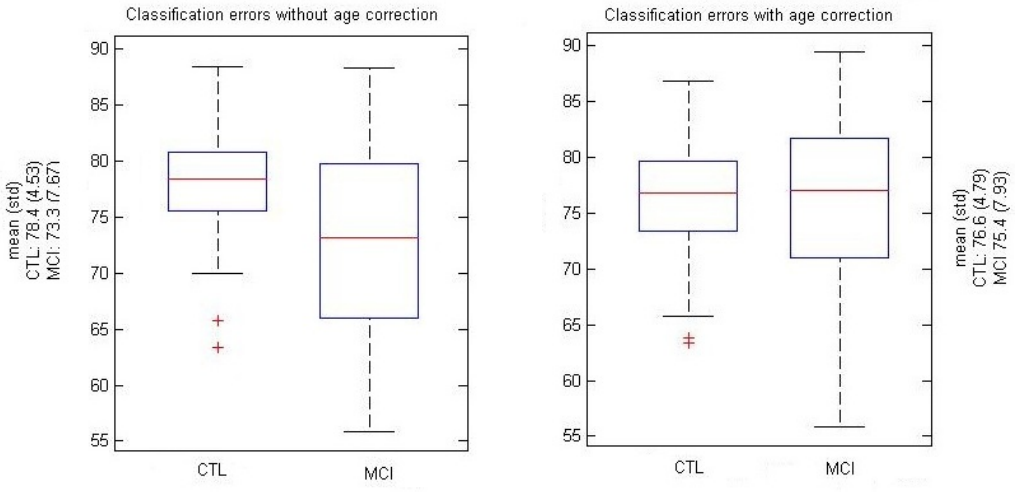
Figura 5.2: Comparision of LS and Robust regression in two brain regions: (a) left hippocampus and (b) left amygdala, with the ADNI1 Database.



(a) Classification errors without age correction. P-value= 3.82 E-41.

(b) Classification errors with age correction. P-value= 0.7505.

Figura 5.3: Age characteristics of the misclassified CTL and AD subjects with and without age correction, with the ADNI1 Database.



(a) Classification errors without age correction. P-value= 2.738 E-54.

(b) Classification errors with age correction. P-value= 0.0004.

Figura 5.4: Age characteristics of the misclassified CTL and MCI subjects with and without age correction, with the ADNI1 Database.

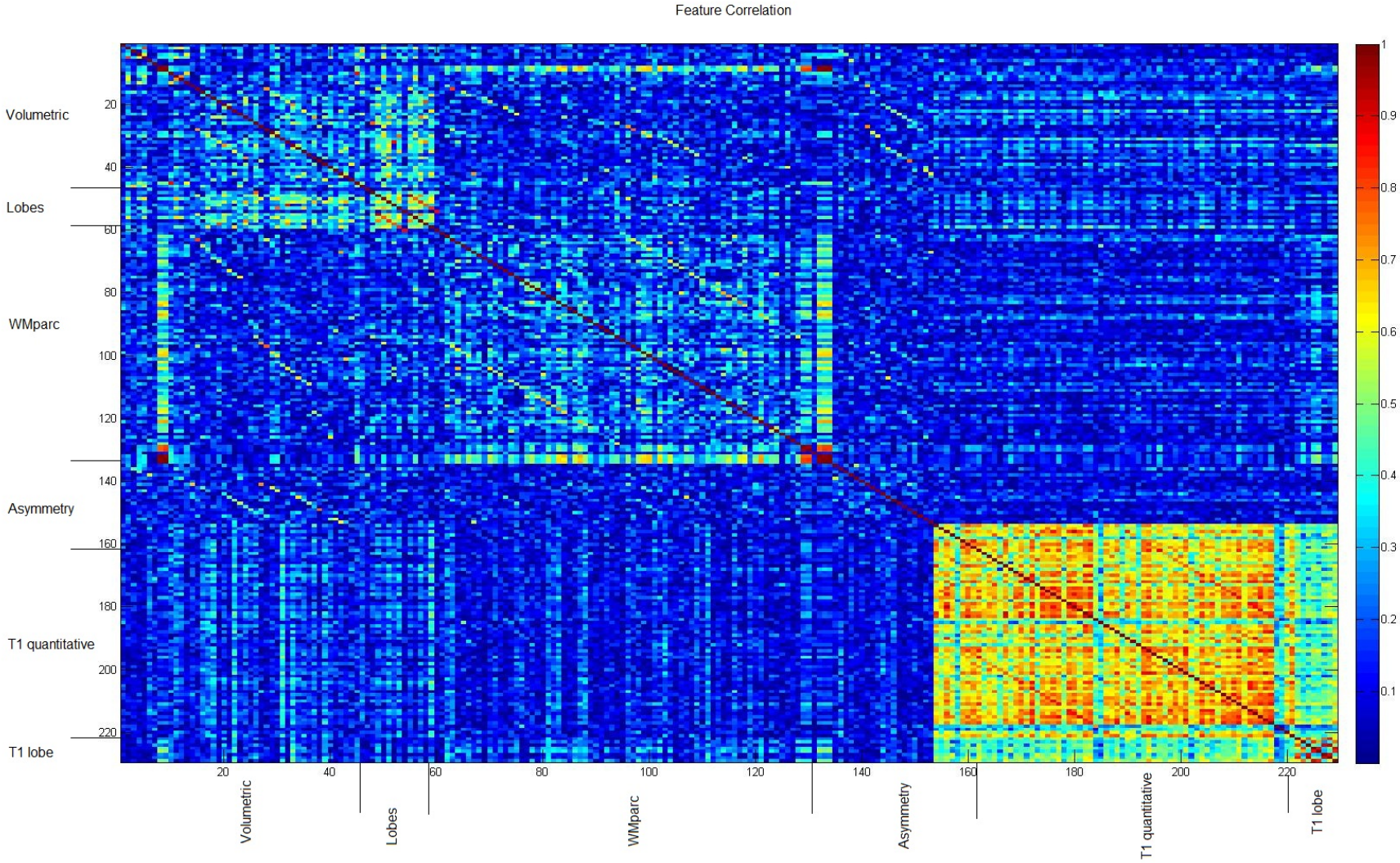


Figura 5.5: Feature Correlation with the Expanded Database.

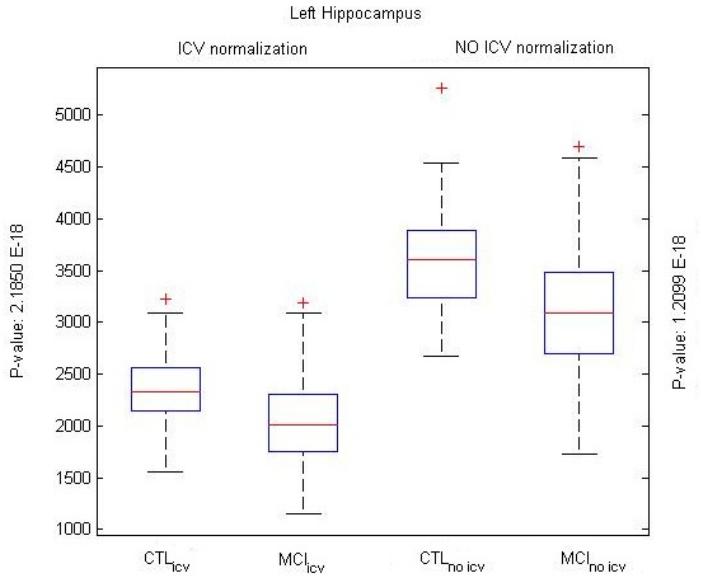


Figura 5.6: Group comparision of the Left Hippocampus volume when normalizing (left) and when not normalizing (right) by ICV, with the ADNI1 Database.

5.2.3. ICV

The ICV normalization is a commonly used step in the literature to reduce intersubject variablity due to the gender and the head size. In our opinion, this pre-processing step is useful when applied to large brain structures, but probably some useful information is lost when dealing with small structures. In order to asses this hypothesis, all the analysis have been performed twice: once normalizing by ICV and once without doing the normalization.

A preliminar test have been done, which consists in analyzing by a *t*-test the mean differences among groups of the whole feature set, comparing when normalizing by ICV and when not, using the CTL and MCI subjects of the ADNI 1 Database. The results are presented below:

- 39 out of 48 brain regions presented more differences in mean when normalizing by ICV.
- 9 regions presented more differences in mean when not normaliying by ICV: Right parahippocampal, 3rd Ventricle, Right Rostral Anterior Cingulate, White Matter Hypointensities, Left Hippocampus, Left Medial Orbitofrontal, Left Temporal Pole, Left WM No Hypointensities and Right Caudal Anterior Cingulate. Nevertheless, only the differences in the Left Hippocampus and the Left Temporal Pole are statistically significant, and so they are when applying ICV normalization.

Although an SVM model has to be trained to find out the real feature weights for the classification, this results suggest that more class separation could be achieved when normalizing by ICV as there are more statistical differences among groups. However, the fact that *usual*

suspects like the Left Hippocampus or the Left Temporal Pole present more differences when not normalizing reaffirms our hypothesis that some information can be lost when applying this pre-processing step. Thus, all the results are presented twice, once normalizing and once not normalizing.

5.2.4. Comparision Global Performance with the ADNI1 Database and literature results

Finally, three analyisis have been performed: CTL versus AD, CTL vs MCI and MCI vs AD SVM classifications. The purpose is to compare our methodology to the literature when the same database is used. In each classification, 155 subjects have been used to train the model and 30 to test its performance, applying the procedure explained in Section 4.3. Let us recall that only region volumes have been used for this analysis (see Chapter 4.2). The results are presented in the Table 5.3. The obtained pure results are, accuracy (CI=lower bound-upper bound):

- CTL vs AD: 88.3 % (77.8 %- 94.2 %)
- CTL vs MCI: 81.7 % (70.1 %-89.4 %)
- MCI vs AD: 60.0 % (47.4 %-71.4 %)

In all the cases the best results have been got when normalizing by ICV.

Those preliminary studies were required to check the overall perfomance and to compare our method to the methods explained in the literature. The obtained results are consistent with the accuracies and disriminative features reported by other studies. An accuracy of 90 % is what we expected from the CTL vs AD classification. An accuracy higher than 80.0 % for CTL vs MCI was not expected for the volumetric univariate analysis. However, we suspect that classification results over 80 % when dealing with MCI are near the upper confidence interval bound. Thus, the typical expected result would be found in the range 70.0 to 80.0 %, being lower and upper bound respectively.

The CTL vs AD results are comparable to those obtained in the literature and presented in Chapter 2. On the one hand, *Fan et al., [12]* have done a region based approach using the ADNI database, and the CV results are worse (82.0 %) than those obtained with our method. On the other hand, many VB approaches have been done presenting similar results as in our work: [5, 9, 17], which have obtained pure accuracies of 88.0 % , 89.0 % and CV accuracy 85.0 % respectively. Moreover, two VB studies presented better accuracies than those in our work: [4] and [12], which have outcome accuracies of 96.0 % and 94.3 % respectively. The results presented by the first one are extremely good but, in our opinion, some considerations should be taken into account. First of all, the fact that tiny databases have been used for the analyisis do not let to figure out what is really expected to happen when using other bigger databases. Another important point is that the MMSE scores of the AD subjects are significatively lower

than those from other CTL vs AD studies, what means that those subjects are more severely diseased and, therefore, easier to classify. Refering to [12], there is only one possible remark to do: the CTL group is more than two years younger than the AD group. As the AD atrophy pattern advances in the same way as age does, it probably introduces more differences across groups and would make easier the classification task to the SVM. In our opinion, this results are probably the upper bound of the accuracy that is expected to result from CTL vs AD classifications. Refering to the most discriminative features for the CTL vs AD classification, the top ranked ones are: Left Hippocapus, Left Entorhinal, Right Middle Temporal, Left Inferior Lateral Ventricle, Right Caudal Anterior Cingulate, Left Amygdala, Left Isthmus Cingulate, Right Hippocampus, Right Amygdala and Right Fusiform. These results are consistent with the regions obtained in the literature, as many studies have reported significative differences in the Hippocampus[1, 2, 5, 12], the Entorhinal Cortex[5, 12], the Amygdala[1, 2, 5, 12] and the ventricles[2], as well as other structures.

The CTL vs MCI results are also comparable to [1, 12], but showing slightly lower results than those in our work, 71.0% and 76.0% CV respectively. Also three VB based approaches have been presented in the literature: [5, 12], that obtained accuracies of 81.5% (pure accuracy) and 81.8% (CV accuracy) respectively are higher than those obtained in our work. One point should be taken into consideration when analyzing the results from [5]: the MCI subjects used for the classification had converted to AD within 18 months, and MCI AD-converter group is expected to present more differences respect the CTL group than a non-converter MCI group. Moreover, doing so, the MCI group becomes more homegenious. It has to be taken into account that the MCI group is, by far, more variable than the CTL and the AD groups, and the classification results strongly depend on the homogeneity of the classification groups.

Refering to the top ranked features for the CTL vs MCI classification, the most discriminatives have been: Left Hippocampus, Left Amygdala, Right Hippocampus, Left Middle Temporal, Right Amygdala, Right Entorhinal, Left Entorhinal, Right Temporal Pole, Left Caudal Middle Frontal and Right Accumbens Area. These results are consistent with the regions obtained in the literature, as different studies have reported differences, for instance, in the Hippocampus[2, 5, 12] or in the Amygdala[2, 5].

In the Appendix B, the complete lists of features sorted by its discriminative power are presented.

Cuadro 5.3: Classification results using the ADNI 1 Databse. The results are presented as *Accuracy pure [Condifence interval: lower bound- upper bound](Accuracy Cross-Validation).*

Test		Sensitivity pure (CV)	Specificity pure(CV)	Accuracy pure [CI] (CV)
CTL vs AD	ICV	86.7 % (91.0 %)	90.0 % (89.0 %)	**88.3 %** [77.8- 94.2] (90.0 %)
	No ICV	80.0 % (90.7 %)	93.3 % (90.3 %)	**86.7 %** [79.9- 95.3](90.5 %)
CTL vs MCI	ICV	80.0 % (70.3 %)	83.3 % (76.1 %)	**81.7 %** [70.1- 89.4] (73.2 %)
	No ICV	76.7 % (72.9 %)	66.7 % (79.4 %)	**71.7 %** [59.2- 81.5] (76.2 %)
MCI vs AD	ICV	66.7 % (74.8 %)	53.3 % (67.1 %)	**60.0 %** [47.4- 71.4] (71.0 %)
	No ICV	66.7 % (72.9 %)	53.3 % (65.8 %)	**60.0 %**[47.4- 71.4] (69.4 %)

6 Results

Results for the Lausanne Database are presented in first section, while the Expanded database is considered in the second section. Then, the Multivariate analyisis is presented using again the Lausanne Database. Finally, a stacking strategy is presented.

6.1. Analysis Lausanne Database

First, 4 univariate analyisis have been performed, feeding the classifier with GM region volumes, lobe volumes, asymmetry and WM volumes separately. Second, we tested whether adding multiscale data enhances the performance of the classifier. Third, the classifier has been fed combining multivariatre measures. The training have been performed using 21 subjects, while 8 subjects have been used for the pure testing. The results are presented in the Table 6.1 and discussed below.

The best pure accuracy when doing a univariate approach has been got when using the classical regional GM volume data, which yielded to accuracies [CI] of 62.5 % [38.7-81.5 %] when normalizing and 68.8 % [44.4-85.9 %] when not normalizing by ICV. The lobe volumes, the asymmetry and the WM volumes outcame accuracies of 62.5 % [38.7-81.5 %], 56.3 % [33.2-76.9 %] and 62.5 % [38.7-81.5 %] (when not normalizing by ICV) respectively. Multiscale information did not improve the accuracy when combined with GM region volumes but it did when added to WM volumes, up to 68.8 % [44.4-85.9 %] and 75.0 % [50.5-89.8 %] when normalizing and when not normalizing by ICV. Regarding to when adding the asymmetry coefficients, the best results were a 75.0 % [50.5-89.8 %] of accuracy when combined with GM regional volumes, both normalizing and not normalizing by ICV. Finally, when adding different variables and scales, the best results have been gotten when combining regional GM and Lobe volumes and the asymmetry coefficients, which yielded to a pure accuracy of 75.0 % [50.5-89.8 %] both normalizing and not normalizing by ICV. Only 5 cases have passed the chance threshold (more than 50.0 % in the lower interval confidance bound), which are WM+ lobes (no ICV), Cortical GM Vol+ Asy (ICV & no ICV) and Cortical GM Vol+ ASY+ Lobe (ICV & no ICV).

The classification process outcomes the pure and CV accuracy, sensitivity, specificity and complete list of features ranked by its discriminative power. An example of this list[1] can be seen in the Figure 6.1, which refers to the GM regional volumes classification with ICV normalization. The x axis (numbered) refers to the importance of the feature in the classification (being 1 the most important, 2 the second most important and so on), and the colours refer to how many times this feature have been placed in this position. In this figure it is shown that the most discriminative features are Right and Left Entorhinal, the Right and Left Hippocampus, the Right Inferior Parietal, the Righ Inferior Lateral Ventricle, the Left Caudal Middle Frontal and the WM Hypointensities. These results are consistent with previous studies in the literature.

Finally, considering all the performed analysis, those features that were shown as top-ranked for the classification are:

- **GM regional volumes**: Right and Left Entorhinal, Left and Right Hippocampus, Right Inferior Parietal, Right Inferior Lateral Ventricle, Left Caudal Middle Frontal, WM Hypointensities, Left Fusiform and Left Parahippocampal.
- **GM lobe volumes**: Right Limbic and Right Frontal.
- **Asymmetry**: Caudal Anterior Cingulate, Fusiform and Lateral Ventricle.
- **WM**: Right entorhinal, Right Paracentral, Right Precentral, Left Caudal Middle Frontal, Left Rostral Anterior Cingulate, Left Transverse Temporal, Left Rostral Anterior Cingulate and Left Middle Temporal.

These features are consistent with those found in the ADNI1 Database and in the literature, that also found the Hippocampus and the Entorhinal, for instance, as two of the most discriminative. In regard to the volumetric analysis, apart from those regions, the Inferior Parietal was cited as important to the classification by [3, 12], the Inferior Lateral Ventricle by [2] and the Caudal Middle Frontal by [3]. As to the Lobe Volumes, discrimination power of the Frontal Lobe was also reported by [1]. It was unexpected not to find the Amygdala as one of the most discriminant as our study on the ADNI1 Database and a lot of studies in the literature have reported changes in this structure in MCI[2, 3, 5, 13, 16]. The interpretation of this results is not easy and requires further studies. In our opinion, this result can be due to the differences across the Lausanne and the ADNI 1 Database. Apart from the mean age, which is corrected, there is another important difference: the MMSE mean score of the MCI Lausanne Database participants was almost one point higher than the ADNI1 MCI MMSE mean score[2]. It means that the Lausanne Database MCI individuals are less impaired than the ADNI 1 Database MCI individuals, and therefore, in an earlier stage of the disease. The fact that the Amygdala is not so affected in this database could mean that this structure is affected in a more advanced stage. Nevertheless, this issue requires further investigation to determinate its real cause.

[1] These output figures are shown in the Appendix B.2.

[2] This difference in MMSE mean score is statistically significant at level $p \leq 0{,}01$.

Finally, regarding to the ICV normalization, it is not clear whether it can yield to better accuracies or not, as similar results are outcome from the ICV and no ICV classifications.

Cuadro 6.1: Classification results using the Lausanne Databse. The results are presented as *Accuracy pure (Accuracy Cross-Validation) [Condifence interval: lower bound- upper bound]*. GM Region Volumes is refered as Vol, Lobe Volumes as Lobe and the Asymmetry coefficients as ASY.

Input		Sensitivity pure (CV)	Specificity pure(CV)	Accuracy pure [CI] (CV)
Cort GM Region Vol	ICV	62.5% (61.9%)	62.5% (66.7%)	**62.5%** [38.6- 81.5] (64.3%)
	No ICV	62.5% (61.9%)	75.0% (76.2%)	**68.8%** [44.4- 85.8] (69.0%)
Lobe Volumes	ICV	62.5% (61.9%)	62.5% (71.4%)	62.5% [38.6- 81.5] (66.7%)
Asymmetry	-	37.5% (57.1%)	75.0% (71.4%)	56.3%[33.2- 76.9] (64.3%)
WM volumes	ICV	62.5% (71.4%)	50.0% (71.4%)	56.3% [33.2- 76.9] (71.4%)
	No ICV	75.0% (66.7%)	50.0% (61.9%)	62.5% [38.6- 81.5](64.3%)
Vol+Lobe	ICV	50.0% (52.4%)	62.5% (66.7%)	56.3%[33.2- 76.9] (59.6%)
	No ICV	62.5% (47.6%)	75.0% (66.7%)	68.8% [44.4 - 85.8](57.2%)
WM+Lobe	ICV	75.0% (71.4%)	62.5% (76.2%)	**68.8%** [44.4- 85.8] (73.8%)
	No ICV	75.0% (71.4%)	75.0% (61.9%)	**75.0%** [50.5- 89.8] (66.7%)
Vol+ASY	ICV	75.0% (61.9%)	75.0% (57.1%)	**75.0%** [50.5- 89.8] (59.5%)
	No ICV	75.0% (57.0%)	75.0% (66.7%)	**75.0%**[50.5- 89.8] (61.9%)
Vol+WM	ICV	50.0% (61.9%)	62.5% (71.4%)	56.3%[33.2- 76.9] (66.7%)
	No ICV	50.0% (52.4%)	62.5% (74.2%)	56.3% [33.2- 76.9] (63.3%)
WM+ASY	ICV	50.0% (76.2%)	50.0% (66.7%)	50.0% [28.0- 72.0] (71.5%)
	No ICV	75.0% (71.4%)	62.5% (61.9%)	68.8%[44.4- 85.8] (66.7%)
Vol+WM+ASY	ICV	50% (61.9%)	62.5% (66.7%)	56.3%[33.2- 76.9] (64.3%)
	No ICV	62.5% (57.1%)	75.0% (76.2%)	68.8% [44.4 - 85.8](66.7%)
WM+ASY+Lobe	ICV	50.0% (71.4%)	50.0% (71.4%)	50.0% [28.0- 72.0] (71.4%)
	No ICV	50.0% (57.1%)	62.5% (66.7%)	56.3% [33.2- 76.9](61.9%)
Vol+ASY+Lobe	ICV	75% (47.6%)	75.0% (66.7%)	**75.0%**[50.5- 89.8] (57.2%)
	No ICV	75.0% (52.4%)	75.0% (66.6%)	**75.0%** [50.5- 89.8] (59.5%)
All	ICV	75.0% (71.4%)	62.5% (76.2%)	68.8% [38.6- 81.5] (73.8%)
	No ICV	25.0% (52.4%)	62.5% (71.4%)	43.8% [23.1- 66.8](61.9%)

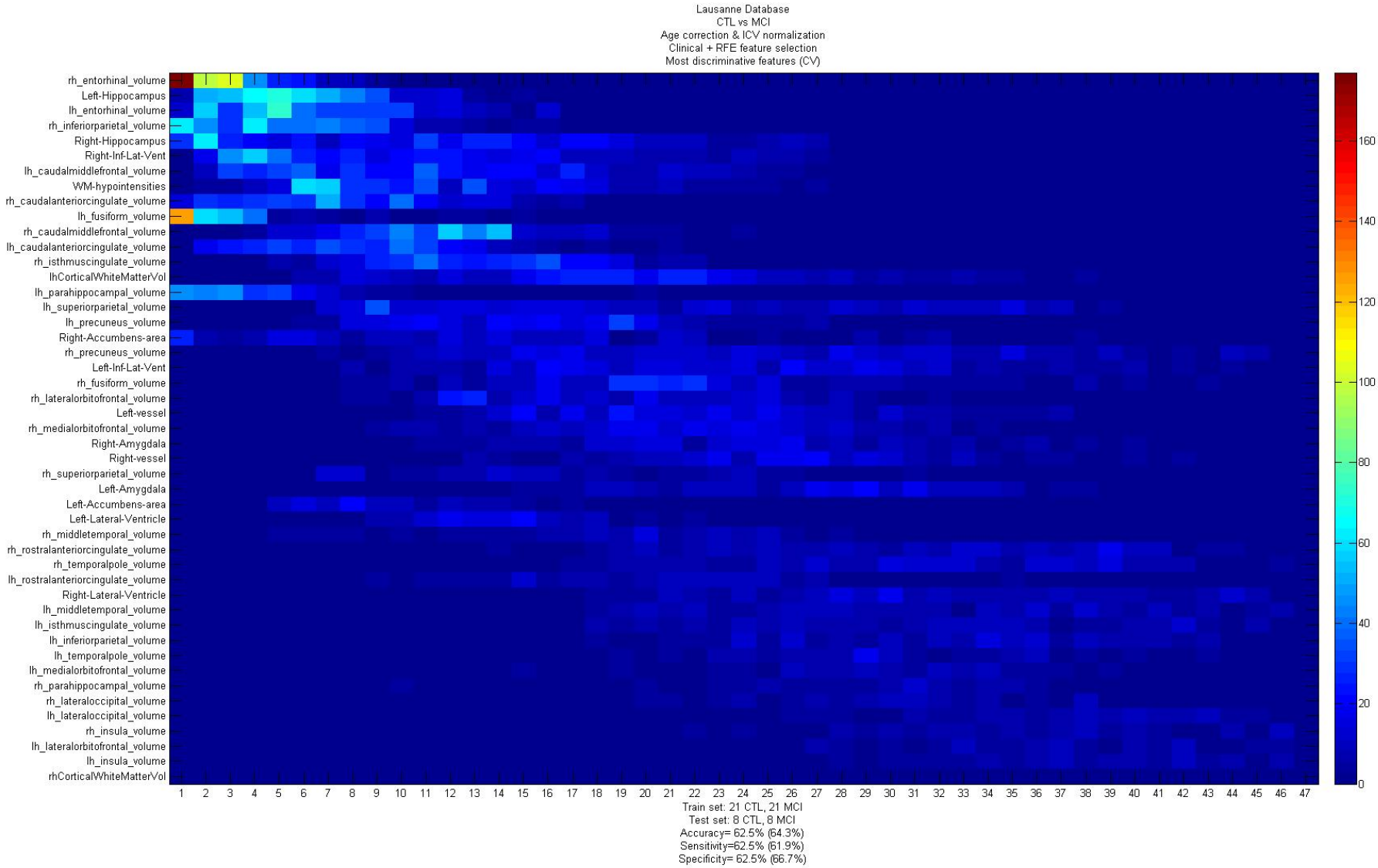


Figura 6.1: Output Classification Figure. Lausanne Database most discriminative Features: GM region volumes with ICV

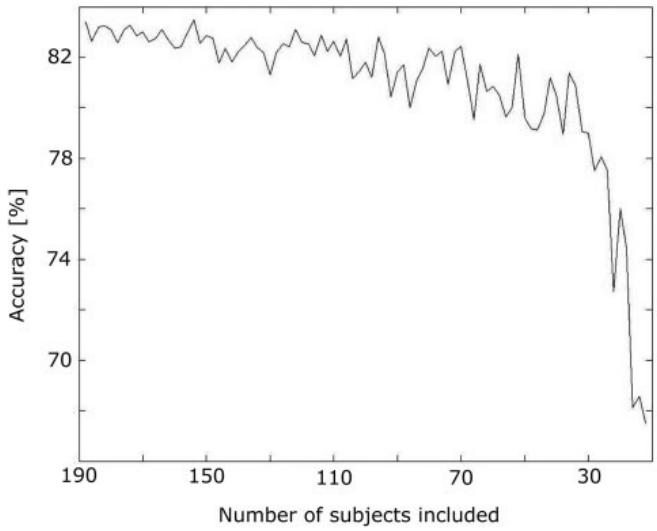


Figura 6.2: Accuracy behaviour in CTL vs AD classification when reducing the number of subjects per class, got from [59].

These results suggest that when combining different input variables, higher accuracies can be achieved. Even so, some considerations should be taken into account. First of all, it can be seen that the results are slightly lower than those obtained in the literature. This may be due to many reasons. First of all, the fact that the MMSE score of the MCI participants is higher than in the ADNI 1 Database makes the classification task harder, as they are less imparied and more difficult to distinguish from the CTL group. Moreover, we have realized that the accuracy results were very sensitive to which individuals were selected for the training and the testing set. This is probably due to the fact that the database is quite small. *Klöppel et al., [4],[59]* also reported variable results when reducing the number of subjects, as well as a worsening of the accuarcy. This accuracy behaviour can be seen in the Figure 6.2. Appreciate that the accuracy dramatically goes down and becomes more variable when reducing the number of subjects. We have also checked what happens if reducing the database. To do so, only 29 subjects per class have been left in the ADNI 1 Database. The CTL vs MCI GM volumetric classification process overcame a pure accuracy of a 75 %, 6.7 % points lower than the case when taking the whole database (185 individuals per class). Also the CV accuracy has gone down to 66.7 %, 6.5 % lower. The results when using this reduced database are presented in the Table 6.2 and are compared to the Table 5.3.

Cuadro 6.2: Comparison of the classification results using the ADNI 1 Databse with 29 subjects and GM volumetric features normalized by ICV. The results are presented as *Accuracy pure [Condifence interval: lower bound- upper bound](Accuracy Cross-Validation).*

Test		Sensitivity pure (CV)	Specificity pure(CV)	Accuracy pure [CI] (CV)
CTL vs MCI	29 subjects	75.0 % (71.4 %)	75.0 % (61.9 %)	**75.0 %** [50.5- 89.8] (**66.7 %**)
CTL vs MCI	185 subjects	80.0 % (70.3 %)	83.3 % (76.1 %)	**81.7 %** [70.1- 89.4] (**73.2 %**)

Thus, although the results suggest that higher accuarcy is expected when adding more variables to the analysis rather than doing univariate approaches, further analysis should be required to confirm this tendency. To overcome the problems that arise when dealing with small databases, we decided to expand the Lausanne Database, adding 29 subjects from each class as explained in Section 4.1, and run the same process explained above. The results are shown in Section 6.2.

6.2. Analysis Expanded Database

In this Section, the Expanded Database has been used to perform the analysis in exactly the same way as in Section 6.1. The results are presented in the Table 6.3 and summarized in the paragraphs below.

As with the Lausanne Database, the best results when doing a univariate approach have been gotten when using the Cortical GM Region Volumes, but in this case when not normalizing by ICV, which yielded to a pure accuracy [CI] of 70.9 % [50.8- 85.4]. The best CV accuracy has been outcome when using no ICV normalized GM Region Volumes: 77.2 %. Adding multiscale information, i.e., lobe volumes, raised the pure accuracy of the ICV normalized GM Region Volume classification to 75.0 % [55.1- 88.0] but did not when not normalizing. The pure results were not improved for both WM analysis, although the CV accuracy raised. When combining different variables, the Asymmetry improved the accuracy to 75.0 % [55.1-88.0] when added to the ICV normalized GM Region Volumes. Also combining WM and GM Region Volumes yielded to better accuracies than when taken them alone: 70.9 % [50.8- 85.1] of pure accuracy both when normalizing and not normalizing. Moreover, the CV accuracies are higher than those obtained when taken the variables alone, suggesting that some robustness is added if the two variables are combined. Finally, combining both different scales and variables yielded to better accuracies than when taking only separate variables. The best results were obtained when combining GM Region Volumes, Lobe Volumes and Asymmetry, that yielded to 75.0 % [55.1-88.0] and 70.9 % [50.8- 85.1] of pure accuracy respectively when normalizing and when not normalizing by ICV. Moreover, the CV accuracies of both cases were 75.0 % and 77.2 %, what shows that robustness is added to the classifier. Finally, it is important to appreciate that 11 cases in this study have overcome the chance threshold in the pure testing, which represents an improvement respect the analysis with the Lausanne Database. The confidence interval has also become thinner, due to the fact that more subjects are considered for this pure testing.

Regarding to the most discriminative features for the classification, considering all the performed analysis, those features that were shown as top-ranked are:

- **GM regional volumes**: Left and Right Hippocampus, Left and Right Entorhinal, Right Precuneus, Left Amygdala, Left and Right Inferior Lateral Ventricles, Right Middle Temporal, Right Superior Parietal and Right Temporal Pole.
- **GM lobe volumes**: Left Limbic, and Left and Right Temporal.
- **Asymmetry**: Precuneus and Amygdala
- **WM**: Left Cortical White Matter, Left and Right Entorhinal, Right Precuneus, Left and Right Superior Temporal, Right Middle Temporal, Right Superior Frontal and Right Insula.

Appreciate that these features are not the same that those obtained in the Lausanne Database. As the individuals are not the same, we did not expected that the most discriminative features exactly match either. However, the very most discriminative features are the same: Left and Right Hippocampus and Entorhinal and Inferior Lateral Ventricles. Moreover, other features that do not appear before have been added to the list, for instance the Left Amygdala, the Precuneus or the Right Temporal Pole. The findings on the precuneus are consistent with [3, 12, 13, 16] that also reported MCI-related changes on this structure. So they are the findings with the Amygdala[2, 3, 5, 13, 16]. For further region study, the complete set of output classification figures can be found in the Appendix B.3.

With regard to the ICV normalization, it is quite difficult to determinate which is the best option, although the accuracy is normally slightly higher when normalizing by ICV than when not normalizing. There is still one thing to add refering to the ICV normalization. In the no ICV normalization GM Region Volumes analysis, the best CV results have been obtained when using only 2 features, so the classifier have used only this features to predict the pure test. Those two features were the Left and the Right Hippocampus, which yield to a pure accuracy [CI] of 70.9 % [50.8- 85.1] and a CV accuracy of 77.2 %. Those are great results with only two structures, and it probably would mean that this regions have more discriminative power when not normalizing than when normalizing by ICV. This findings are consistent with the results presented in Subsection 5.2.3, in which more differences have been found in the Left Hippocampus when not normalizing by ICV.

Again, the best accuracies have been obtained when combining different variables instead of when taking them alone. We must be though very prudent with these results but, as the tendency is the same as the results outcome with the Lausanne Database, it suggests that combining variables from different scales (lobe volumes) and sources (asymmetry or WM) would provide robustness and better overall performance. Despite the better results when combining variables, some considerations should be taken into account, as they are certain factors that can add variability to the database. The expanded database was composed by joining two

different databases that come from two different sources (ADNI 1 and Lausanne Database). Moreover, the images were taken with different scanners and with different magnetic field strength (1.5T and 3T). Is for that reason that we have been very rigurous when composing the database, and we have created well-balanced groups, as explained in Chapter 4.1, in order to remove, as much as possible, all this possible adverse effects.

Cuadro 6.3: Classification results using the Expanded Database. The results are presented as *Accuracy pure (Accuracy Cross-Validation) [Condifence interval: lower bound- upper bound]*. GM Region Volumes is refered as Vol, Lobe Volumes as Lobe and the Asymmetry coefficients as ASY.

Input		**Sensitivity pure (CV)**	**Specificity pure(CV)**	**Accuracy pure [CI] (CV)**
Cort GM Region Vol	ICV	66.7% (63.0%)	66.7% (82.6%)	66.7% [46.7- 82.0] (72.8%)
	No ICV	50.0% (71.7%)	91.7% (82.6%)	**70.9%** [50.8- 85.1] (77.2%)
Lobe Volumes	ICV	50.0% (60.9%)	83.3% (67.4%)	66.7% [46.7- 82.0] (64.2%)
Asymmetry	-	50.0% (60.9%)	50.0% (67.4%)	50.0%[31.4- 68.6] (64.2%)
WM vol	ICV	65.2% (58.3%)	71.7% (58.3%)	68.5% [46.7- 82.0] (58.3%)
	No ICV	50.0% (65.2%)	58.3% (60.9%)	54.2% [35.1- 72.1](63.1%)
Vol+Lobe	ICV	75.0% (60.9%)	75.0% (87.0%)	**75.0%**[55.1- 88.0] (74.0%)
	No ICV	50.0% (71.7%)	91.7% (82.6%)	70.9% [50.8 - 85.1](77.2%)
WM+Lobe	ICV	33.3% (71.7%)	91.7% (69.6%)	62.5% [42.7- 78.8] (70.7%)
	No ICV	33.3% (69.6%)	75.0% (67.4%)	54.2%[35.1- 72.1] (68.5%)
Vol+ASY	ICV	66.7% (67.4%)	83.3% (76.1%)	**75.0%** [55.1- 88.0] (71.8%)
	No ICV	50.0% (71.7%)	91.7% (82.6%)	70.9% [50.8- 85.1] (77.2%)
Vol+WM	ICV	58.3% (71.7%)	83.3% (80.4%)	**70.9%** [50.8- 85.1] (76.1%)
	No ICV	50.0% (71.7%)	91.7% (82.6%)	**70.9%** [50.8- 85.1] (77.2%)
WM+ASY	ICV	58.3% (65.2%)	58.3% (71.7%)	58.3% [38.8- 75.5] (68.5%)
	No ICV	66.7% (73.9%)	50.0% (58.3%)	58.3% [38.8- 75.5] (66.1%)
Vol+WM+ASY	ICV	58.3% (71.7%)	75.0% (82.6%)	66.7%[46.7- 82.0] (77.2%)
	No ICV	50.0% (71.1%)	91.7% (82.6%)	70.9% [50.8- 85.1] (66.7%)
WM+ASY+Lobe	ICV	33.3% (71.7%)	83.3% (73.9%)	58.3% [38.8- 75.5] (72.8%)
	No ICV	50.0% (71.7%)	66.7% (69.6%)	58.3% [38.8- 75.5](70.7%)
Vol+ASY+Lobe	ICV	66.7% (69.6%)	83.3% (80.4%)	**75.0%**[55.1- 88.0] (75.0%)
	No ICV	50.0% (71.7%)	91.7% (82.6%)	70.9% [50.8- 85.1] (77.2%)
All	ICV	58.3% (73.9%)	75.0% (78.3%)	66.7% [46.7- 82.0] (76.1%)
	No ICV	50.0% (71.7%)	91.7% (82.6%)	70.9% [50.8- 85.1] (77.2%)

6.3. Analysis Lausanne Database: Multiscale and Multivariate Approach

We added T_1-quantitative MRI data to the classifier as explained in Section 4.1 to test whether combining information from different image modalities can improve the overall perfomance of the classifier. We have used the Lausanne Database as ADNI 1 do not provide this kind of data. The same simulations than in Section 6.1 and 6.2 have been run, adding the T_1-quantitative values per region and per lobe first separately and then together. The results are presented in the Table 6.4, 6.5, 6.6 and explained below.

Taking T_1 or T_1-lobe separately yielded to pure accuracies of 56.3 % [33.2- 76.9], which are significantly lower than when only taking the GM region volume results. Moreover, any combination of them did not improve the maximum classification rate of 75 % obtained in Section 6.1. The detailed analysis is presented below:

- Adding T_1-lobe data improved the pure accuracy of 7 cases, worsened the accuracy of 10 cases and mantained the accuracy in 5 cases, comparing with Section 6.1. Three cases achieved the maximum classification rate, 75 %, which are WM+Lobe+ T_1-lobe, Vol+WM+T_1-lobe and Vol+WM+ASY+T_1-lobe.
- Adding Regional T_1 data improved and worsened the pure accuracy of 9 cases, while 4 cases mantained the classification rates, comparing with Section 6.1. In this analysis, any case reached the 75 %.
- Adding Regional T_1 and T_1-lobe data together improved the pure accuracy of 6 cases, worsened the pure accuracy of 14 cases and mantained the pure classification rates of 4 cases, comparing with Section 6.1. Again, none of them reached the 75 %.

Refering to the T_1-quantitative most discriminative features, as far as we know, there are no studies in the literature that have used this image modality, so any comparison can be done. We have found that there are some that are normally top-ranked, which are listed below:

- Lobe T_1: Parietal GM and Parietal WM.
- Regional T_1: Right Bankssts, Left Entorhinal, Left Frontal Pole, Left Parsopercularis, Left Medial Orbitofrontal and Left and Right Precuneus.

We have noticed that the obtained accuracy results were again very variable and sensitive to the pure test set, which is, basically, due to the dimensionality problem. There are, by far, more features than examples and the training procedure seems to create a model that does not generalize well. It is the same problem that arised in Section 6.1 but more accentuated, as we have added more features and the $\frac{F}{N}$ relation have increased even more.

Thus, the interpretation of this results is very difficult. Although it suggests that no benefit is obtained by adding more information from different image modalities to the data set, more studies with larger databases are required to draw robust conclusions.

Cuadro 6.4: Classification results using the Lausanne Database when adding T_1-lobe. The results are presented as *Accuracy pure (Accuracy Cross-Validation) [Condifence interval: lower bound- upper bound]*. GM Region Volumes is refered as Vol, Lobe Volumes as Lobe, the Asymmetry coefficients as ASY, the brain region T_1-quantitative values as $_1$ and the lobe T_1-quantitative values as T_1-L.

Input		**Sensitivity pure (CV)**	**Specificity pure(CV)**	**Accuracy pure [CI] (CV)**
T_1	-	50.0% (61.9%)	62.5% (57.1%)	56.3%[33.2- 76.9] (59.5%)
T_1 - L	-	62.5% (61.9%)	50.0% (57.1%)	56.3%[33.2- 76.9] (59.5%)
VOL+T_1 - L	ICV	50% (71.4%)	75.0% (76.2%)	62.5% [38.6-81.5] (73.8%)
	No ICV	50.0% (61.9%)	62.5% (85.7%)	56.3% [33.2-76.9] (73.8%)
Lobe Volumes+T_1 - L	ICV	25.0% (61.9%)	37.5% (71.4%)	31.3% [14.2-55.6] (66.7%)
Asymmetry+T_1 - L	-	37.5% (52.4%)	62.5% (57.1%)	50.0% [28.0-72.0] (54.8%)
WM+T_1 - L	ICV	62.5% (71.4%)	62.5% (81%)	62.5% [38.6-81.5] (76.2%)
	No ICV	75.0% (57.1%)	50.0% (61.9%)	62.5% [38.6-81.5] (59.5%)
Vol+Lobe+T_1 - L	ICV	50.0% (66.7%)	62.5% (76.2%)	56.3% [33.2-76.9] (71.4%)
	No ICV	50.0% (66.7%)	62.5% (81%)	56.3% [33.2-76.9] (73.8%)
WM+Lobe+T_1 - L	ICV	75.0% (66.7%)	75.0% (71.4%)	**75.0%** [50.5-89.8] (69.0%)
	No ICV	62.5% (61.9%)	62.5% (57.1%)	62.5% [38.6-81.5] (59.5%)
Vol+ASY+T_1 - L	ICV	50.0% (61.9%)	75.0% (61.9%)	62.5% [38.6-81.5] (61.9%)
	No ICV	50.0% (61.9%)	62.5% (85.7%)	56.3% [33.2-76.9] (73.8%)
Vol+WM+T_1 - L	ICV	75.0% (57.1%)	75% (71.4%)	**75.0%** [50.5-89.8] (64.3%)
	No ICV	50% (61.9%)	75.0% (76.2%)	62.5% [38.6-81.5] (69.0%)
WM+ASY+T_1 - L	ICV	50.0% (61.9%)	50.0% (76.2%)	50.0% [28.0-72.0] (69.0%)
	No ICV	75.0% (76.2%)	50.0% (57.1%)	62.5% [38.6-81.5] (66.7%)
Vol+WM+ASY+T_1 - L	ICV	75.0% (61.9%)	75.0% (76.2%)	**75.0%** [50.5-89.8] (69.0%)
	No ICV	62.5% (61.9%)	75.0% (76.2%)	68.8% [44.4-85.8] (69.0%)
WM+ASY+Lobe+T_1 - L	ICV	50.0% (66.7%)	62.5% (71.4%)	56.3% [33.2-76.9] (69.0%)
	No ICV	75.0% (57.1%)	62.5% (57.1%)	68.8% [44.4-85.8] (57.1%)
Vol+ASY+Lobe+T_1 - L	ICV	50.0% (61.9%)	62.5% (57.1%)	56.3% [33.2-76.9] (59.5%)
	No ICV	62.5% (61.9%)	62.5% (66.7%)	62.5% [38.6-81.5] (64.3%)

Cuadro 6.5: Classification results using the Lausanne Database when adding brain region T_1 values. The results are presented as *Accuracy pure (Accuracy Cross-Validation) [Condifence interval: lower bound- upper bound]*. GM Region Volumes is refered as Vol, Lobe Volumes as Lobe, the Asymmetry coefficients as ASY and the brain region T_1-quantitative values as T_1.

Input		Sensitivity pure (CV)	Specificity pure(CV)	Accuracy pure [CI] (CV)
VOL+T_1	ICV	75.0% (61.9%)	62.5% (66.7%)	68.8% [44.4-85.8] (64.3%)
	No ICV	37.5% (57.1%)	62.5% (71.4%)	50.0% [28.0-72.0] (64.3%)
Lobe Volumes+T_1	ICV	50.0% (57.1%)	50.0% (66.7%)	50.0% [28.0-72.0] (61.9%)
Asymmetry+T_1	-	62.5% (66.7%)	75.0% (57.1%)	68.8% [44.4-85.8] (61.9%)
WM+T_1	ICV	62.5% (71.4%)	62.5% (66.7%)	62.5% [38.6-81.5] (69.0%)
	No ICV	37.5% (66.7%)	75.0% (71.4%)	56.3% [33.2-76.9] (69.0%)
Vol+Lobe+T_1	ICV	62.5% (66.7%)	62.5% (61.9%)	62.5% [38.6-81.5] (64.3%)
	No ICV	62.5% (52.4%)	75.0% (66.7%)	68.8% [44.4-85.8] (59.5%)
WM+Lobe+T_1	ICV	50.0% (66.7%)	50.0% (66.7%)	50.0% [28.0-72.0] (66.7%)
	No ICV	62.5% (57.1%)	50.0% (76.2%)	56.3% [33.2-76.9] (66.7%)
Vol+ASY+T_1	ICV	50.0% (52.4%)	75.0% (66.7%)	62.5% [38.6-81.5] (59.5%)
	No ICV	75.0% (52.4%)	62.5% (71.4%)	68.8% [44.4-85.8] (61.9%)
Vol+WM+T_1	ICV	62.5% (52.4%)	62.5% (76.2%)	62.5% [38.6-81.5] (64.3%)
	No ICV	50.0% (57.1%)	62.5% (76.2%)	56.3% [33.2-76.9] (66.7%)
WM+ASY+T_1	ICV	75.0% (66.7%)	62.5% (76.2%)	**68.8%** [44.4-85.8] (71.4%)
	No ICV	62.5% (66.7%)	75.0% (76.2%)	68.8% [44.4-85.8] (71.4%)
Vol+WM+ASY+T_1	ICV	50.0% (71.4%)	62.5% (61.9%)	56.3% [33.2-76.9] (66.7%)
	No ICV	62.5% (57.1%)	75.0% (76.2%)	68.8% [44.4-85.8] (66.7%)
WM+ASY+Lobe+T_1	ICV	62.5% (71.4%)	62.5% (71.4%)	62.5% [38.6-81.5] (71.4%)
	No ICV	62.5% (66.7%)	75.0% (76.2%)	68.8% [44.4-85.8] (71.4%)
Vol+ASY+Lobe+T_1	ICV	50.0% (52.4%)	87.5% (81%)	68.8% [44.4-85.8] (66.7%)
	No ICV	62.5% (52.4%)	62.5% (61.9%)	62.5% [38.6-81.5] (57.1%)

Cuadro 6.6: Classification results using the Lausanne Database when adding all T_1 data. The results are presented as *Accuracy pure (Accuracy Cross-Validation) [Condifence interval: lower bound- upper bound]*. GM Region Volumes is refered as Vol, Lobe Volumes as Lobe, the Asymmetry coefficients as ASY, the brain region T_1-quantitative values as T_1 and the lobe T_1-quantitative values as T_1-L.

Input		Sensitivity pure (CV)	Specificity pure(CV)	Accuracy pure [CI] (CV)
VOL+T_1 - all	ICV	50.0 % (61.9 %)	62.5 % (61.9 %)	56.3 % [33.2-76.9] (61.9 %)
	No ICV	37.5 % (52.4 %)	62.5 % (66.7 %)	50.0 % [28.0-72.0] (59.5 %)
Lobe Volumes+T_1 - all	ICV	50.0 % (52.4 %)	62.5 % (57.1 %)	56.3 % [33.2-76.9] (54.8 %)
Asymmetry+T_1 - all	-	75.0 % (61.9 %)	50.0 % (52.4 %)	62.5 % [38.6-81.5] (57.1 %)
WM+T_1 - all	ICV	75.0 % (66.7 %)	37.5 % (71.4 %)	56.3 % [33.2-76.9] (69.0 %)
	No ICV	62.5 % (61.9 %)	37.5 % (66.7 %)	50.0 % [28.0-72.0] (64.3 %)
Vol+Lobe+T_1 - all	ICV	37.5 % (57.1 %)	62.5 % (57.1 %)	50.0 % [28.0-72.0] (57.1 %)
	No ICV	37.5 % (52.4 %)	75.0 % (61.9 %)	56.3 % [33.2-76.9] (57.1 %)
WM+Lobe+T_1 - all	ICV	75.0 % (66.7 %)	37.5 % (66.7 %)	56.3 % [33.2-76.9] (66.7 %)
	No ICV	37.5 % (61.9 %)	50.0 % (66.7 %)	43.8 % [23.1-66.8] (64.3 %)
Vol+ASY+T_1 - all	ICV	62.5 % (57.1 %)	75.0 % (57.1 %)	68.8 % [44.4-85.8] (57.1 %)
	No ICV	37.5 % (57.1 %)	62.5 % (61.9 %)	50.0 % [28.0-72.0] (59.5 %)
Vol+WM+T_1 - all	ICV	50.0 % (52.4 %)	50.0 % (76.2 %)	50.0 % [28.0-72.0] (64.3 %)
	No ICV	37.5 % (57.1 %)	75.0 % (76.2 %)	56.3 % [33.2-76.9] (66.7 %)
WM+ASY+T_1 - all	ICV	62.5 % (66.7 %)	62.5 % (71.4 %)	62.5 % [38.6-81.5] (69.0 %)
	No ICV	75.0 % (76.2 %)	62.5 % (61.9 %)	68.8 % [44.4-85.8] (69.0 %)
Vol+WM+ASY+T_1 - all	ICV	62.5 % (61.9 %)	75.0 % (76.2 %)	68.8 % [44.4-85.8] (69.0 %)
	No ICV	62.5 % (66.7 %)	62.5 % (71.4 %)	62.5 % [38.6-81.5] (69.0 %)
WM+ASY+Lobe+T_1 - all	ICV	75.0 % (66.7 %)	62.5 % (66.7 %)	68.8 % [44.4-85.8] (66.7 %)
	No ICV	75.0 % (71.4 %)	62.5 % (66.7 %)	68.8 % [44.4-85.8] (69.0 %)
Vol+ASY+Lobe+T_1 - all	ICV	50.0 % (47.6 %)	75.0 % (61.9 %)	62.5 % [38.6-81.5] (54.8 %)
	No ICV	50.0 % (57.1 %)	75.0 % (61.9 %)	62.5 % [38.6-81.5] (59.5 %)
All	ICV	62.5 % (61.9 %)	75.0 % (71.4 %)	68.8 % [44.4-85.8] (66.7 %)
	No ICV	50.0 % (61.9 %)	87.5 % (71.4 %)	68.8 % [44.4-85.8] (66.7 %)

6.4. Stacking

Finally, we tried to create a new different classifier by stacking single classifiers (see Figure 6.3). The basic idea is that, instead of feeding the classifier with all the features, M classifiers are created, each of them fed with a feature type. Thus, a classification process is performed separately using only GM region volumes, Lobe Volumes, WM region volumes, Asymmetry Coefficients and T_1-quantitative measurements, generating M predicted labels $L_m = \{-1, 1\}$ per subject, one for each classification process. The final decision label will be the sum of the L_m labels. If the sum is equal or lower than zero, the subject is labeled as MCI. Otherwise the subject is labeled as CTL. This kind of approach is suitable for the type of scenario where different kind of features are available and it is likely that combining classifiers the accuracy raises. This process have been performed for the Lausanne and for the Expanded Database and is presented in the subsections that follow.

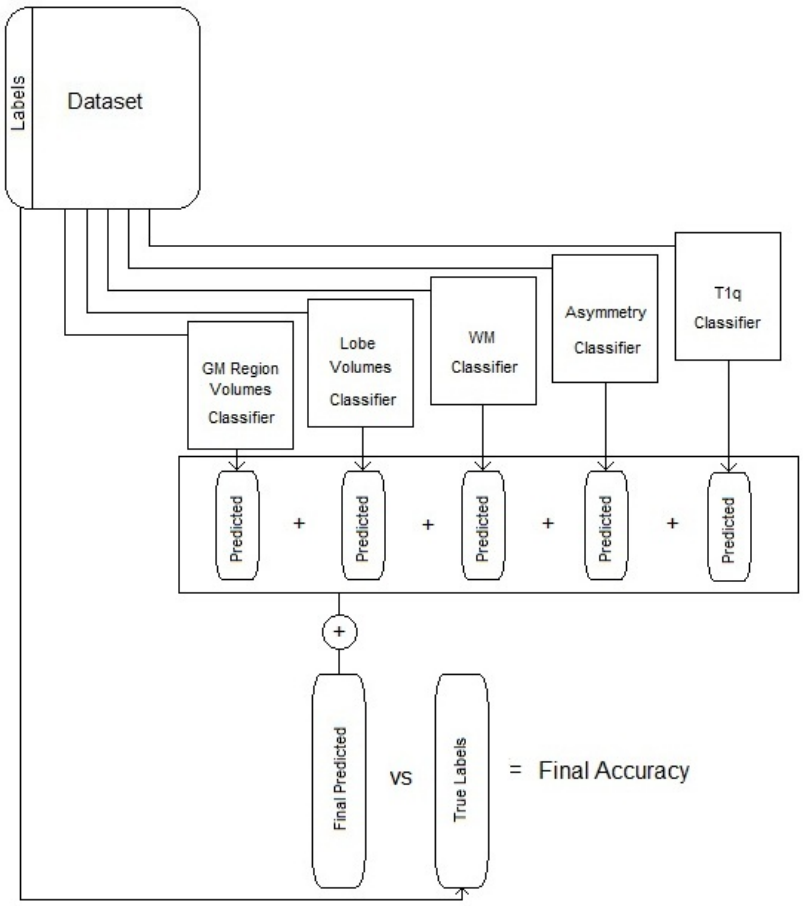
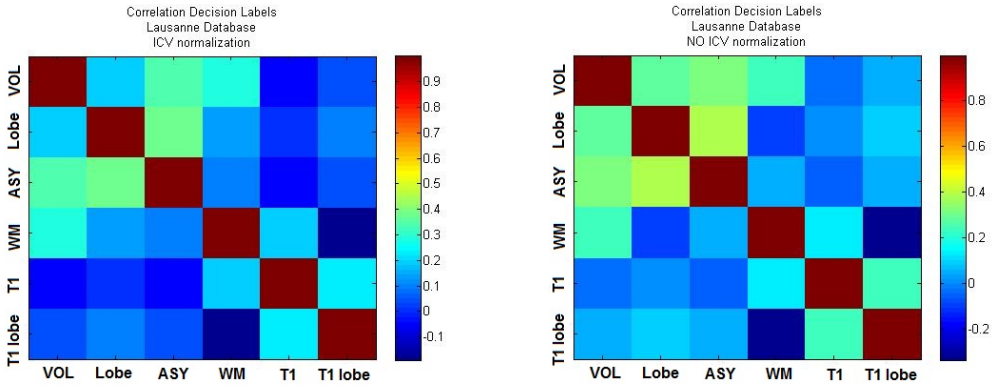


Figura 6.3: Classifier stacking flow diagram

6.4.1. Lausanne Database

A pre-processing step is required before stacking the classifiers, which is checking the correlation of the decision labels from each single classifier in order to know which combination is likely to work better. If the label correlation of two different classifiers is high, it means that there is not a lot of extra information added when taking both insted of taking them alone. But if the label correlation is very low, it means that they are strongly independent and combining them will produce a noisy output. The desired correlation values will be in the [0.2-0.5] range. The correlation of the Lausanne Database CV decision labels is showed in the Figure 6.4. Appreciate that there is 0.3-0.4 of correlation coefficient in Vol- ASY- WM, ASY- Vol- Lobe, WM- Vol- T_1 and T_1- WM- T_1-lobe. So this four combinations have been tested, as well as combining all of them. The results are presented in the Table 6.7 and explained below. No combination increased the pure accuracy 75.0 % obtained in Section 6.1. Even so, the CV results show a tendancy to increase respect the single analysis, up to 83.4 % when combining all the classifiers. However, the fact that the database is so small do not let extract conclusions and further studies with larger databases are required to determinate if the stacking strategy could add robustness or not.



(a) Correlation Decision Labels Lausanne Database with ICV.

(b) Correlation Decision Labels Lausanne Database without ICV.

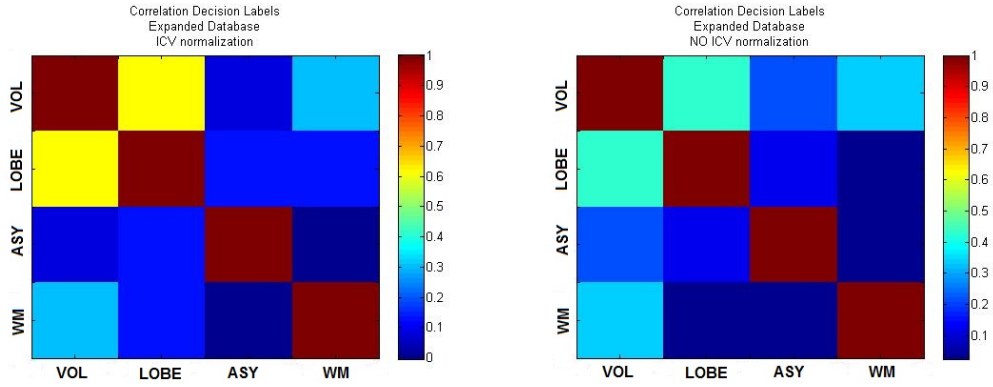
Figura 6.4: Correlation Decision Labels Lausanne Database with (a) and without (b) ICV normalization.

Cuadro 6.7: Lausanne Database multivariate classification results when stacking classifiers. The results are presented as *Accuracy pure (Accuracy Cross-Validation) [Condifence interval: lower bound- upper bound]*. GM Region Volumes is refered as Vol, Lobe Volumes as Lobe, the Asymmetry coefficients as ASY, the brain region T_1-quantitative values as T_1 and the lobe T_1-quantitative values as T_1-L.

Combination		**Sensitivity pure (CV)**	**Specificity pure(CV)**	**Accuracy pure [CI] (CV)**
Vol- ASY- WM	ICV	50.0% (71.4%)	62.5% (81%)	56.3% [33.2- 76.9] (76.2%)
	No ICV	62.5% (66.7%)	62.5% (85.7%)	62.5% [38.6- 81.5] (76.2%)
ASY- Vol- Lobe	ICV	75.0% (71.4%)	62.5% (71.4%)	68.8% [44.4- 85.8] (71.4%)
	No ICV	75.0% (71.4%)	62.5% (76.2%)	68.8% [44.4- 85.8] (73.8%)
WM- Vol- T_1	ICV	62.5% (76.2%)	75.0% (71.4%)	68.8% [44.4- 85.8] (73.8%)
	No ICV	62.5% (66.6%)	75.0% (71.4%)	68.8% [44.4- 85.8] (69.0%)
T_1- WM- T_1-lobe	ICV	75.0% (66.6%)	50.0% (61.9%)	62.5% [38.6- 81.5] (64.3%)
	No ICV	75.0% (66.6%)	50.0% (66.6%)	62.5% [38.6- 81.5] (66.6%)
All	ICV	75.0% (85.7%)	37.5% (78.6%)	56.3% [33.2- 76.9] (**82.2%**)
	No ICV	75.0% (90.5%)	37.5% (76.2%)	56.3% [33.2- 76.9] (**83.4%**)

6.4.2. Expanded Database

Exactly the same procedure have been done for the Expanded Database. First of all, the decision label correlation have been computed and is shown in the Figure 6.5. In this case the combinations Vol+Lobe+WM and Vol+Lobe+ASY+WM have been tried. The results are presented in the Table 6.8 and explained below. Again, any combination achieved better pure nor CV accuracy than in the previous sections. This database is twice as large as the Lausanne database, so the results are less variable, what means that probably, the stacking strategy would not be a suitable approach for this problem. Nevertheless, more experiments are required with larger databases to validate whether this concept can be useful or not.



(a) Correlation Decision Labels Expanded Database with ICV.

(b) Correlation Decision Labels Expanded Database without ICV.

Figura 6.5: Correlation Decision Labels Expanded Database with (a) and without (b) ICV normalization.

Cuadro 6.8: Expanded Database classification results when stacking classifiers. The results are presented as *Accuracy pure (Accuracy Cross-Validation) [Condifence interval: lower bound-upper bound]*. GM Region Volumes is refered as Vol, Lobe Volumes as Lobe, the Asymmetry coefficients as ASY, the brain region T_1-quantitative values as T_1 and the lobe T_1-quantitative values as T_1-L.

Combination		Sensitivity pure (CV)	Specificity pure(CV)	Accuracy pure [CI] (CV)
Vol- Lobe- WM	ICV	66.7% (63.0%)	75.0% (80.4%)	70.9% [50.8- 85.1] (71.8%)
	No ICV	50.0% (54.3%)	75.0% (65.2%)	62.5% [42.7- 78.8] (59.8%)
All	ICV	83.3% (52.2%)	58.3% (69.6%)	70.9% [50.8- 85.1] (60.9%)
	No ICV	66.7% (47.8%)	58.3% (50.0%)	62.5% [42.7- 78.8] (48.9%)

7 Conclusion

7.1. Discussion

This study was designed to face the problem of the early diagnosis of the AD and to find out which regions are affected in MCI. Our hypothesis was that adding more information to the classifiers than only the widely used GM brain region volumes could improve classification accuracies. Thus, apart from T1-weighted MRI WM and GM brain region data, we have combined multiscale data (lobe volumes) and information from other MR image modalities (T_1-quantitative MRI), as well as Asymmetry coefficients. The set of participants used in this work was from the ADNI 1 and the Lausanne Database. To carry out this work we have used SVM classifier, which has been trained with a part of the whole database and tested in the other. All the training steps have been performed only in the training set, in order to get a pure accuracy that is directly what is expected be obtained in a real clinical environment. After performing all the analysis presented in the previous sections, our conclusions are summarized below.

Age correction

Although many studies in the literature have reported sistematically youngest AD and oldest CTL misclassifications, any of them attempted to correct this undesirable effect. Inspired by a recent study presented by *Dukart et al.,[17]*, we have applied a region-based age-correction pre-processing step before feeding the classifier, which, as far as we know, is a completely new concept. Similarly to the results reported by [17], the classification accuracy improved a 2 % in the CTL versus AD classification. Also the results were better when classifying CTL versus MCI, 1.5 %, 3.8 % and 4.5 % of accuracy improvement with the ADNI 1, Lausanne and Expanded Databases respectively. There is even more beyond this accuracy improvements: the mean age differences between groups of misclassified subjects have become statistically not significant for CTL versus AD and have been reduced for CTL versus MCI classifications. These results demonstrate that the youngest AD and oldest CTL are no longer sistematically misclassified. As regards to MCI, although there were still differences in mean age among misclassified

groups, the gap in mean age among groups have also been reduced, what means that have a positive effect for the classification. Despite these slightly lower improvements for MCI, which were expected as it is a more heterogenious group, we strongly believe that this pre-processing step is fully advantageous and totally required for AD, MCI and CTL classifications problems.

ICV normalization

In this study we have performed all the analysis twice: once normalizing by ICV and once not, and the results were very similar in both cases. Moreover, the fact that *usual suspect* structures like the Left Hippocampus or the Left Temporal Pole presented more differences in mean when not normalized by ICV suggests that some useful information would be smoothed when applying this pre-processing step. We believe that global structures, as brain lobes, should be normalized in order not to classify by head size. Nevertheless, more studies are required to determinate whether it is useful to normalyze all the structures by ICV or, on the other side, only some or none small structures should be normalized.

Univariate Accuracy results

Our presented accuracy results are comparable to those obtained in the literature when using the same database: 88.3%, 81.7% and 60.0% for the CTL vs AD, CTL vs MCI and MCI vs AD classifications respectively. On the other hand, the results with the Lausanne and the Expanded database were lower, a 68.8% and a 70.9% when using only GM regional volumes. Our interpretation of this results is that the Lausanne MCI participants are probably less imparied and, therefore, more difficult to classify than the ADNI 1 ones, as the MMSE score of the first group is almost one point higher than the second. Moreover, the fact that the database is smaller makes harder the generalization task to the SVM, according to reported results for small databases.

One final consideration should be taken into account when comparing our results to those obtained in previous studies. Since the normal ageing effects are not treated in the literature, their results could be slightly biased and probably less robust than ours. Proof of this fact would be the systematic oldest CTL and younger AD subjects misclassification.

Is it worthy to add more variables to the analysis?

The best accuracy results with both the Lausanne and the Expanded Database have been gotten when combining different variables instead of when taking them alone. This tendency suggests that better overall performances can be achieved when taking into account multiple variables. Although the best results have been obtained when combining GM region volumes, GM lobe volumes and the asymmetry coefficients (multiscale and multivariate approach), it is still not clear which scheme would normally perform better. Thus, further studies are required with larger databases to confirm this tendancy and to find out the best variable combination.

Respect adding WM volumetric information, the results suggest that more discriminant power is achieved when using the GM region volumes.

In regard to the combination of different image modalities, the results were found to be very unstable, mainly due to the database size. Although no clear benefit was obtained adding T_1-quantitative data, studies with larger databases are required to determinate its real usefulness for the classification.

Most discriminative features

We have found that the most discriminative features for the CTL vs AD classification were the Left and Right Hippocampus, the Left and Right Amygdala, the Left Entorhinal, the Left Inferior Lateral Ventricle, the Right Middle Temporal, Right Caudal Anterior Cingulate, Left Isthmus Cingulate and Right Fusiform.

Refering to the MCI classification, the most discriminative GM region volumes in all databases were the Left and the Right Hippocampus, the Left and the Right Entorhinal, the Inferior Lateral Ventricles and the Left Caudal Middle Frontal. For the ADNI 1 database, we also found the Left Middle Temporal, the Left and Right Amygdala and the Right Accumbens Area. For the Lausanne Database, the Amygdala did not appear to be significative, what was a very unexpected finding, as almost all the literature AD and MCI classification studies reported changes in the Amygdala. As the Lausanne Database MCI individuals seemed to be less imparied that the ADNI 1 ones, more studies are required to determinate whether this fact is due to chance or otherwise this structure is not early affected by the disease. In the Lausanne Database we also found the Left Fusiform, the Left Parahippocampal and the WM Hypointensities in the top-ranked set of features. Finally, in the Expanded Database the Right Precuneus, Right Superior Parietal and Right Temporal pole have been found to be discriminative.

As regard to the GM lobe volumes, the most discriminative were the Right Limbic and Right Frontal for the Lausanne Database, while the Left Limbic and Left and Right Temporal for the Expanded Database. For the Asymmetry Coefficients, the most discriminative have been the Caudal Anterior Cingulate, the Fusiform and the Lateral Ventricle for the Lausanne Database, and the Precuneus and Amygdala for the Expanded Database. Refering to WM, the most discriminative regions were the Right entorhinal, Right Paracentral, Right Precentral, Left CaudalMiddle Frontal, Left Rostral Anterior Cingulate, Left Transverse Temporal, Left Rostral Anterior Cingulate and Left Middle Temporal for the Lausanne Database, while the Left Cortical WhiteMatter, Left and Right Entorhinal, Right Precuneus, Left and Right Superior Temporal, Right Middle Temporal, Right Superior Frontal and Right Insula for the Expanded Database. Finally, for the T_1- quantitative data, the most discriminative lobes were the Parietal, both for GM and WM, while the most discriminative regions were the Right Bankssts, Left Entorhinal, Left Frontal Pole, Left Parsopercularis, Left Medial Orbitofrontal and Left and Right Precuneus.

7.2. Future Work

In this Section the possible improvements and future research lines are summarized.

- **Classifier**. The LIBLINEAR Library provides many options to configure the classifier, for instance the type of solver. There exist many configurations, for example L2-regularized L2-loss support vector classification (dual)) or L1-regularized L2-loss support vector classification. In this work we have used the first one which is the default option as it yielded to better accuracies. The library authors also pointed that in most cases, L1 regularization does not give higher accuarcy. Nevertheless, more tests could be done to determine if it could be useful in any situation. For further information, the reader can refer to [39].

- **Feature selection**. The RFE method is simple and easy to implement a feature selection. However, the feature selection step is, in our opinion, a very critical issue as normally there are no large database available to perform the analysis. There exist a lot of feature selection techniques, as F- score or methods using the gradient of the weight vector. Thus, other more sophisticated techniques could be tested. For more information, the reader can refer to [30, 31, 32, 33].

- **Pre-processing**. In this work it has been evidenced that our age-effect correction pre-processing step provides many advantages when used prior to the classification, while being extremely simple. We believe that this findings merits further resarch. On the one hand, we think that individual region studies should be done to precisely characterize the evolution of every single brain area. On the other hand, to be applicable in a real clinical enviornment, the regression coefficients should be universal and well- standarized so, more studies with multiple databases are required.

- **Input features**: The presented results show a tendancy to increase when combining types of feataures. The fact that the best accuracy has been got when combining GM region volumes, GM lobe volumes (multiscale) and asymmetry coefficients (multivariate) is very exciting and reaffirms our previous hypothesis. However, this fact should be checked with larger databases to whether the same improvement is obtained for larger datasets. Moreover, other features can be used for future analysis, as the WM-Asymmetry or the Lobe Asymmetry.

A Complete list of regions

In this Appendix, the complete list of regions provided by different types of FreeSurfer files are shown.

A.1. FreeSurfer *aseg* file

- Left Lateral
- Ventricle
- Left Inf Lat Vent
- Left Cerebellum White Matter
- Left Cerebellum Cortex
- Left Thalamus Proper
- Left Caudate
- Left Putamen
- Left Pallidum
- 3rdVentricle
- 4th Ventricle
- Brain Stem
- Left Hippocampus
- Left Amygdala
- CSF
- Left Accumbens area
- Left VentralDC
- Left vessel
- Left choroid plexus
- Right Lateral Ventricle
- Right Inf Lat Vent
- Right Cerebellum White Matter
- Right Cerebellum Cortex
- Right Thalamus Proper
- Right Caudate
- Right Putamen
- Right Pallidum
- Right Hippocampus
- Right Amygdala
- Right Accumbens area
- Right VentralDC

- Right vessel
- Right choroid plexus
- 5th Ventricle
- WM hypointensities
- Left WM hypointensities
- Right WM hypointensities
- non-WM hypointensities
- Left non-WM hypointensities
- Right non-WM hypointensities
- Optic Chiasm
- CC-Posterior
- CC-Mid-Posterior
- CC-Central
- CC-Mid-Anterior
- CC-Anterior
- lhCortexVol
- rhCortexVol
- CortexVol
- lhCorticalWhiteMatterVol
- rhCorticalWhiteMatterVol
- CorticalWhiteMatterVol
- SubCortGrayVol
- TotalGrayVol
- SupraTentorialVol
- IntraCranialVol

A.2. FreeSurfer *aparc* file

- lh bankssts volume
- lh caudalanteriorcingulate volume
- lh caudalmiddlefrontal volume
- lh cuneus volume
- lh entorhinal volume
- lh fusiform volume
- lh inferiorparietal volume
- lh inferiortemporal volume
- lh isthmuscingulate volume
- lh lateraloccipital volume
- lh lateralorbitofrontal volume
- lh lingual volume
- lh medialorbitofrontal volume
- lh middletemporal volume
- lh parahippocampal volume
- lh paracentral volume
- lh parsopercularis volume
- lh parsorbitalis volume
- lh parstriangularis volume
- lh pericalcarine volume
- lh postcentral volume
- lh posteriorcingulate volume
- lh precentral volume
- lh precuneus volume
- lh rostralanteriorcingulate volume

- lh rostralmiddlefrontal volume
- lh superiorfrontal volume
- lh superiorparietal volume
- lh superiortemporal volume
- lh supramarginal volume
- lh frontalpole volume
- lh temporalpole volume
- lh transversetemporal volume
- lh insula volume
- rh bankssts volume
- rh caudalanteriorcingulate volume
- rh caudalmiddlefrontal volume
- rh cuneus volume
- rh entorhinal volume
- rh fusiform volume
- rh inferiorparietal volume
- rh inferiortemporal volume
- rh isthmuscingulate volume
- rh lateraloccipital volume
- rh lateralorbitofrontal volume
- rh lingual volume
- rh medialorbitofrontal volume
- rh middletemporal volume
- rh parahippocampal volume
- rh paracentral volume
- rh parsopercularis volume
- rh parsorbitalis volume
- rh parstriangularis volume
- rh pericalcarine volume
- rh postcentral volume
- rh posteriorcingulate volume
- rh precentral volume
- rh precuneus volume
- rh rostralanteriorcingulate volume
- rh rostralmiddlefrontal volume
- rh superiorfrontal volume
- rh superiorparietal volume
- rh superiortemporal volume
- rh supramarginal volume
- rh frontalpole volume
- rh temporalpole volume
- rh transversetemporal volume
- rh insula volume

A.3. FreeSurfer *wmparc* file

- wm-lh-bankssts
- wm-lh-caudalanteriorcingulate
- wm-lh-caudalmiddlefrontal
- wm-lh-cuneus
- wm-lh-entorhinal
- wm-lh-fusiform

- wm-lh-inferiorparietal
- wm-lh-inferiortemporal
- wm-lh-isthmuscingulate
- wm-lh-lateraloccipital
- wm-lh-lateralorbitofrontal
- wm-lh-lingual
- wm-lh-medialorbitofrontal
- wm-lh-middletemporal
- wm-lh-parahippocampal
- wm-lh-paracentral
- wm-lh-parsopercularis
- wm-lh-parsorbitalis
- wm-lh-parstriangularis
- wm-lh-pericalcarine
- wm-lh-postcentral
- wm-lh-posteriorcingulate
- wm-lh-precentral
- wm-lh-precuneus
- wm-lh-rostralanteriorcingulate
- wm-lh-rostralmiddlefrontal
- wm-lh-superiorfrontal
- wm-lh-superiorparietal
- wm-lh-superiortemporal
- wm-lh-supramarginal
- wm-lh-frontalpole
- wm-lh-temporalpole
- wm-lh-transversetemporal
- wm-lh-insula
- wm-rh-bankssts
- wm-rh-caudalanteriorcingulate
- wm-rh-caudalmiddlefrontal
- wm-rh-cuneus
- wm-rh-entorhinal
- wm-rh-fusiform
- wm-rh-inferiorparietal
- wm-rh-inferiortemporal
- wm-rh-isthmuscingulate
- wm-rh-lateraloccipital
- wm-rh-lateralorbitofrontal
- wm-rh-lingual
- wm-rh-medialorbitofrontal
- wm-rh-middletemporal
- wm-rh-parahippocampal
- wm-rh-paracentral
- wm-rh-parsopercularis
- wm-rh-parsorbitalis
- wm-rh-parstriangularis
- wm-rh-pericalcarine
- wm-rh-postcentral
- wm-rh-posteriorcingulate
- wm-rh-precentral
- wm-rh-precuneus
- wm-rh-rostralanteriorcingulate
- wm-rh-rostralmiddlefrontal
- wm-rh-superiorfrontal

- wm-rh-superiorparietal
- wm-rh-superiortemporal
- wm-rh-supramarginal
- wm-rh-frontalpole
- wm-rh-temporalpole
- wm-rh-transversetemporal
- wm-rh-insula
- Left-UnsegmentedWhiteMatter
- Right-UnsegmentedWhiteMatter
- BrainSegVol
- lhCorticalWhiteMatterVol
- rhCorticalWhiteMatterVol
- CorticalWhiteMatterVol

A.4. T1-quantitative

- T1-1000 ctx-lh-unknown
- T1-1001 ctx-lh-bankssts
- T1-1002 ctx-lh-caudalanteriorcingulate
- T1-1003 ctx-lh-caudalmiddlefrontal
- T1-1005 ctx-lh-cuneus
- T1-1006 ctx-lh-entorhinal
- T1-1007 ctx-lh-fusiform
- T1-1008 ctx-lh-inferiorparietal
- T1-1009 ctx-lh-inferiortemporal
- T1-1010 ctx-lh-isthmuscingulate
- T1-1011 ctx-lh-lateraloccipita
- T1-1012 ctx-lh-lateralorbitofrontal
- T1-1013 ctx-lh-lingual
- T1-1014 ctx-lh-medialorbitofrontal
- T1-1015 ctx-lh-middletemporal
- T1-1016 ctx-lh-parahippocampal
- T1-1017 ctx-lh-paracentral
- T1-1018 ctx-lh-parsopercularis
- T1-1019 ctx-lh-parsorbitalis
- T1-1020 ctx-lh-parstriangularis
- T1-1021 ctx-lh-pericalcarine
- T1-1022 ctx-lh-postcentral
- T1-1023 ctx-lh-posteriorcingulate
- T1-1024 ctx-lh-precentral
- T1-1025 ctx-lh-precuneus
- T1-1026 ctx-lh-rostralanteriorcingulate
- T1-1027 ctx-lh-rostralmiddlefrontal
- T1-1028 ctx-lh-superiorfrontal
- T1-1029 ctx-lh-superiorparietal
- T1-1030 ctx-lh-superiortemporal
- T1-1031 ctx-lh-supramarginal
- T1-1032 ctx-lh-frontalpole
- T1-1033 ctx-lh-temporalpole
- T1-1034 ctx-lh-transversetemporal
- T1-2000 ctx-rh-unknown
- T1-2001 ctx-rh-bankssts

- T1-2002 ctx-rh-caudalanteriorcingulate
- T1-2003 ctx-rh-caudalmiddlefrontal
- T1-2005 ctx-rh-cuneus
- T1-2006 ctx-rh-entorhinal
- T1-2007 ctx-rh-fusiform
- T1-2008 ctx-rh-inferiorparietal
- T1-2009 ctx-rh-inferiortemporal
- T1-2010 ctx-rh-isthmuscingulate
- T1-2011 ctx-rh-lateraloccipital
- T1-2012 ctx-rh-lateralorbitofrontal
- T1-2013 ctx-rh-lingual
- T1-2014 ctx-rh-medialorbitofrontal
- T1-2015 ctx-rh-middletemporal
- T1-2016 ctx-rh-parahippocampal
- T1-2017 ctx-rh-paracentral
- T1-2018 ctx-rh-parsopercularis
- T1-2019 ctx-rh-parsorbitalis
- T1-2020 ctx-rh-parstriangularis
- T1-2021 ctx-rh-pericalcarine
- T1-2022 ctx-rh-postcentral
- T1-2023 ctx-rh-posteriorcingulate
- T1-2024 ctx-rh-precentral
- T1-2025 ctx-rh-precuneus
- T1-2026 ctx-rh-rostralanteriorcingulate
- T1-2027 ctx-rh-rostralmiddlefrontal
- T1-2028 ctx-rh-superiorfrontal
- T1-2029 ctx-rh-superiorparietal
- T1-2030 ctx-rh-superiortemporal
- T1-2031 ctx-rh-supramarginal
- T1-2032 ctx-rh-frontalpole
- T1-2033 ctx-rh-temporalpole
- T1-2034 ctx-rh-transversetemporal

A.5. ADNI 1 Database selected features

- Volume (Cortical Parcellation) of Right Parahippocampal
- Volume (Cortical Parcellation) of Right Precuneus
- Volume (WM Parcellation) of Third Ventricle
- Volume (Cortical Parcellation) of Right RostralAnteriorCingulate
- Volume (Cortical Parcellation) of Right SuperiorParietal
- Volume (Cortical Parcellation) of Right TemporalPole
- Volume (WM Parcellation) of Left AccumbensArea
- Volume (WM Parcellation) of Right Vessel
- Volume (WM Parcellation) of WMHypoIntensities
- Volume (Cortical Parcellation) of Left Insula

- Volume (WM Parcellation) of Left Amygdala
- Volume (Cortical Parcellation) of Right Insula
- Volume (Cortical Parcellation) of Left CaudalAnteriorCingulate
- Volume (Cortical Parcellation) of Left CaudalMiddleFrontal
- Volume (Cortical Parcellation) of Left Entorhinal
- Volume (Cortical Parcellation) of Left Fusiform
- Volume (WM Parcellation) of Left Hippocampus
- Volume (WM Parcellation) of Left InferiorLateralVentricle
- Volume (Cortical Parcellation) of Left InferiorParietal
- Volume (Cortical Parcellation) of Left IsthmusCingulate
- Volume (Cortical Parcellation) of Left LateralOccipital
- Volume (Cortical Parcellation) of Left LateralOrbitofrontal
- Volume (WM Parcellation) of Left LateralVentricle
- Volume (Cortical Parcellation) of Left MedialOrbitofrontal
- Volume (Cortical Parcellation) of Left MiddleTemporal
- Volume (Cortical Parcellation) of Left Parahippocampal
- Volume (Cortical Parcellation) of Left Precuneus
- Volume (Cortical Parcellation) of Left RostralAnteriorCingulate
- Volume (Cortical Parcellation) of Left SuperiorParietal
- Volume (Cortical Parcellation) of Left TemporalPole
- Volume (WM Parcellation) of Left Vessel
- Volume (WM Parcellation) of Non WMHypoIntensities
- Volume (WM Parcellation) of Right AccumbensArea
- Volume (WM Parcellation) of Right Amygdala
- Volume (Cortical Parcellation) of Right CaudalAnteriorCingulate
- Volume (Cortical Parcellation) of Right CaudalMiddleFrontal
- Volume (Cortical Parcellation) of Right Entorhinal
- Volume (Cortical Parcellation) of Right Fusiform
- Volume (WM Parcellation) of Right Hippocampus
- Volume (WM Parcellation) of Right InferiorLateralVentricle
- Volume (Cortical Parcellation) of RightInferiorParietal
- Volume (Cortical Parcellation) of Right IsthmusCingulate
- Volume (Cortical Parcellation) of Right LateralOccipital

- Volume (Cortical Parcellation) of Right LateralOrbitofrontal
- Volume (WM Parcellation) of Right LateralVentricle
- Volume (Cortical Parcellation) of Right MedialOrbitofrontal
- Volume (Cortical Parcellation) of Right MiddleTemporal
- Volume (WM Parcellation) of Fourth Ventricle

B Complete Results

In this Appendix, the complete list of ranked features for the univariate and the multiscale and multivariate analysis with best accuracy are presented. There are two figures for every classification procedure. The first one is a table where the most discriminative features of the internal cross-validation are showed, sorted by its discriminative power. Thus, the most discriminative feature is placed in the first position and so on. The second table shows the most discriminative features when the final classifier is retrained for the pure testing procedure.

B.1. Complete Results for the ADNI 1 Database

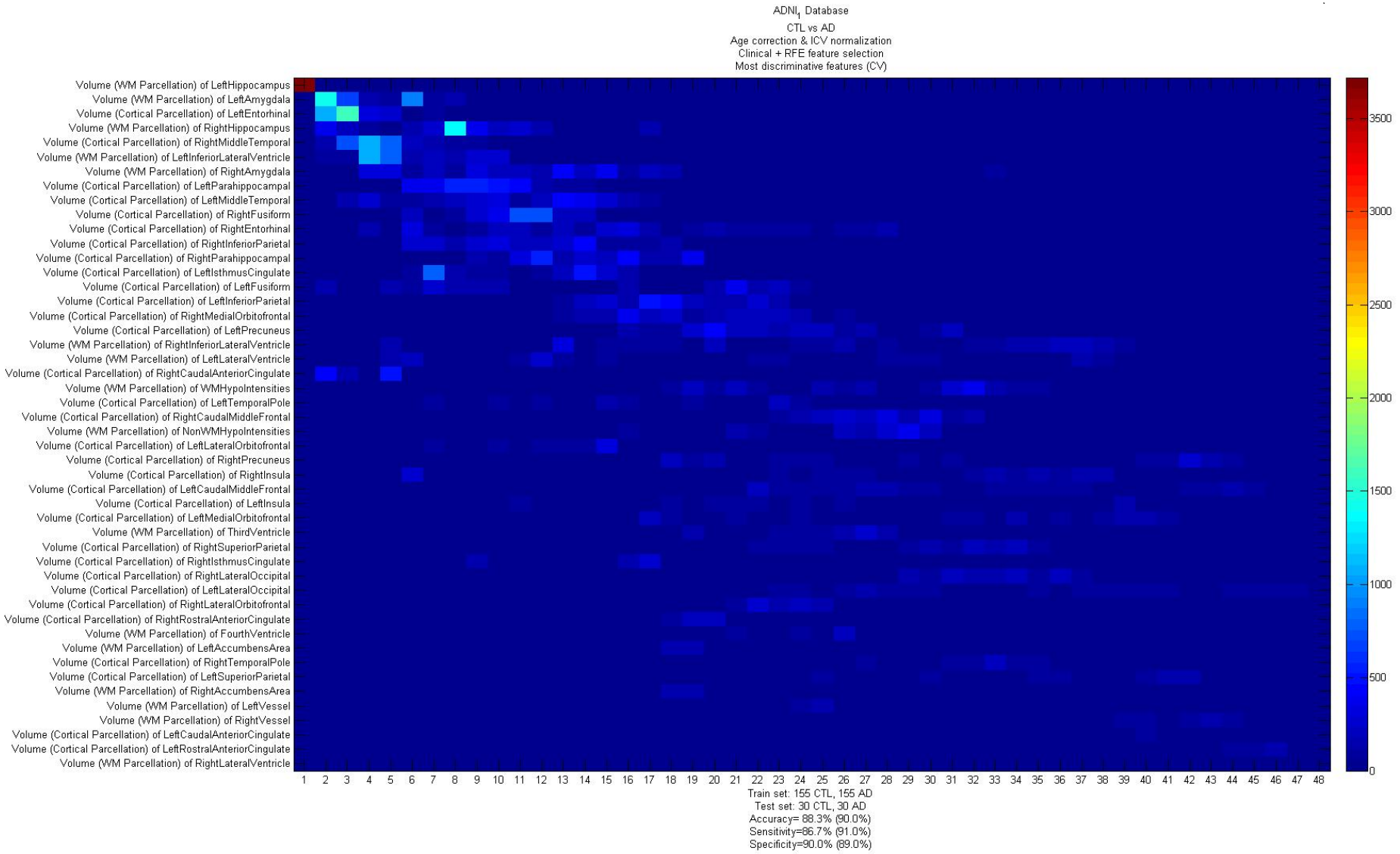


Figura B.1: ADNI1 CTL vs AD CV most discriminative Features

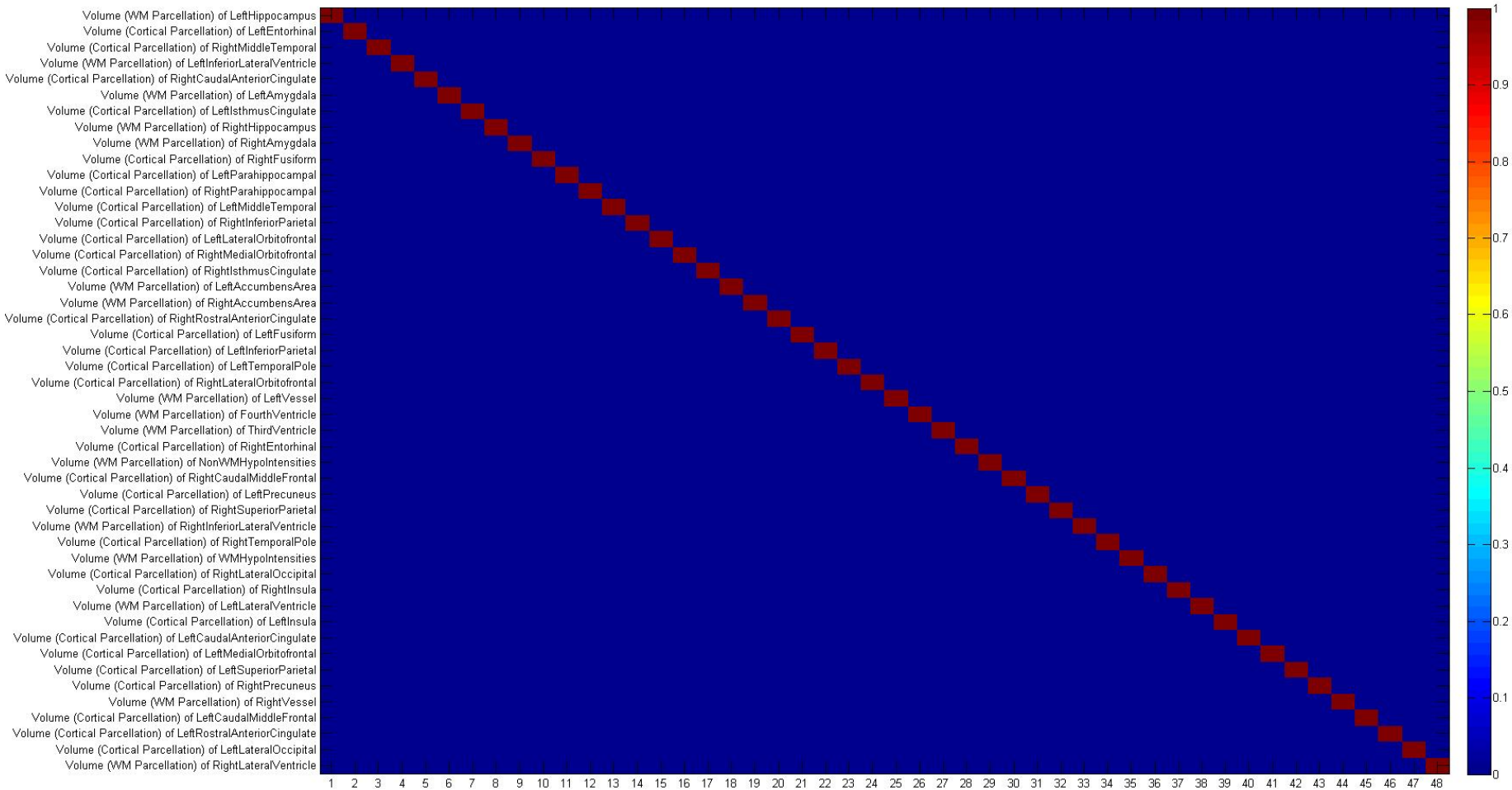


Figura B.2: ADNI1 CTL vs AD pure most discriminative Features

Figura B.3: ADNI1 CTL vs MCI CV most discriminative Features

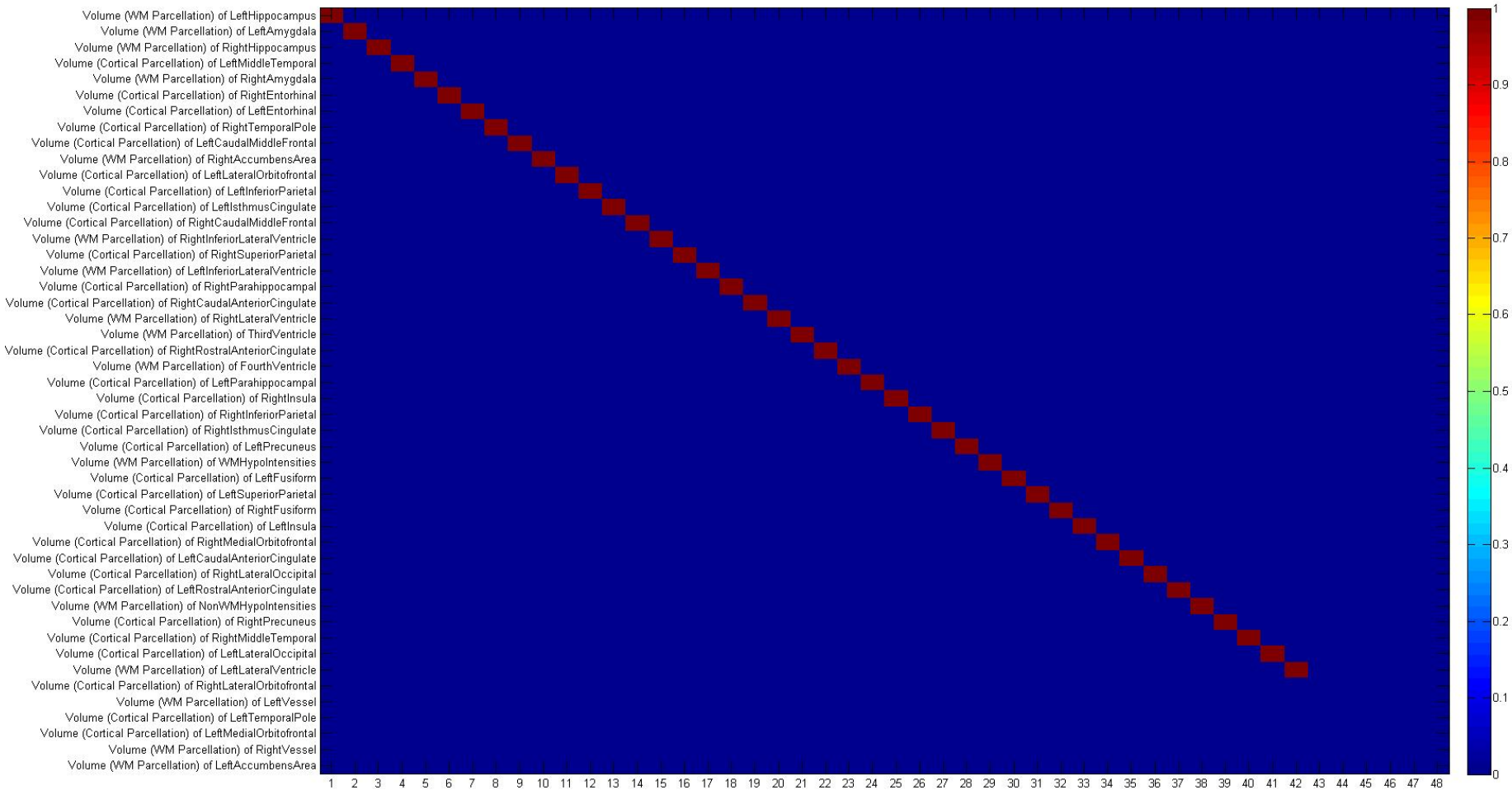


Figura B.4: ADNI1 CTL vs MCI pure most discriminative Features

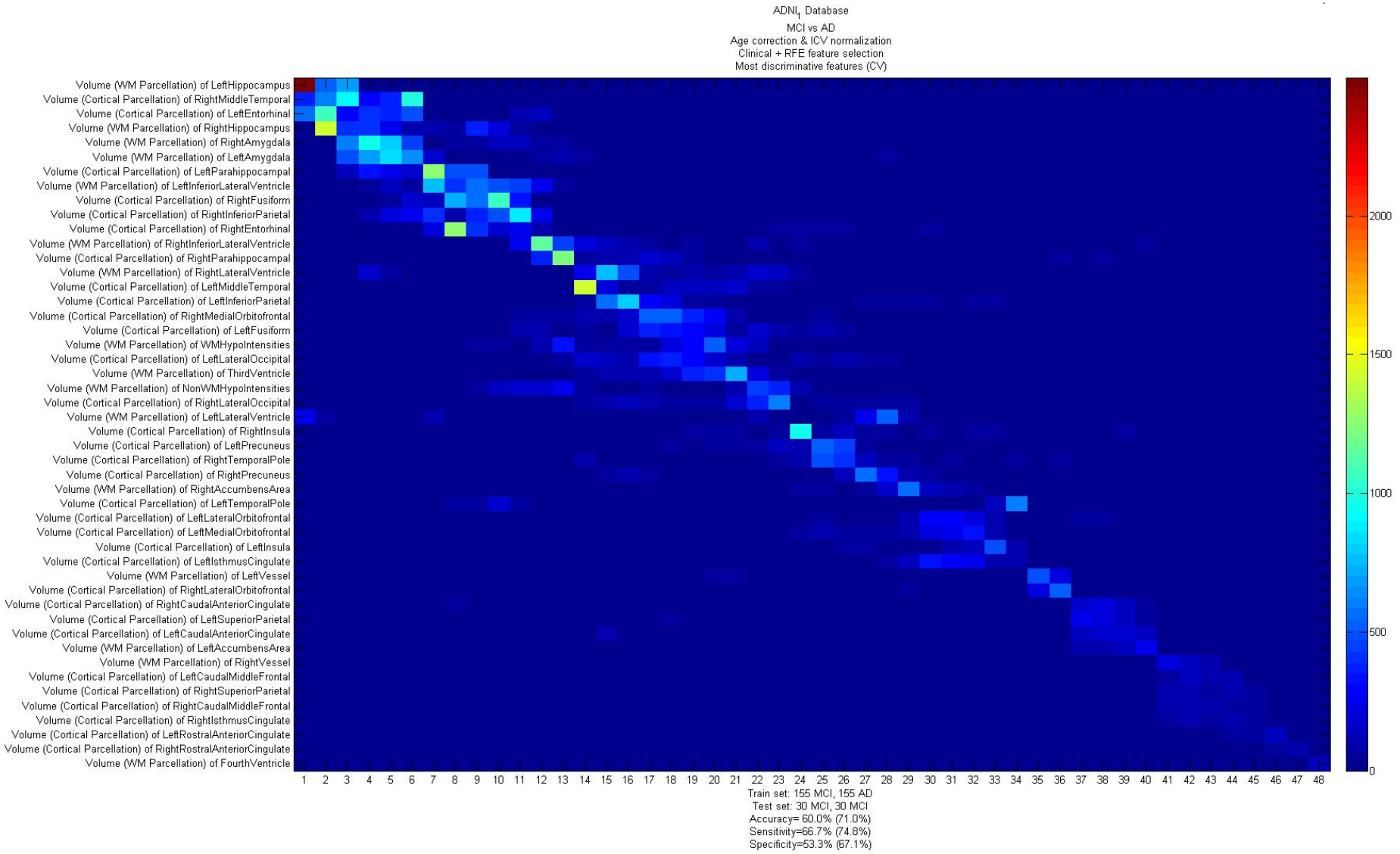


Figura B.5: ADNI1 MCI vs AD CV most discriminative Features

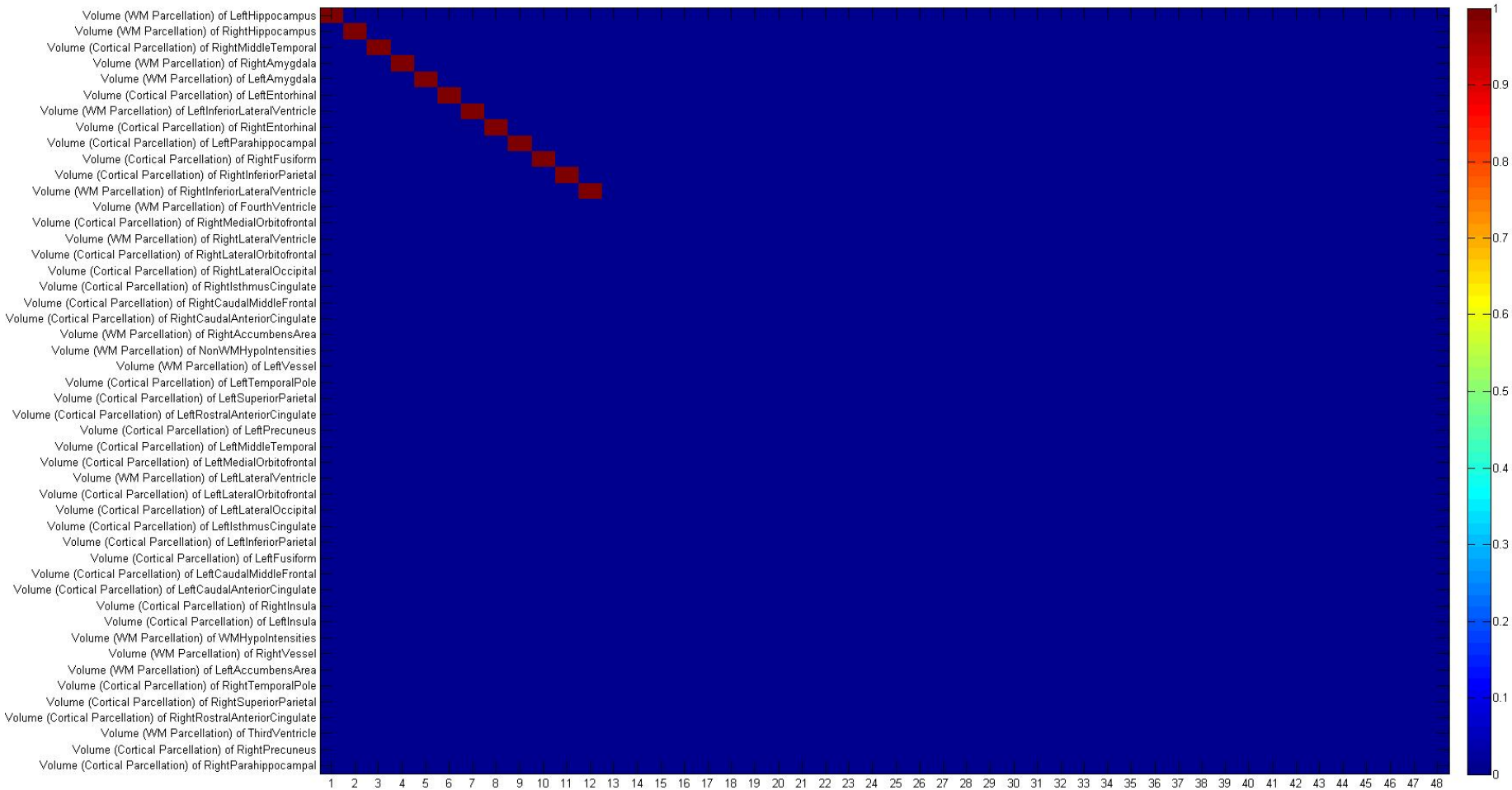


Figura B.6: ADNI1 MCIL vs AD pure most discriminative Features

B.2. Complete Results for the Lausanne Database

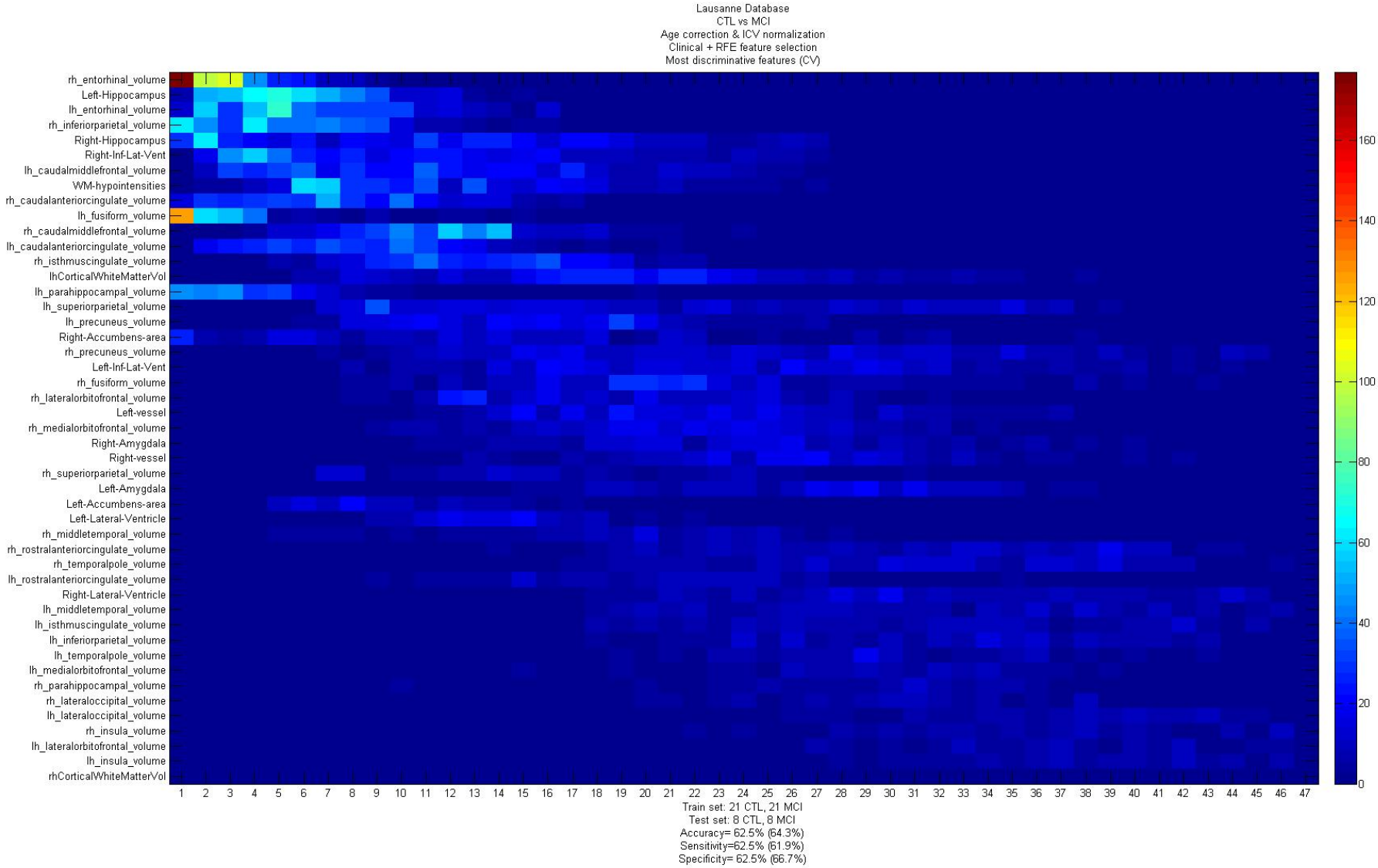


Figura B.7: Lausanne Database CTL vs MCI CV most discriminative Features: GM regional volumes, ICV normalization.

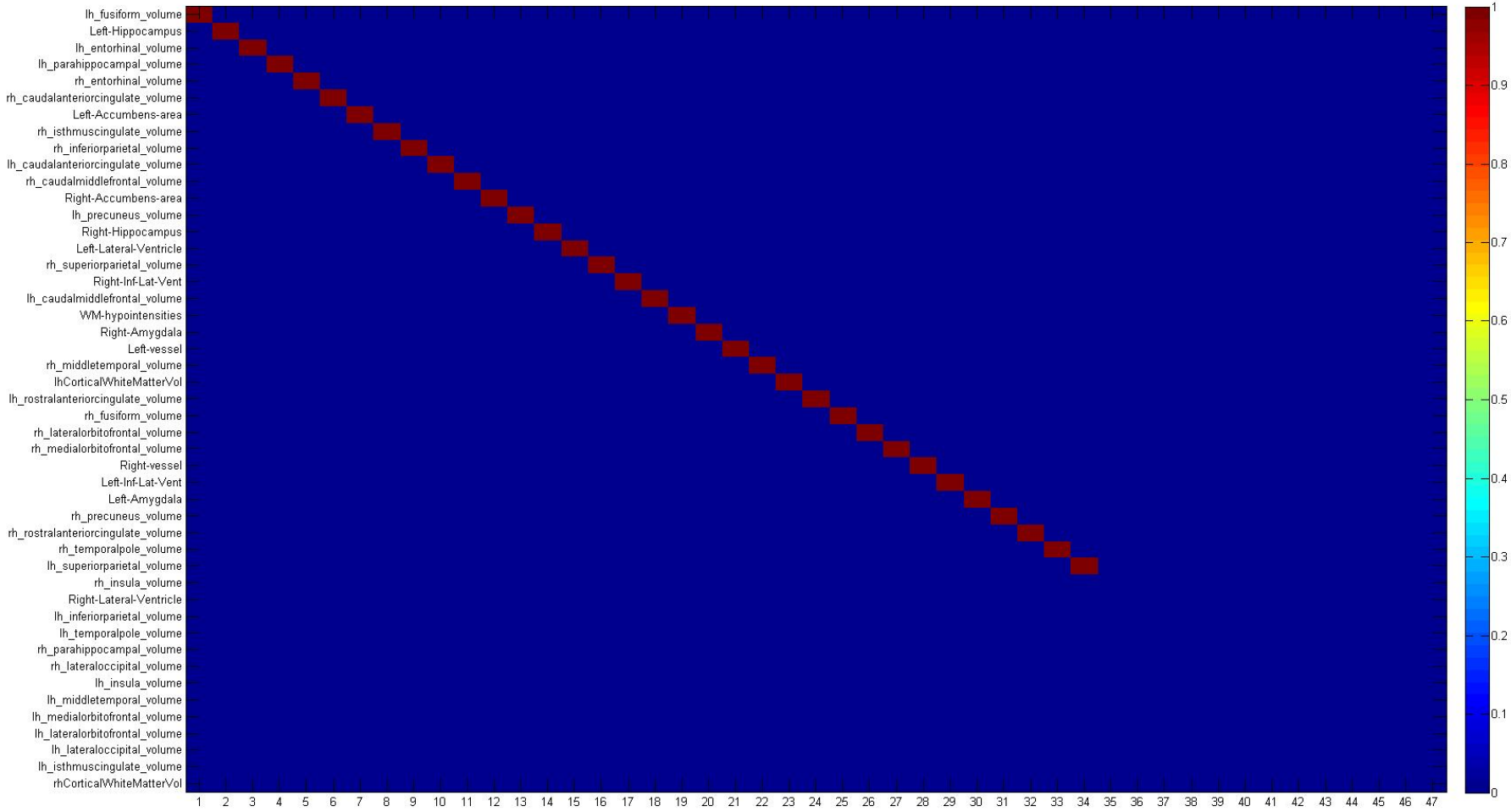


Figura B.8: Lausanne Database CTL vs MCI pure most discriminative Features: GM regional volumes, ICV normalization.

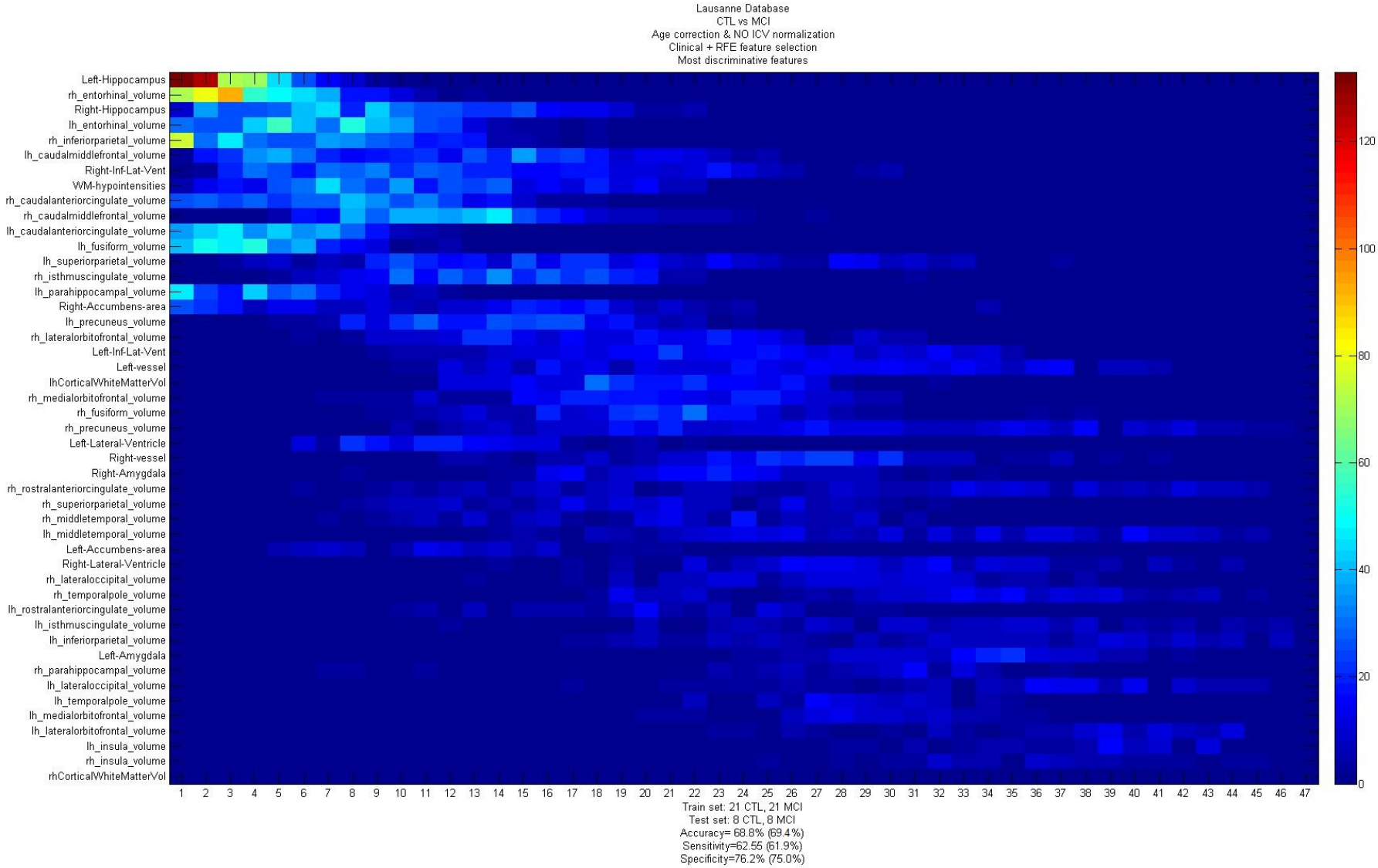


Figura B.9: Lausanne Database CTL vs MCI CV most discriminative Features: GM regional volumes, no ICV normalization.

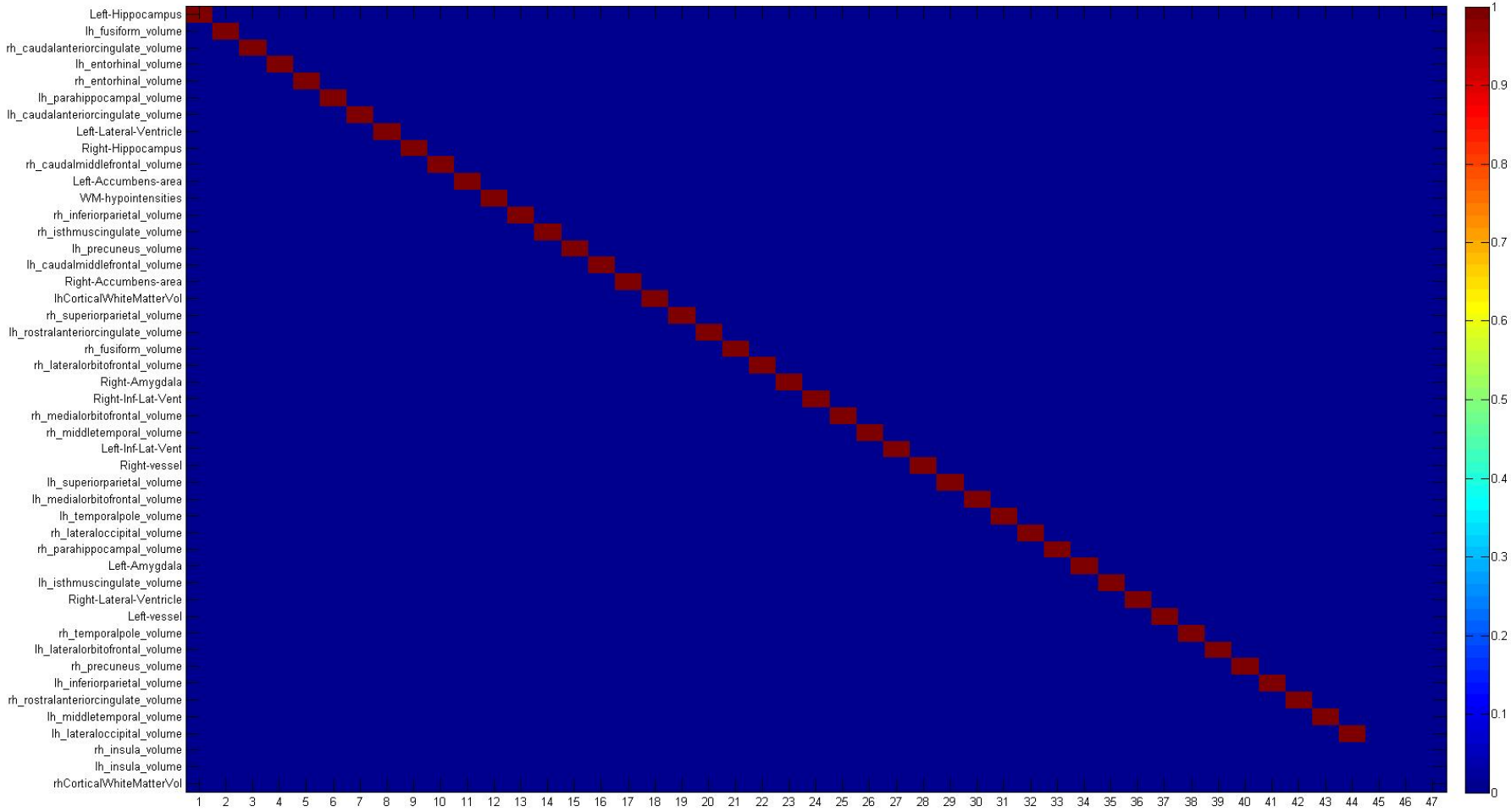


Figura B.10: Lausanne Database CTL vs MCI pure most discriminative Features: GM regional volumes, no ICV normalization.

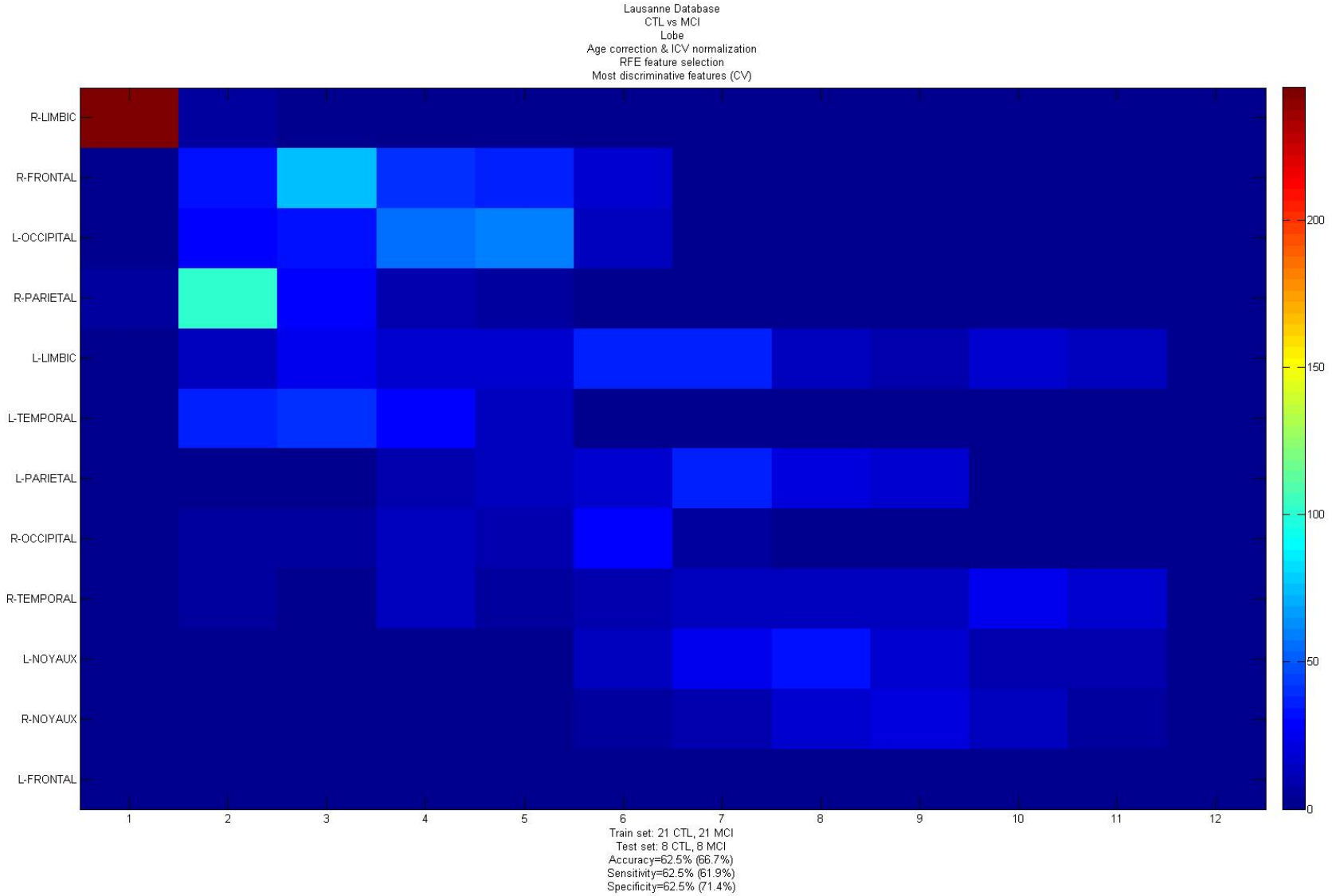


Figura B.11: Lausanne Database CTL vs MCI CV most discriminative Features: GM LOBE volumes, ICV normalization.

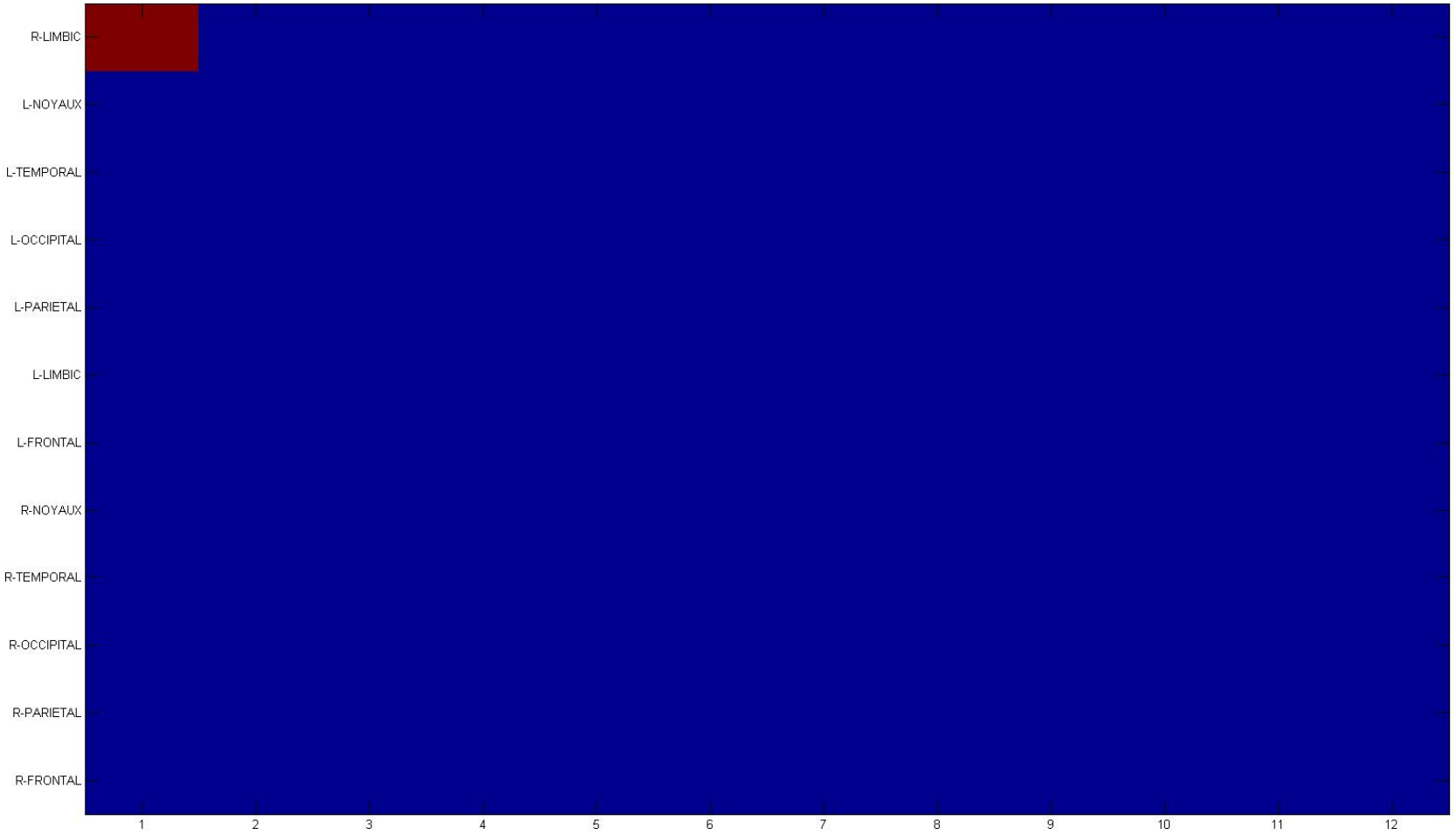


Figura B.12: Lausanne Database CTL vs MCI pure most discriminative Features: GM LOBE volumes, ICV normalization.

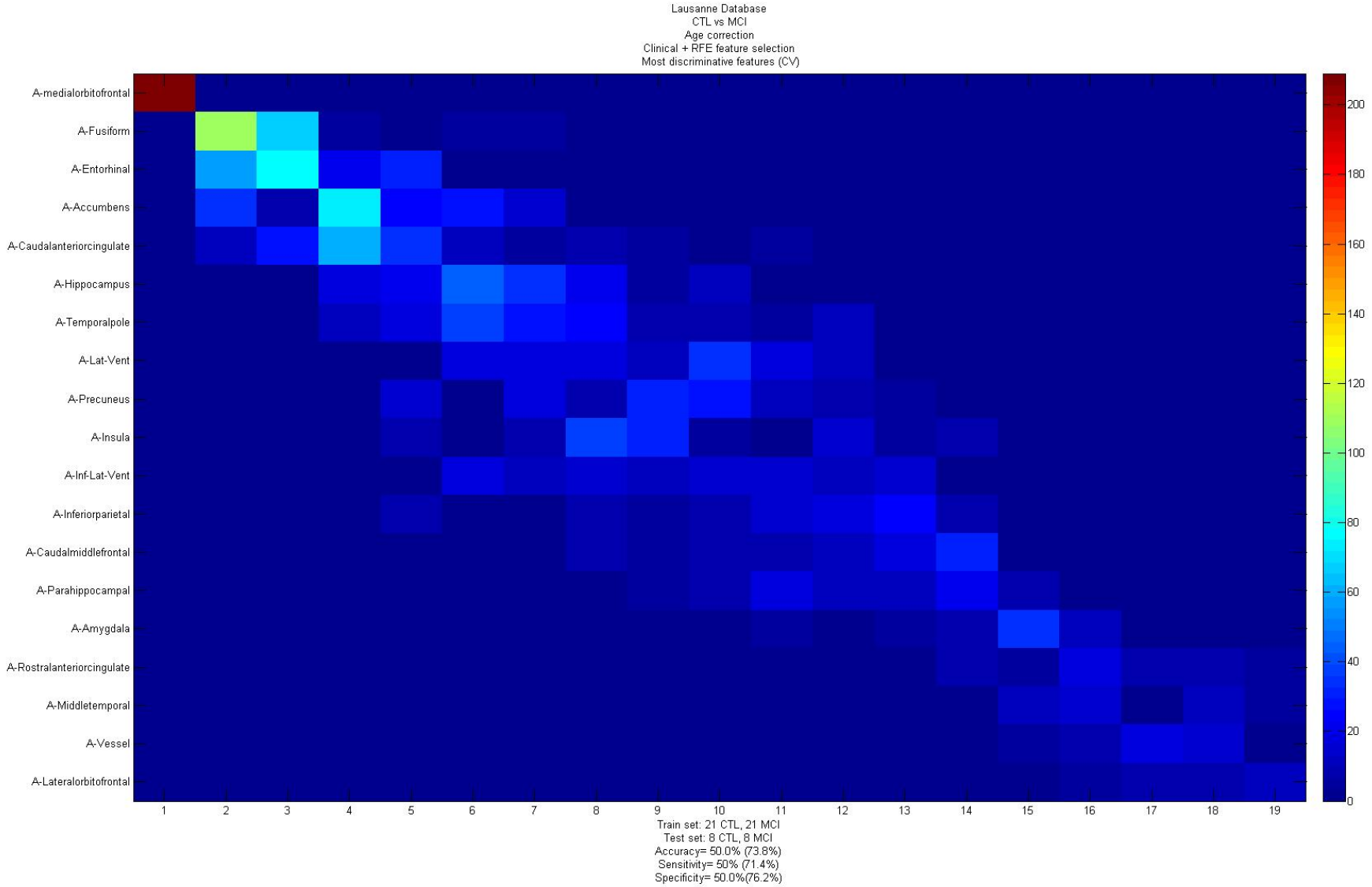


Figura B.13: Lausanne Database CTL vs MCI CV most discriminative Features: Asymmetry coefficient.

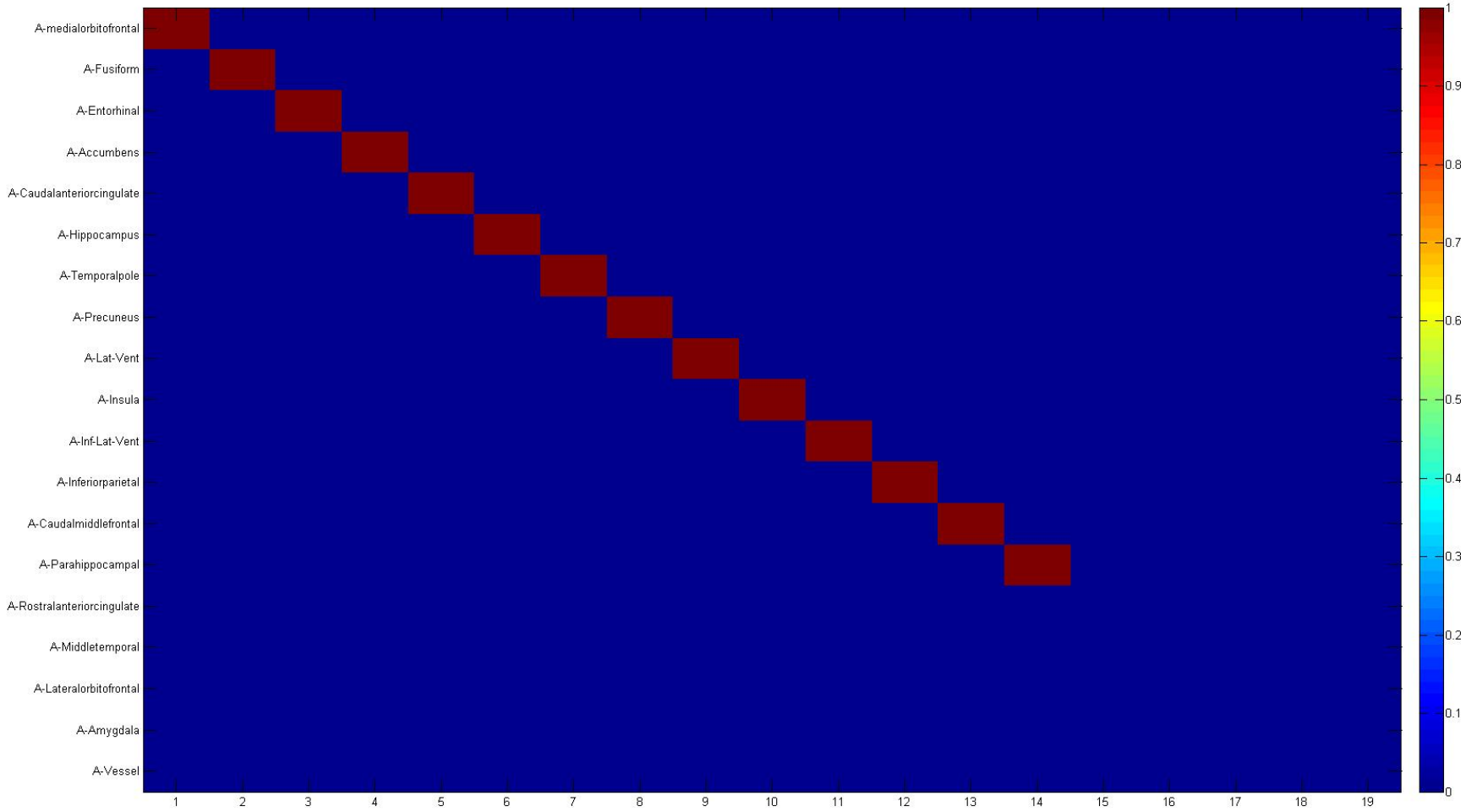


Figura B.14: Lausanne Database CTL vs MCI pure most discriminative Features: Asymmetry coefficient.

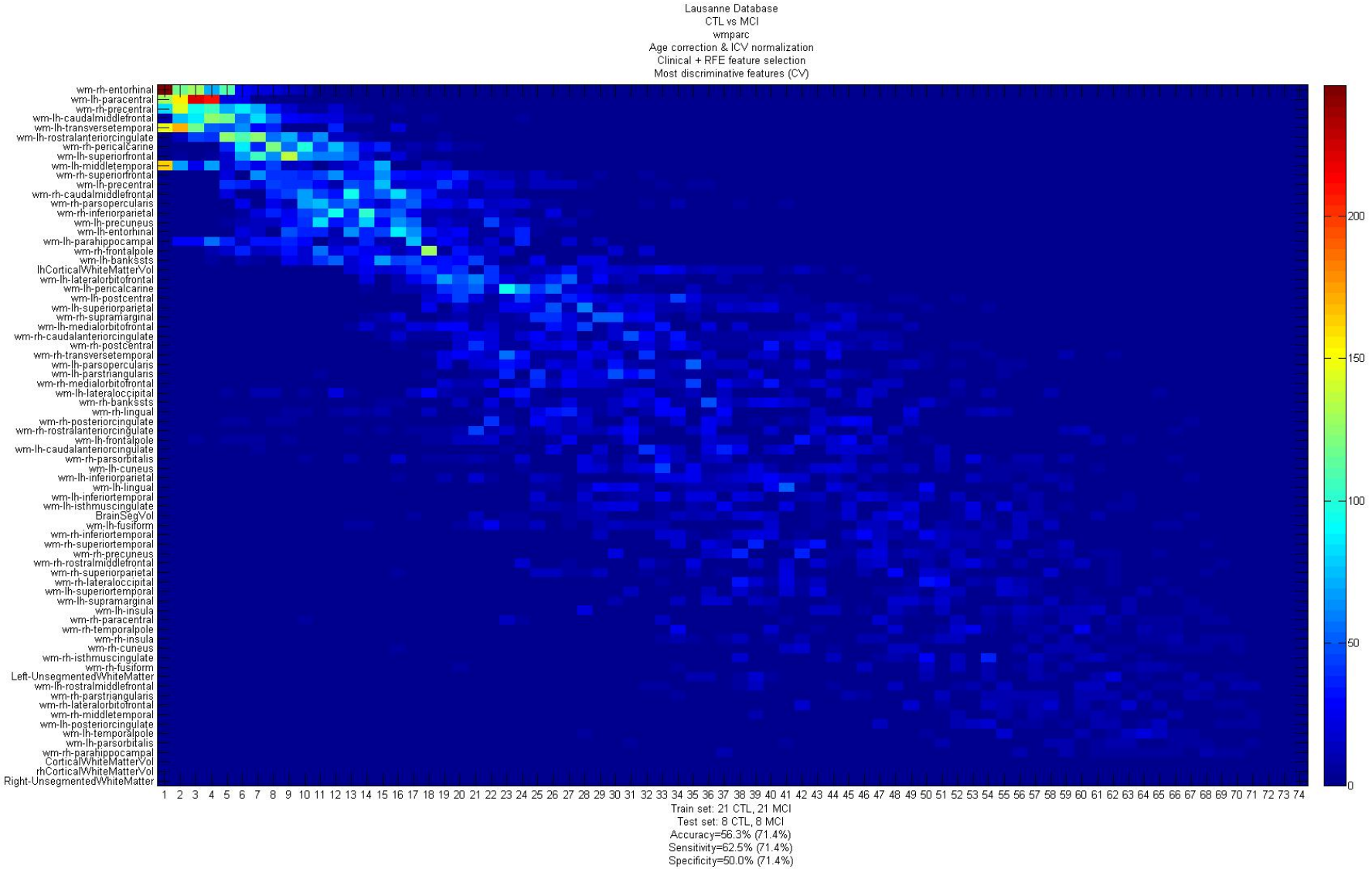


Figura B.15: Lausanne Database CTL vs MCI CV most discriminative Features: WM, ICV normalization.

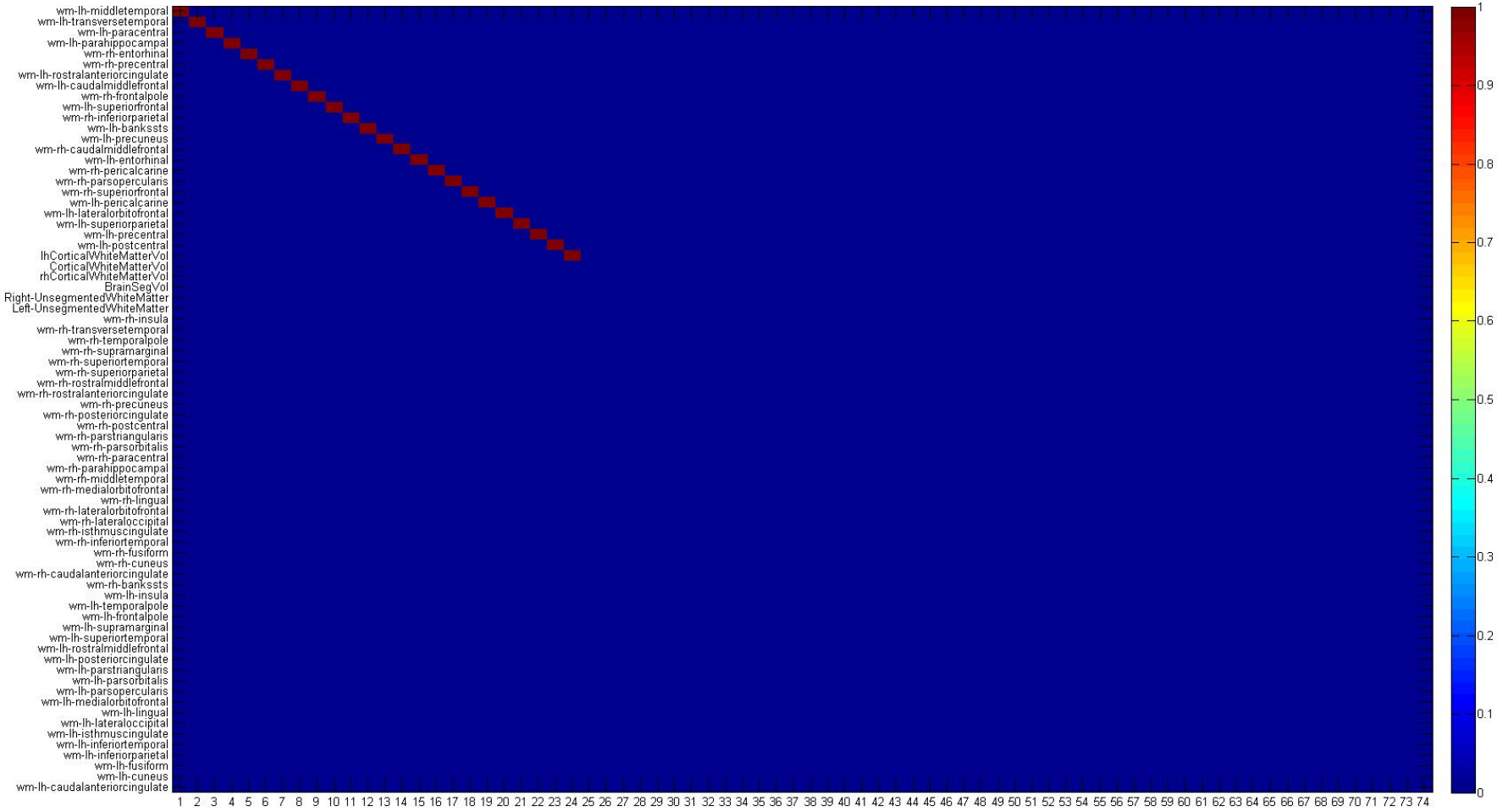


Figura B.16: Lausanne Database CTL vs MCI pure most discriminative Features: WM, ICV normalization.

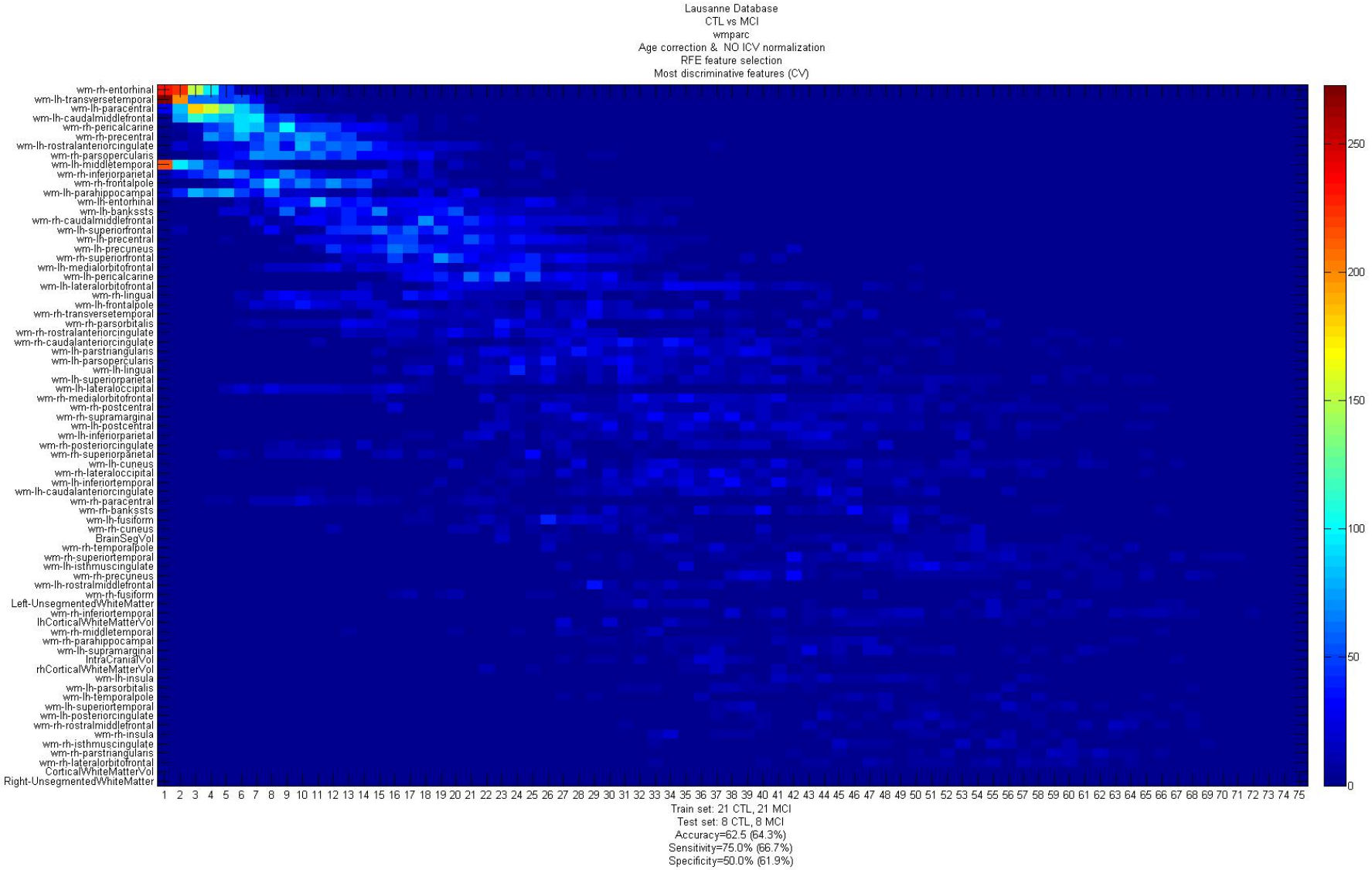


Figura B.17: Lausanne Database CTL vs MCI CV most discriminative Features: WM, no ICV normalization.

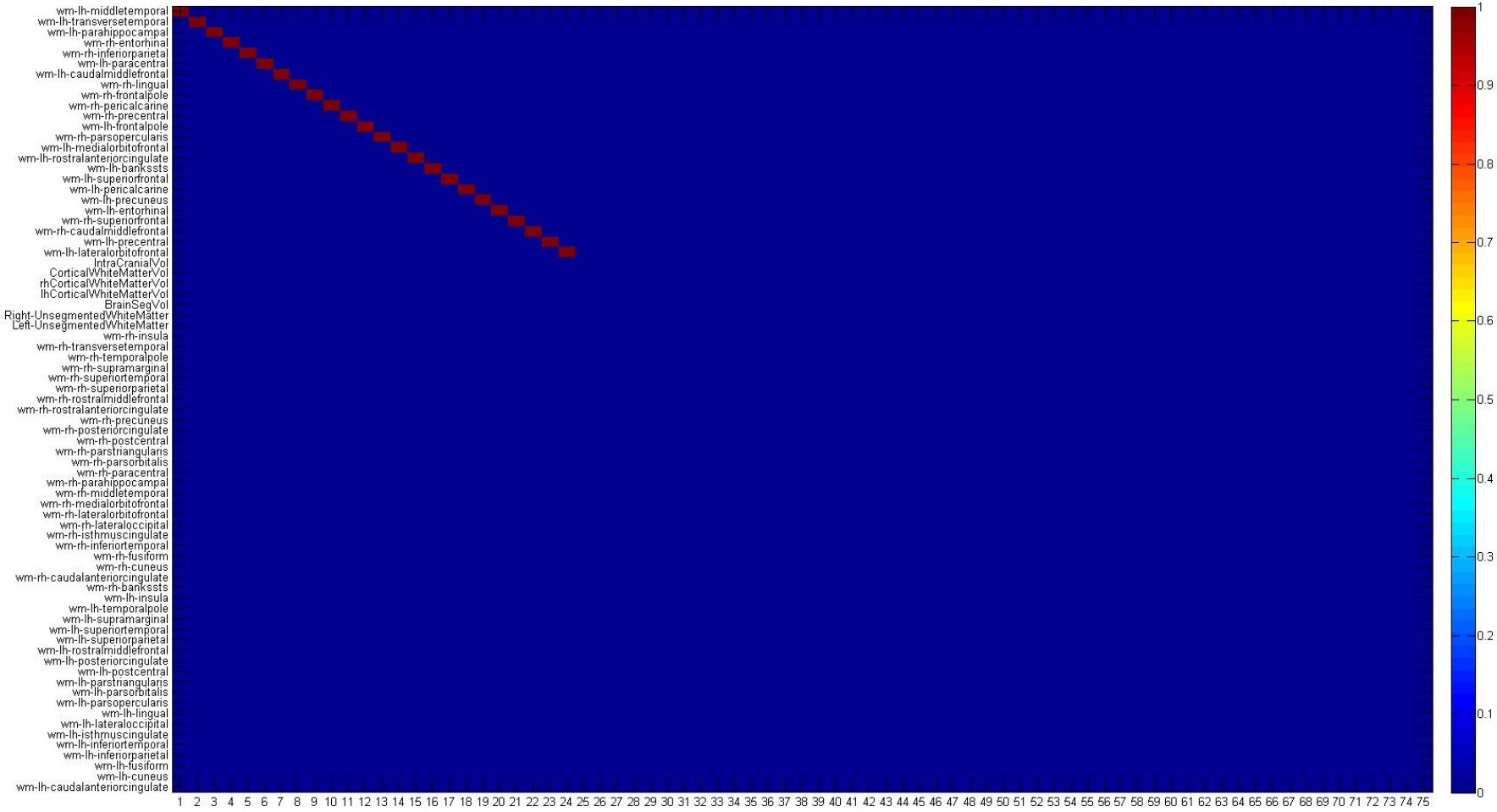


Figura B.18: Lausanne Database CTL vs MCI pure most discriminative Features: WM, no ICV normalization.

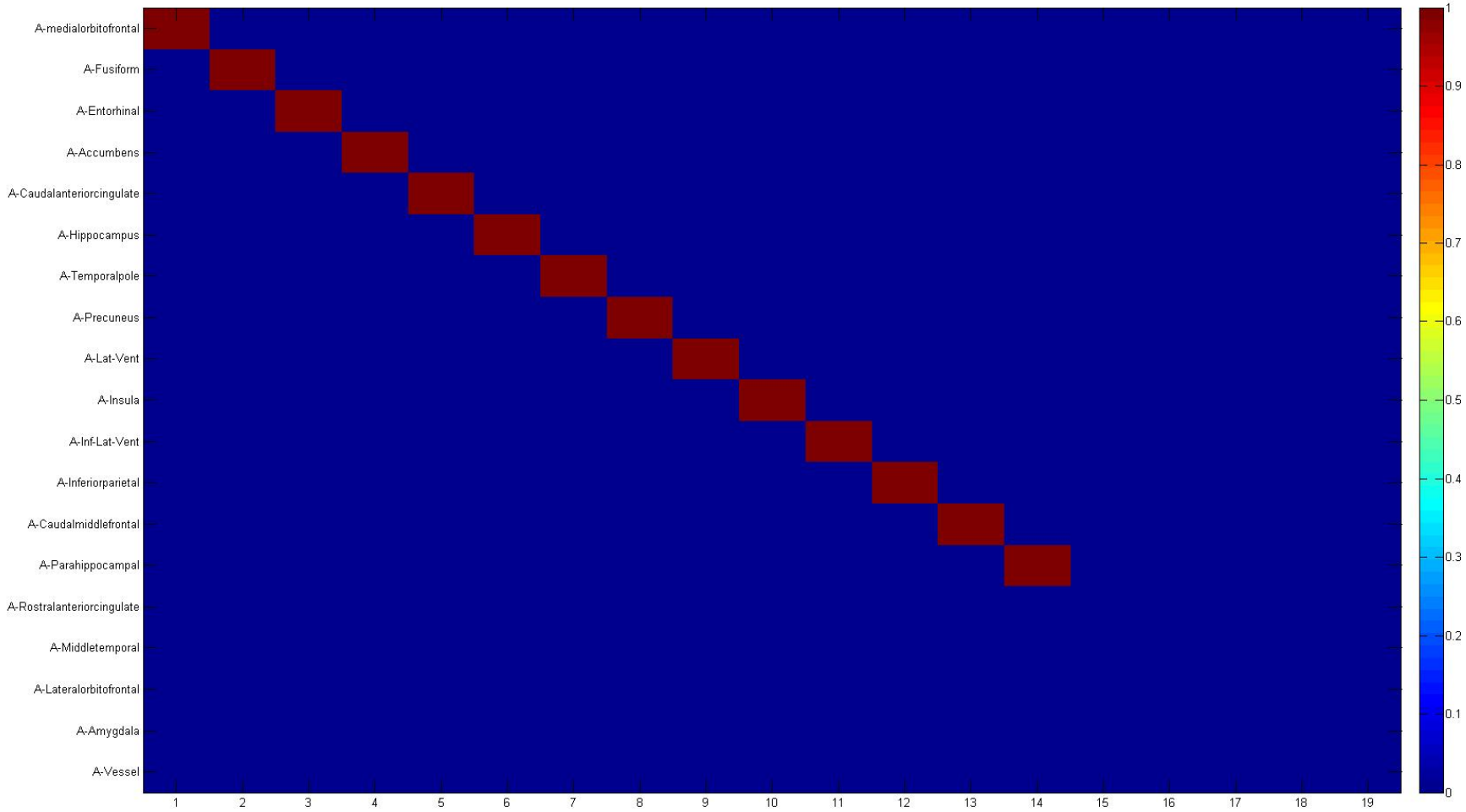


Figura B.19: Lausanne Database CTL vs MCI CV most discriminative Features: WM+ LOBE, ICV normalization.

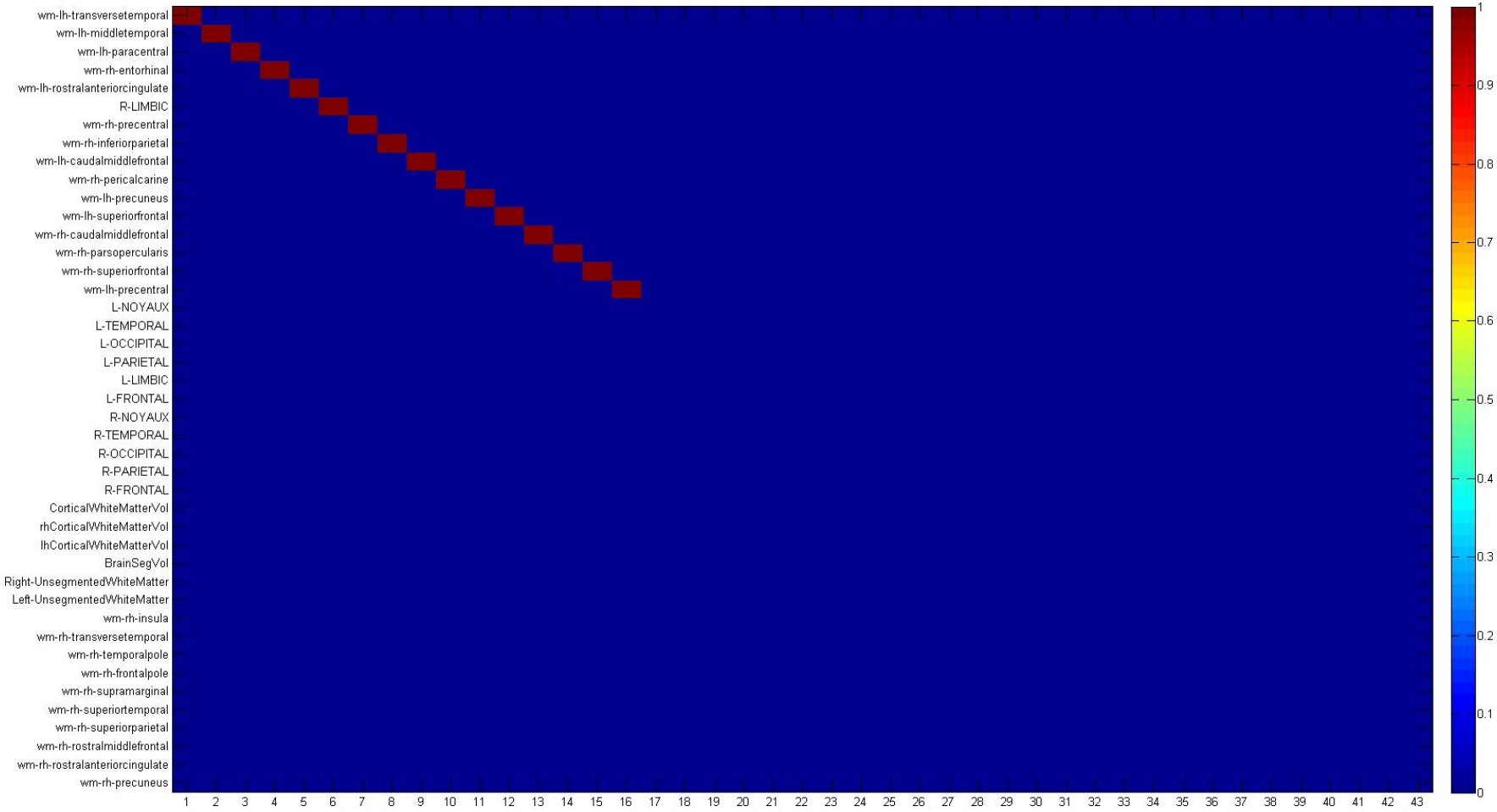


Figura B.20: Lausanne Database CTL vs MCI pure most discriminative Features: WM+ LOBE, ICV normalization.

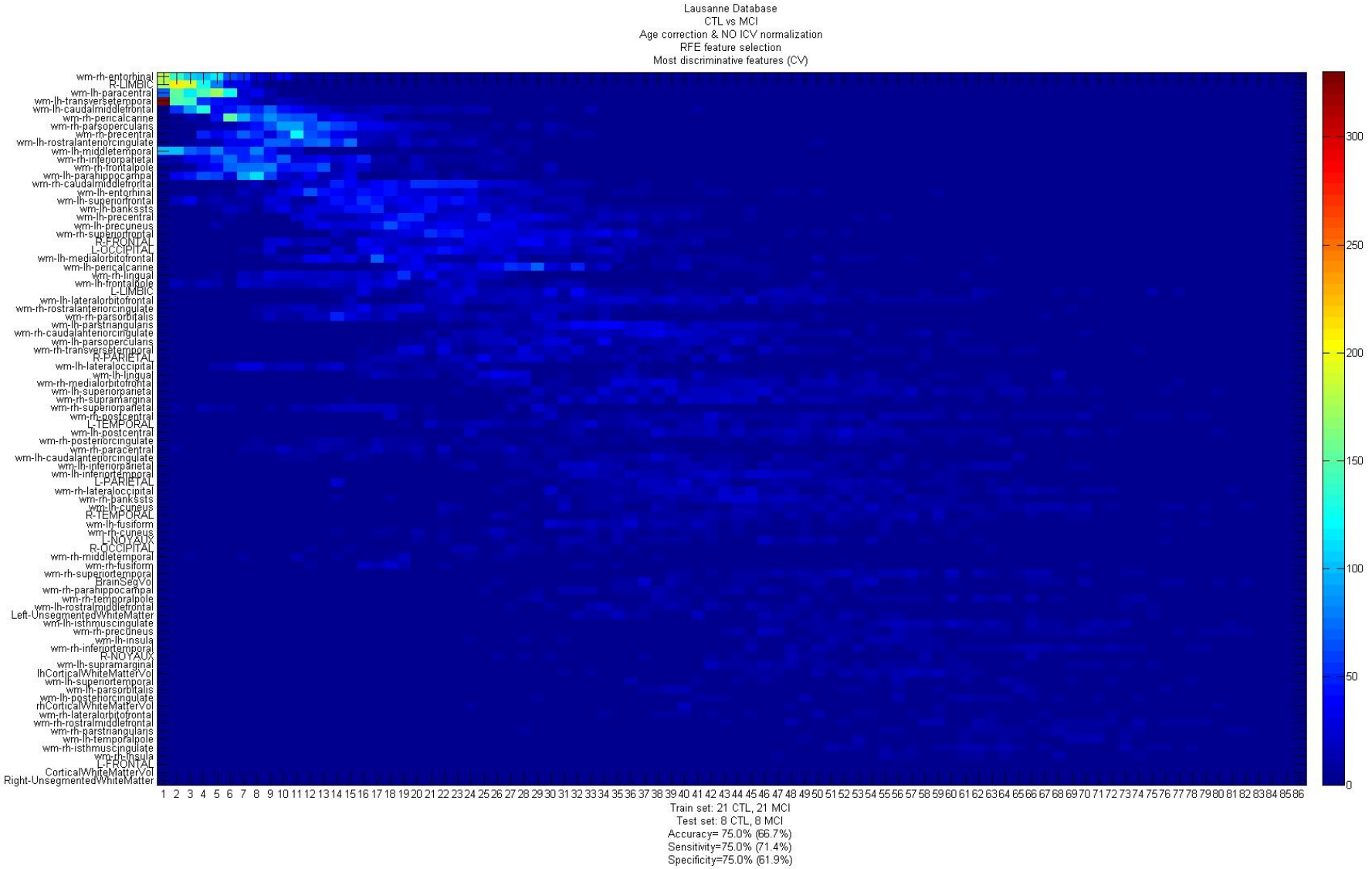


Figura B.21: Lausanne Database CTL vs MCI CV most discriminative Features: WM+ LOBE, no ICV normalization.

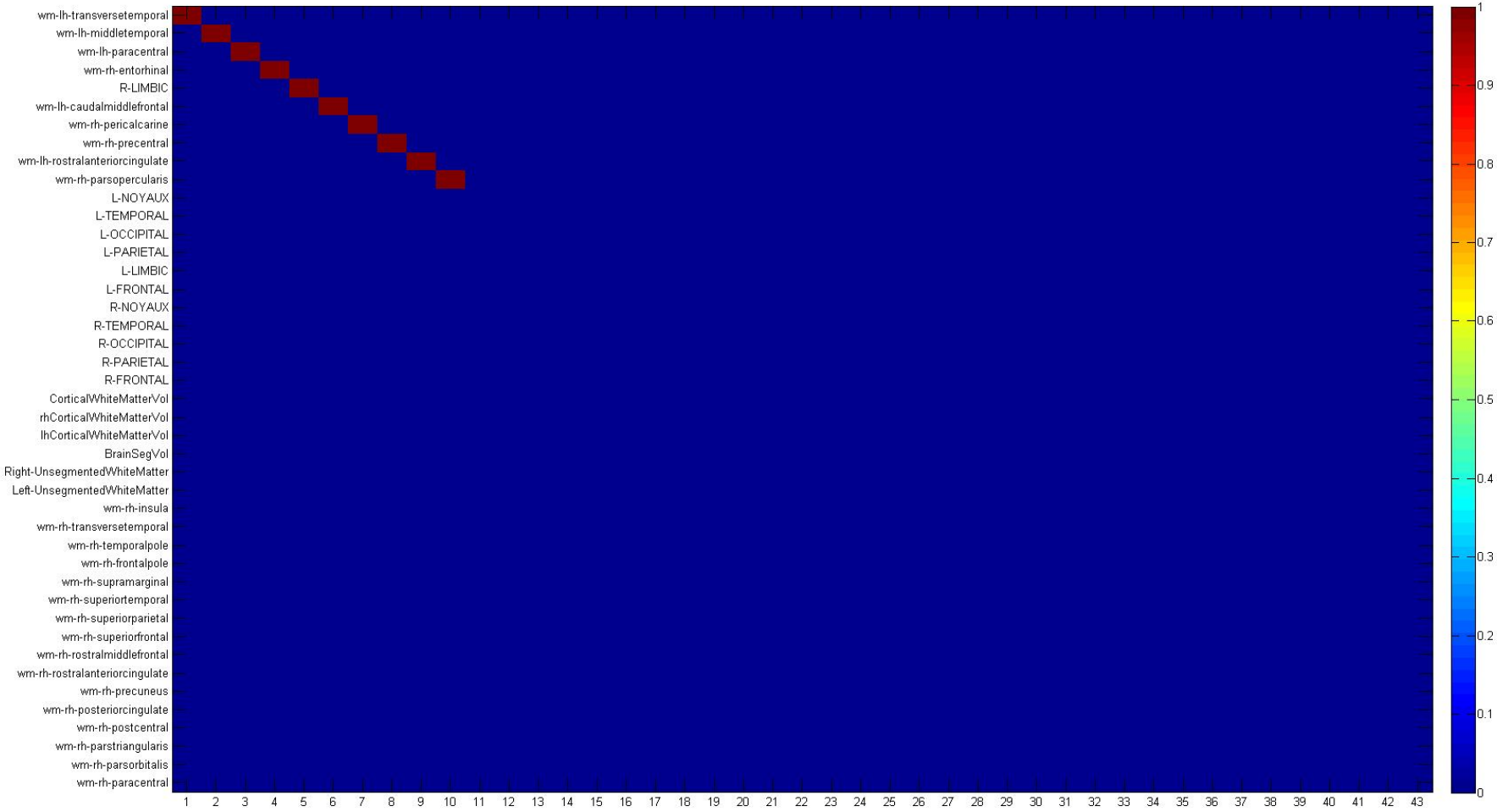


Figura B.22: Lausanne Database CTL vs MCI pure most discriminative Features: WM+ LOBE, ICV normalization.

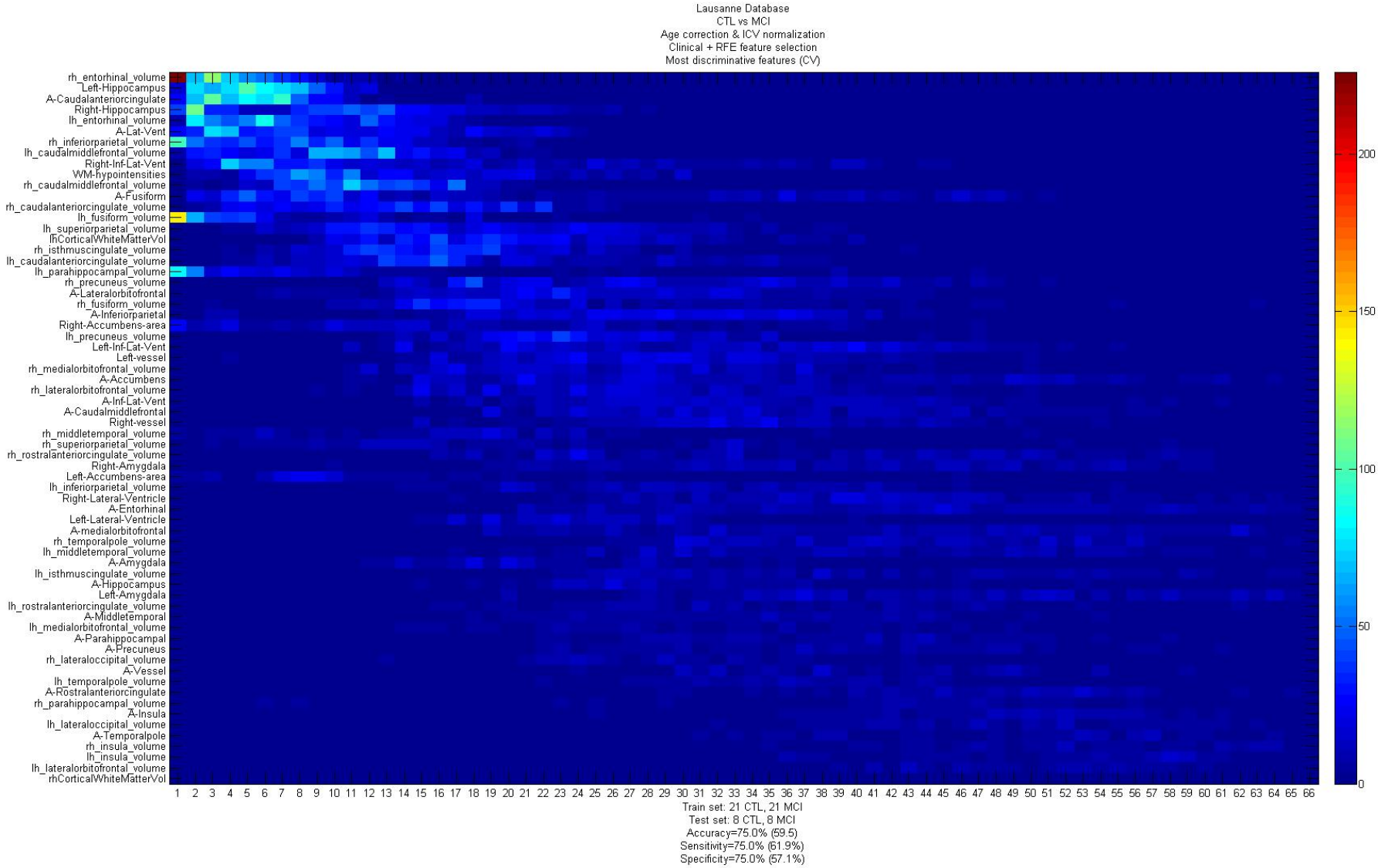


Figura B.23: Lausanne Database CTL vs MCI CV most discriminative Features: GM regional volumes+ ASY, ICV normalization.

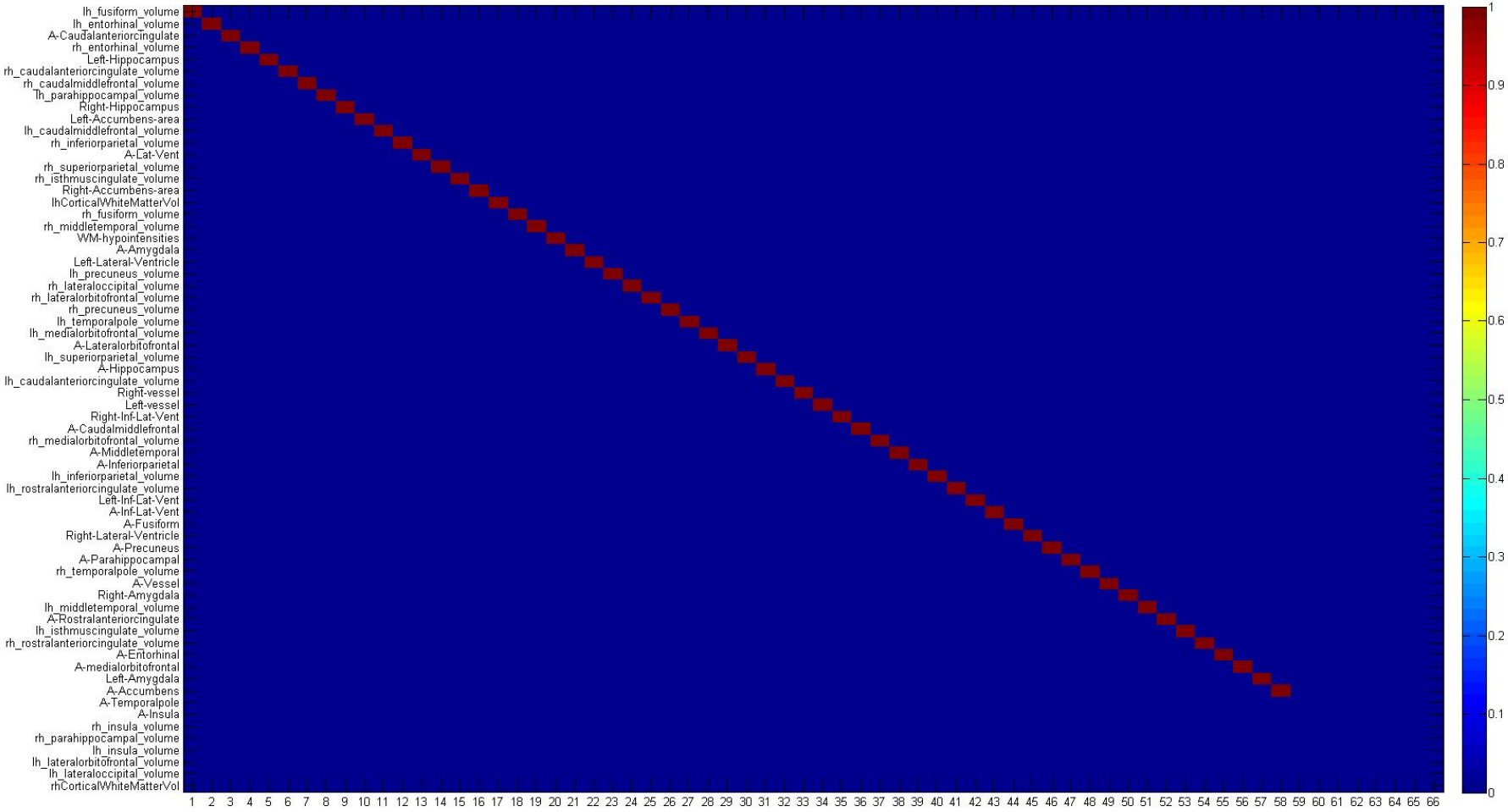


Figura B.24: Lausanne Database CTL vs MCI pure most discriminative Features: GM regional volumes+ ASY, ICV normalization.

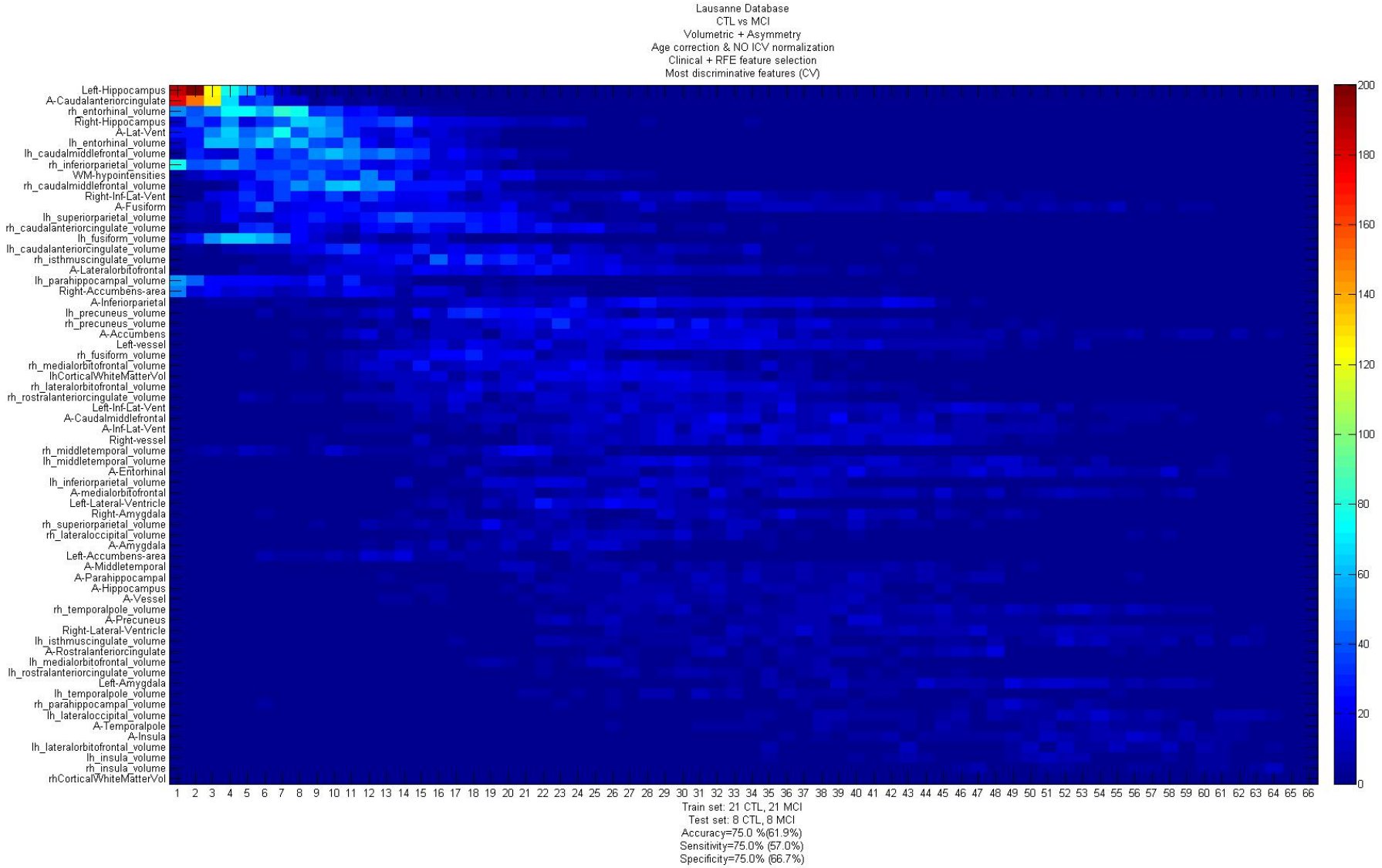


Figura B.25: Lausanne Database CTL vs MCI CV most discriminative Features: GM regional volumes+ ASY, no ICV normalization.

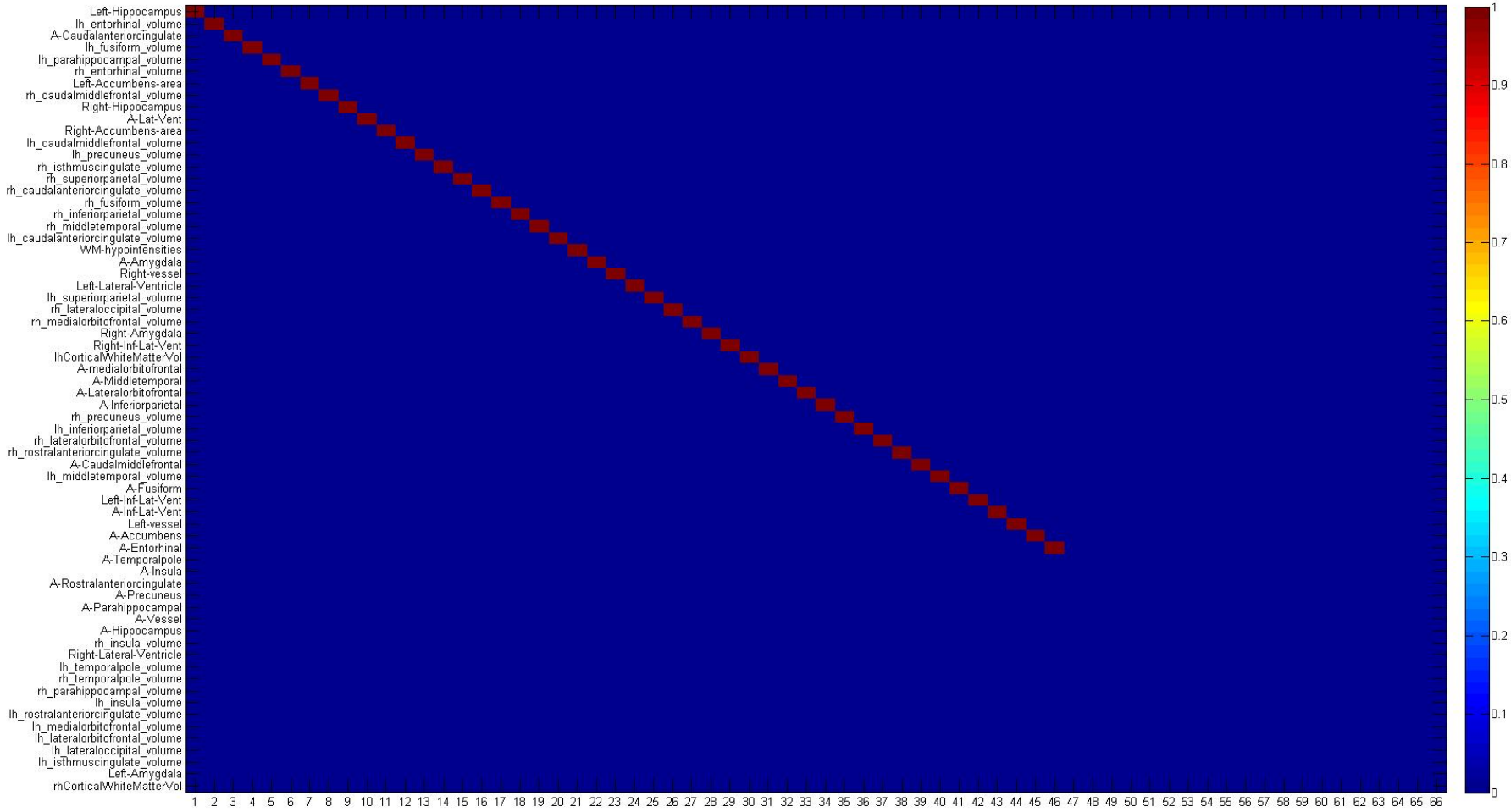


Figura B.26: Lausanne Database CTL vs MCI pure most discriminative Features: GM regional volumes+ ASY, ICV normalization.

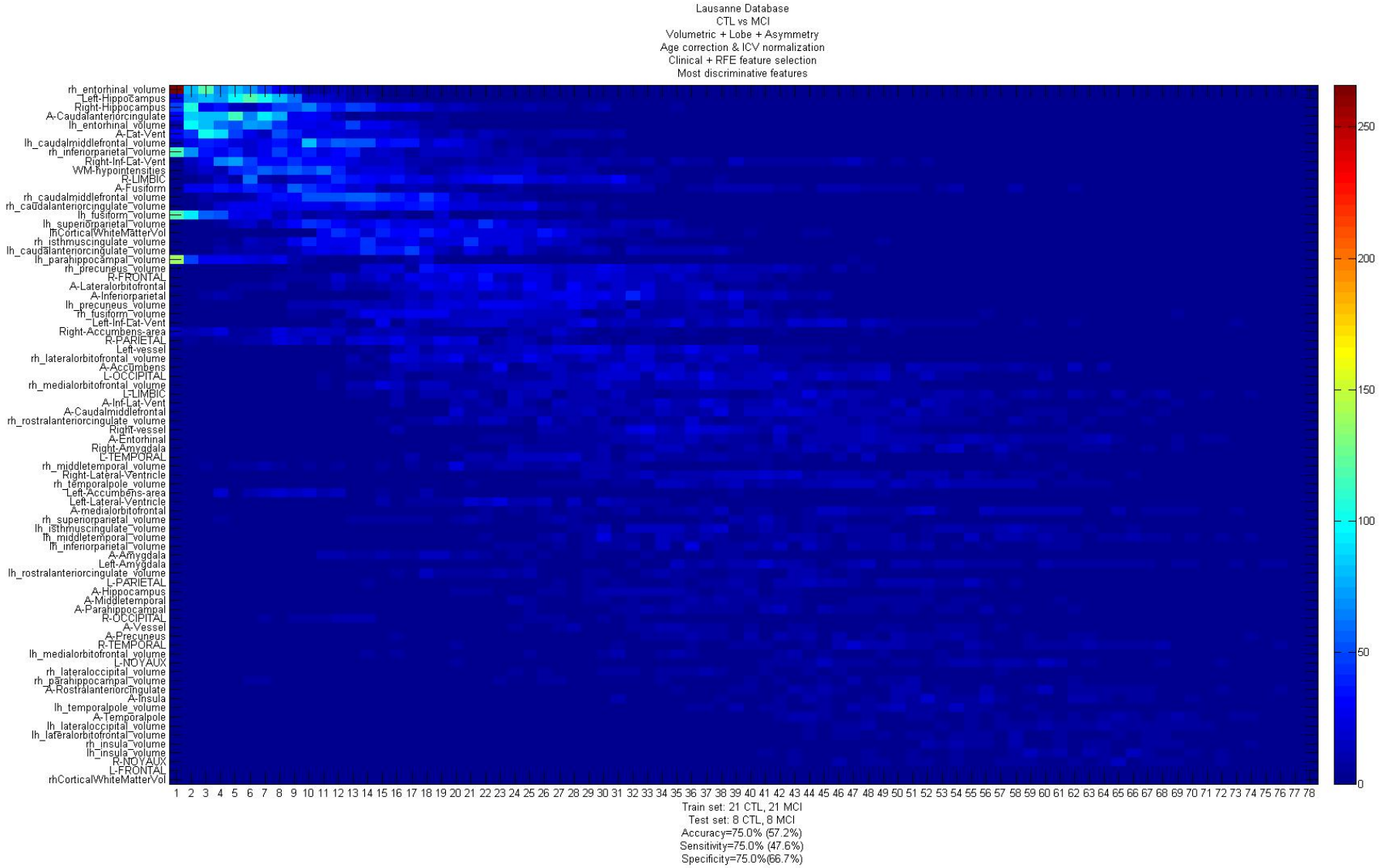


Figura B.27: Lausanne Database CTL vs MCI CV most discriminative Features: GM regional volumes+ ASY+ Lobe, ICV normalization.

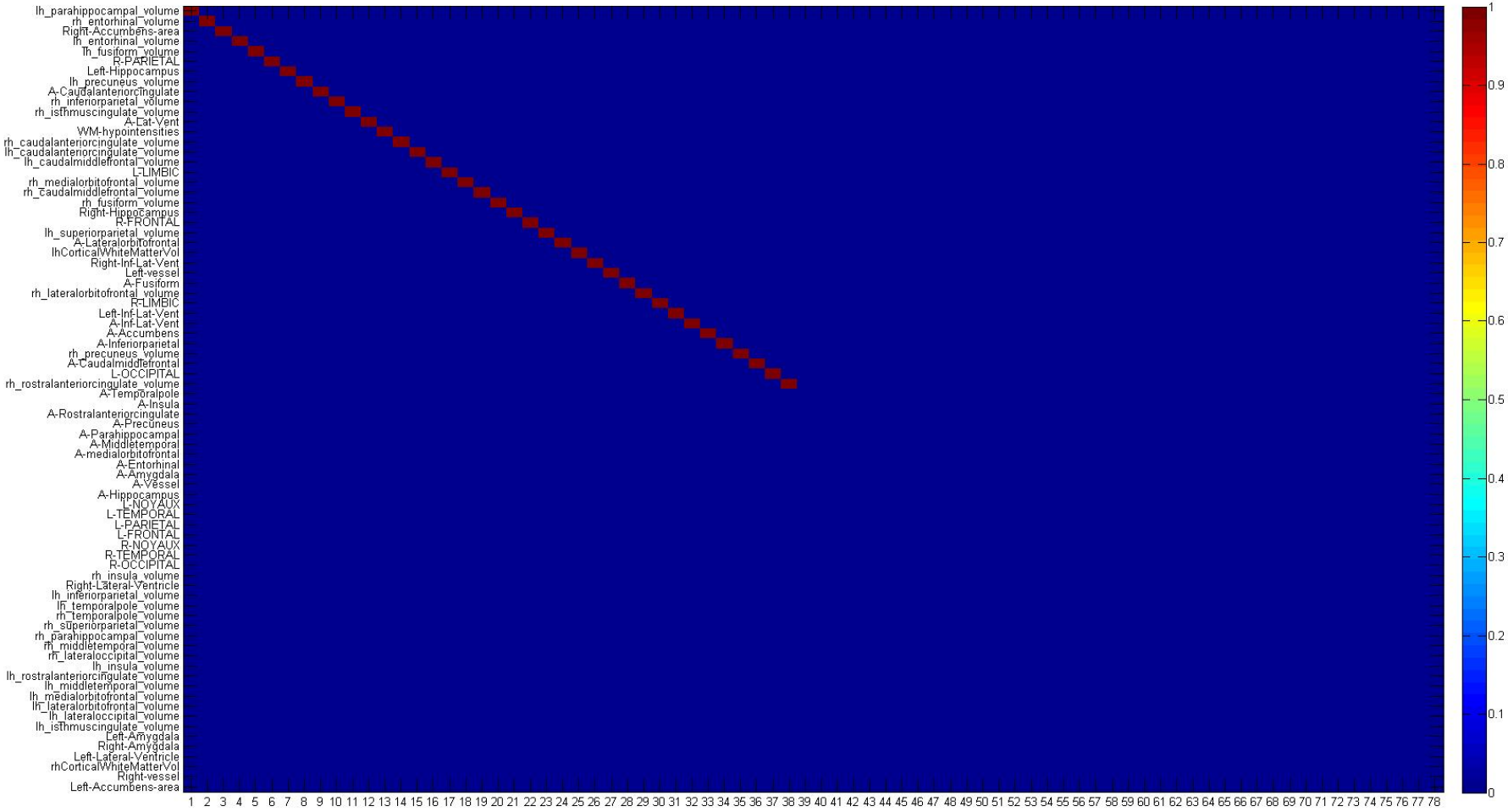


Figura B.28: Lausanne Database CTL vs MCI pure most discriminative Features: GM regional volumes+ ASY+ Lobe, ICV normalization.

B.3. Complete Results for the Expanded Database

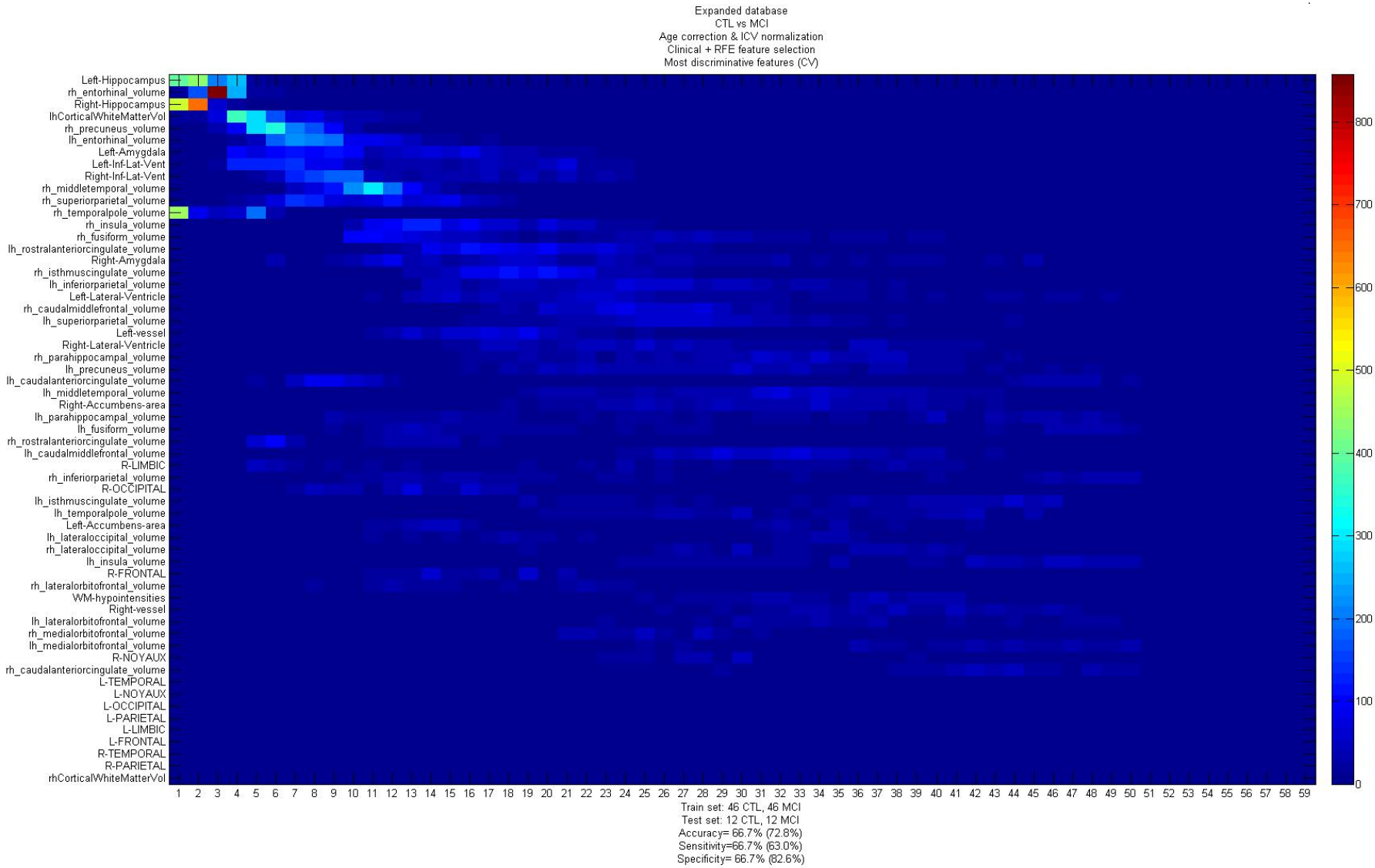


Figura B.29: Expanded Database CTL vs MCI CV most discriminative Features: GM regional volumes, ICV normalization.

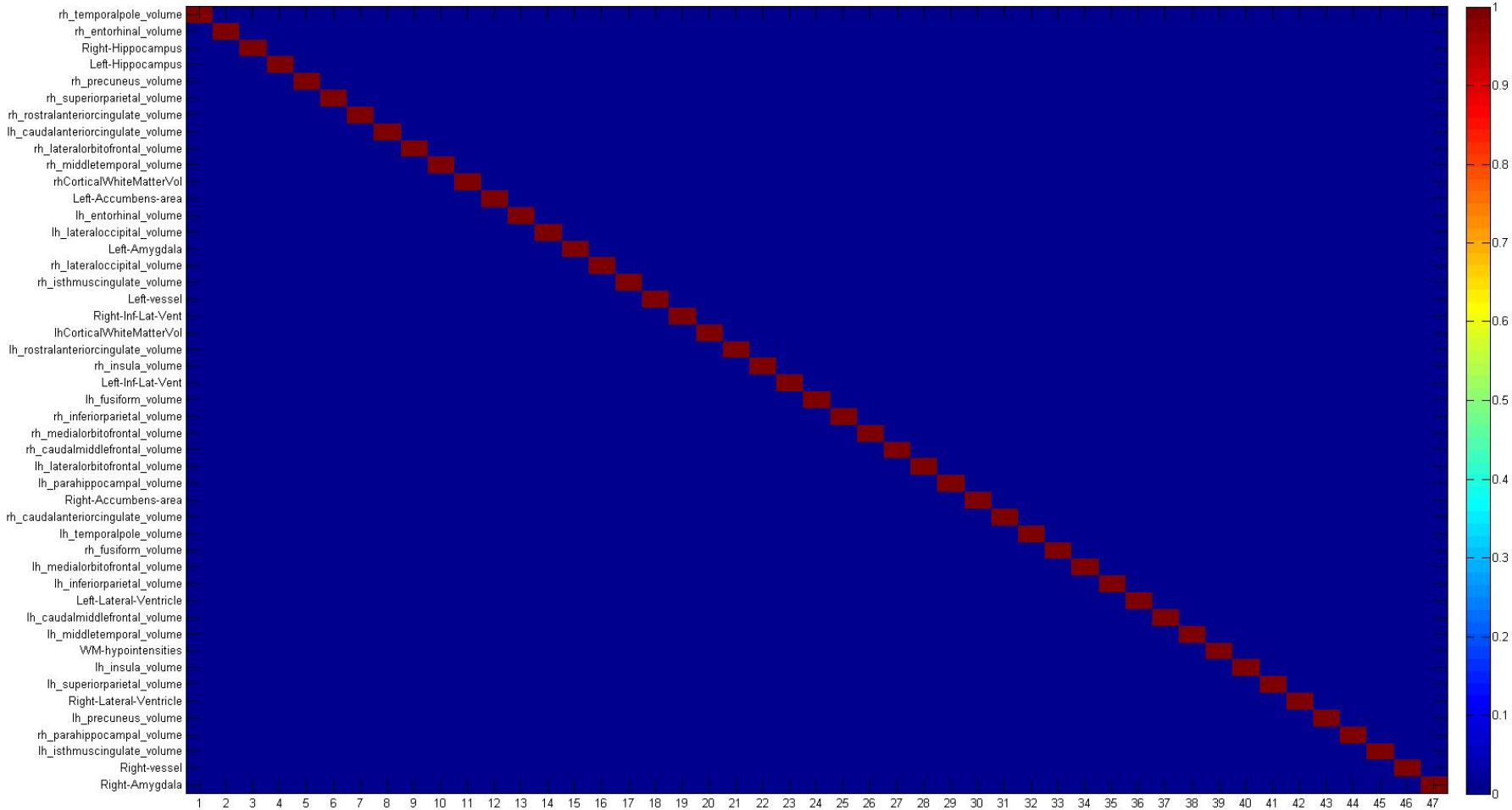


Figura B.30: Expanded Database CTL vs MCI pure most discriminative Features: GM regional volumes, ICV normalization.

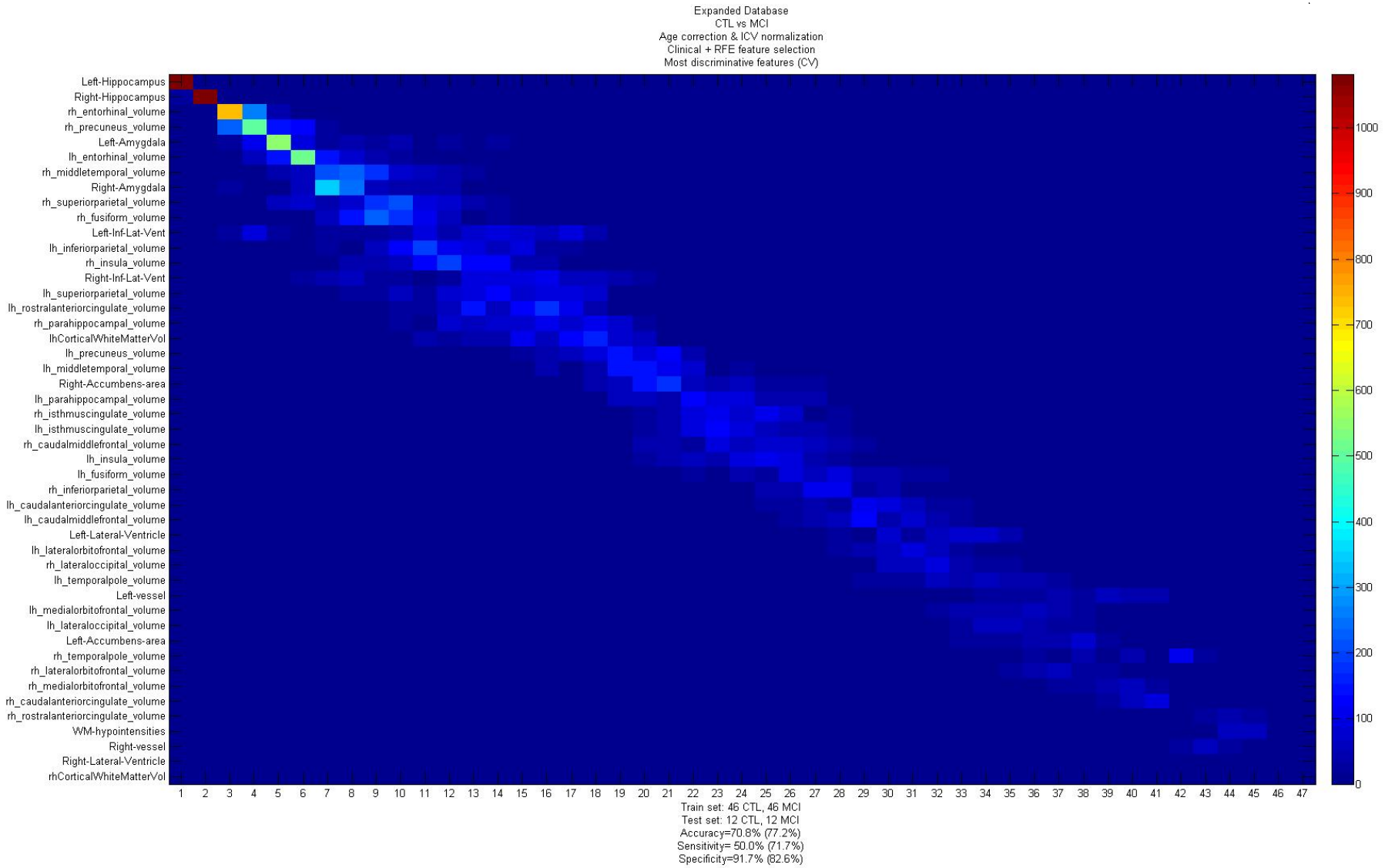


Figura B.31: Expanded Database CTL vs MCI CV most discriminative Features: GM regional volumes, no ICV normalization.

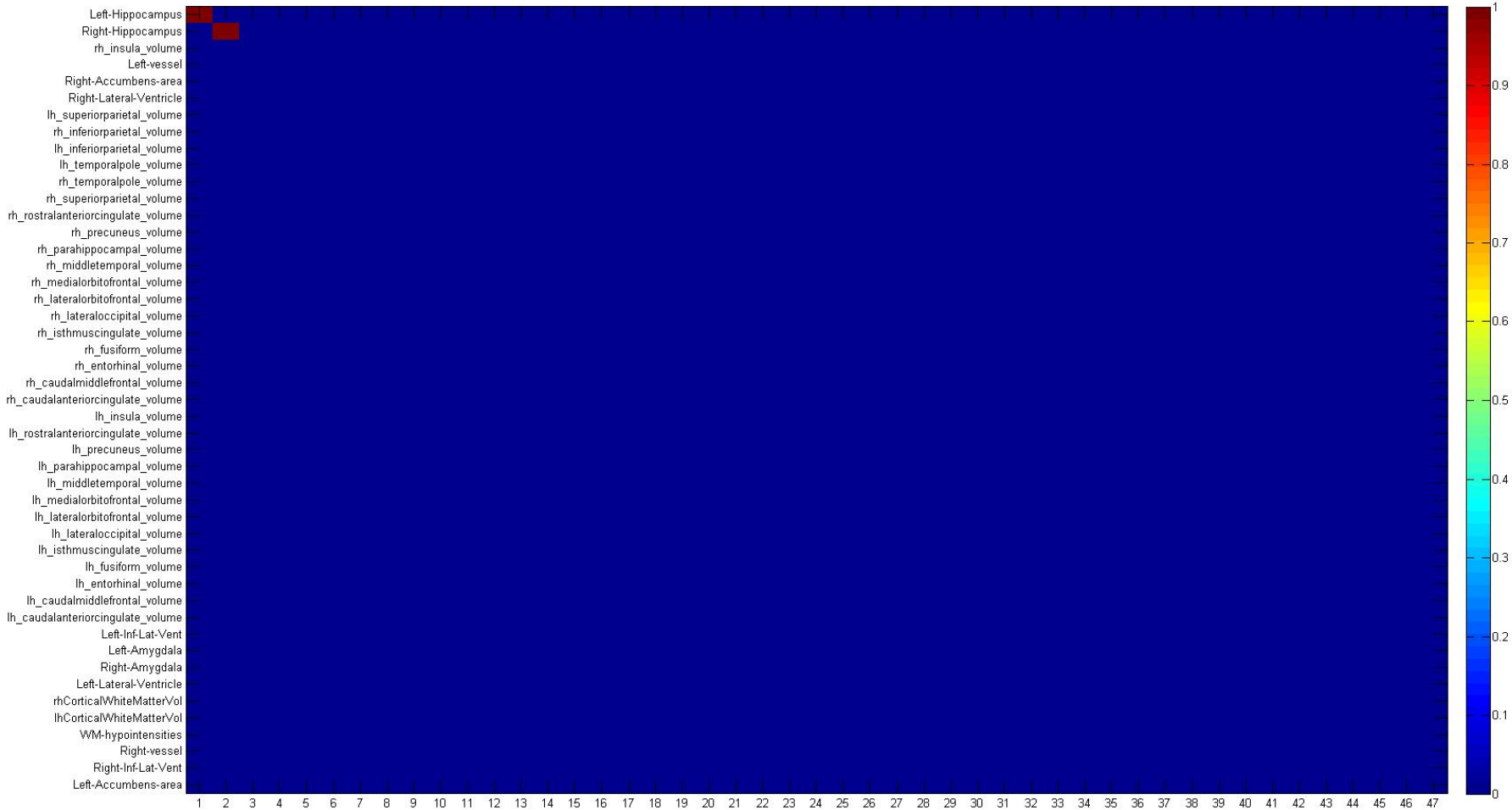


Figura B.32: Expanded Database CTL vs MCI pure most discriminative Features: GM regional volumes, no ICV normalization.

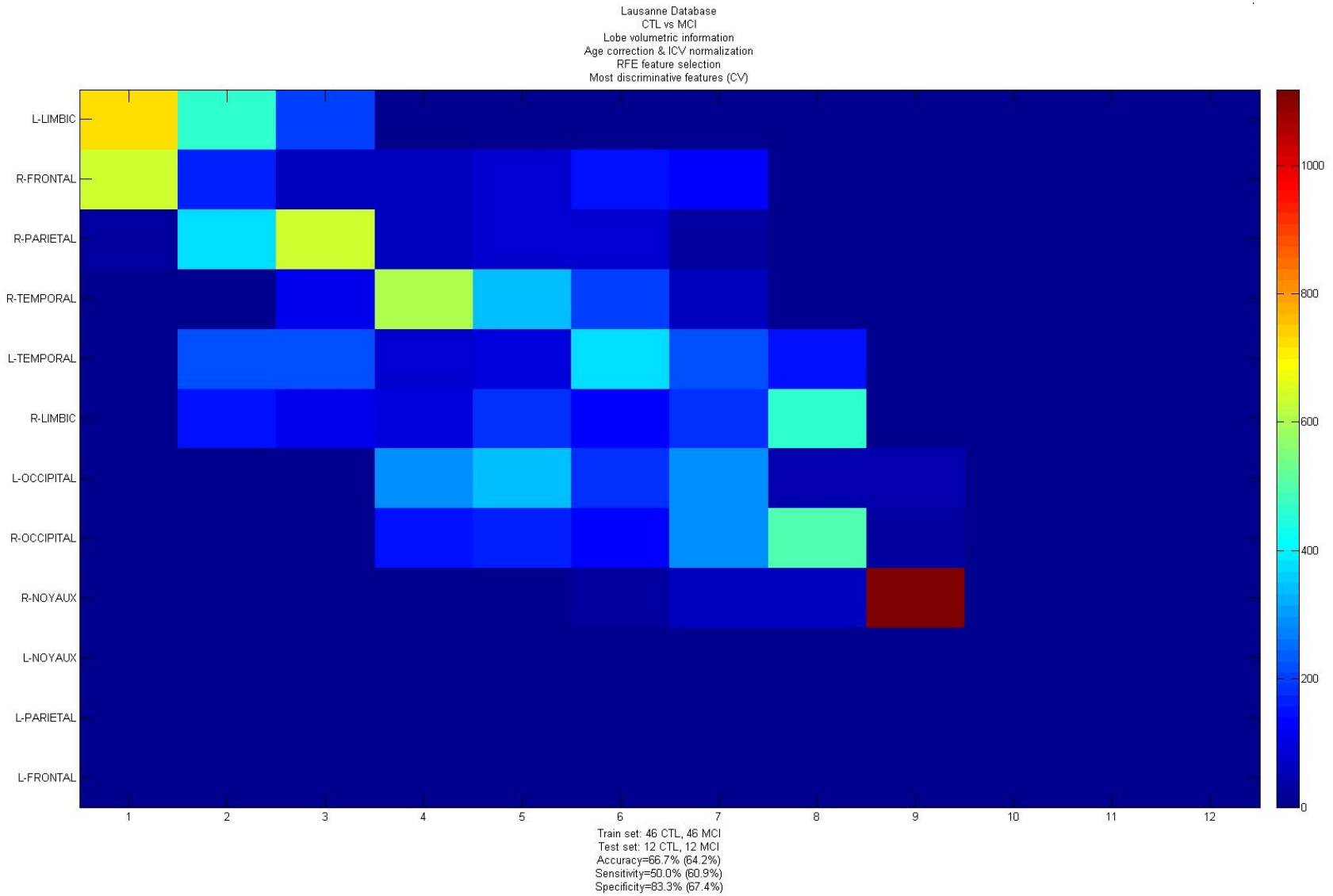


Figura B.33: Expanded Database CTL vs MCI CV most discriminative Features: GM lobe volumes, ICV normalization.

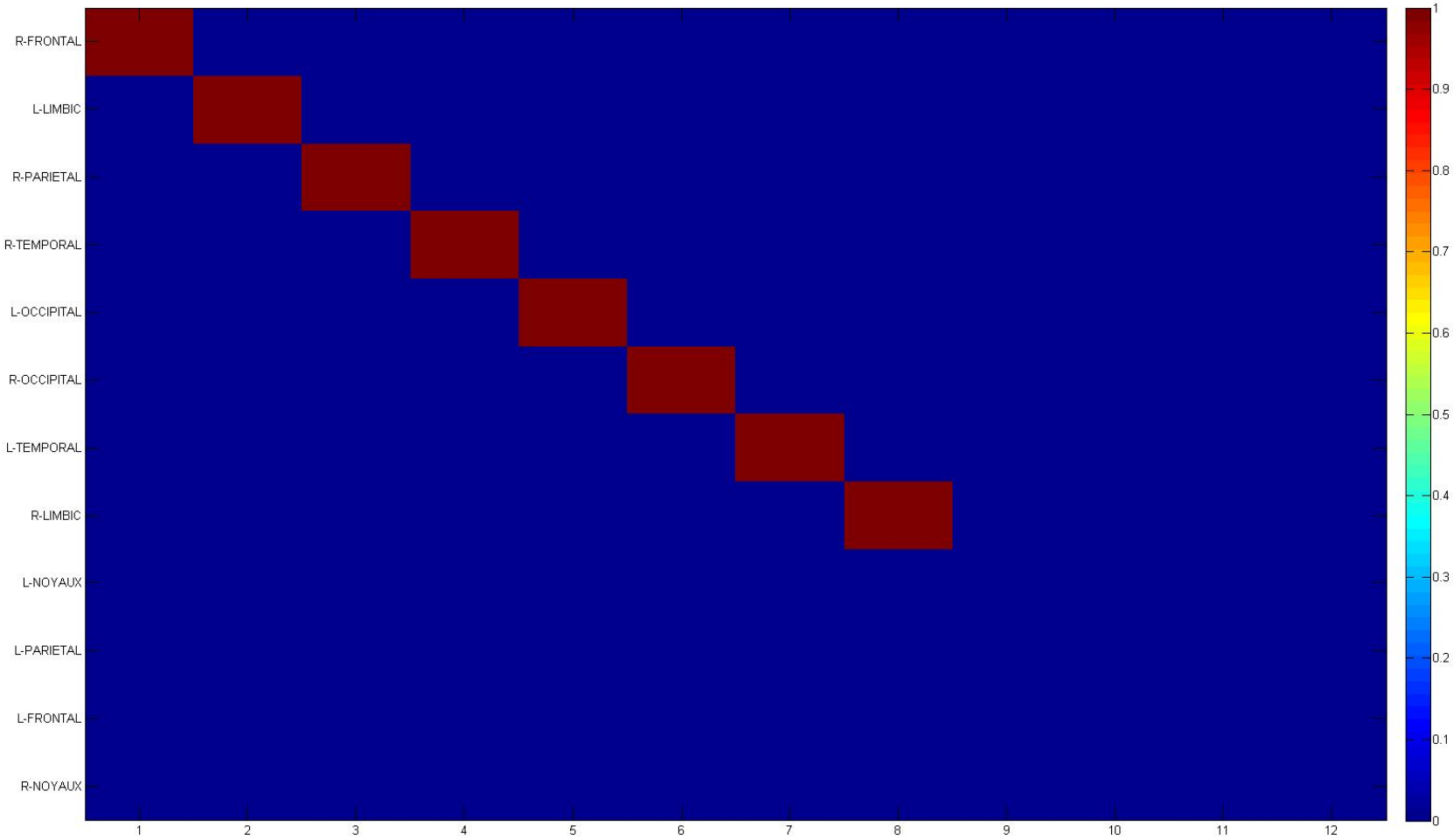


Figura B.34: Expanded Database CTL vs MCI pure most discriminative Features: GM lobe volumes, ICV normalization.

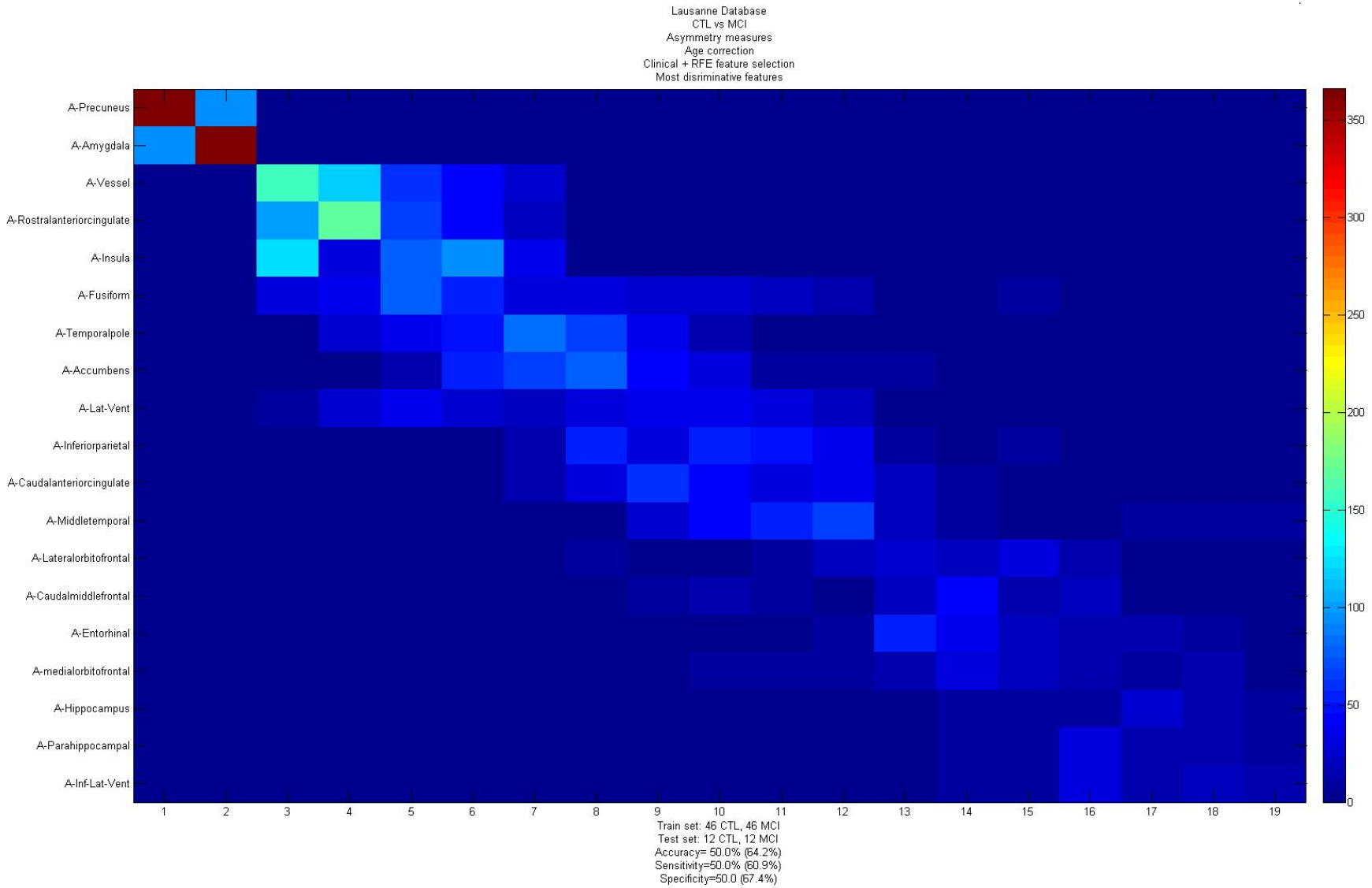


Figura B.35: Expanded Database CTL vs MCI CV most discriminative Features: Asymmetry coefficient.

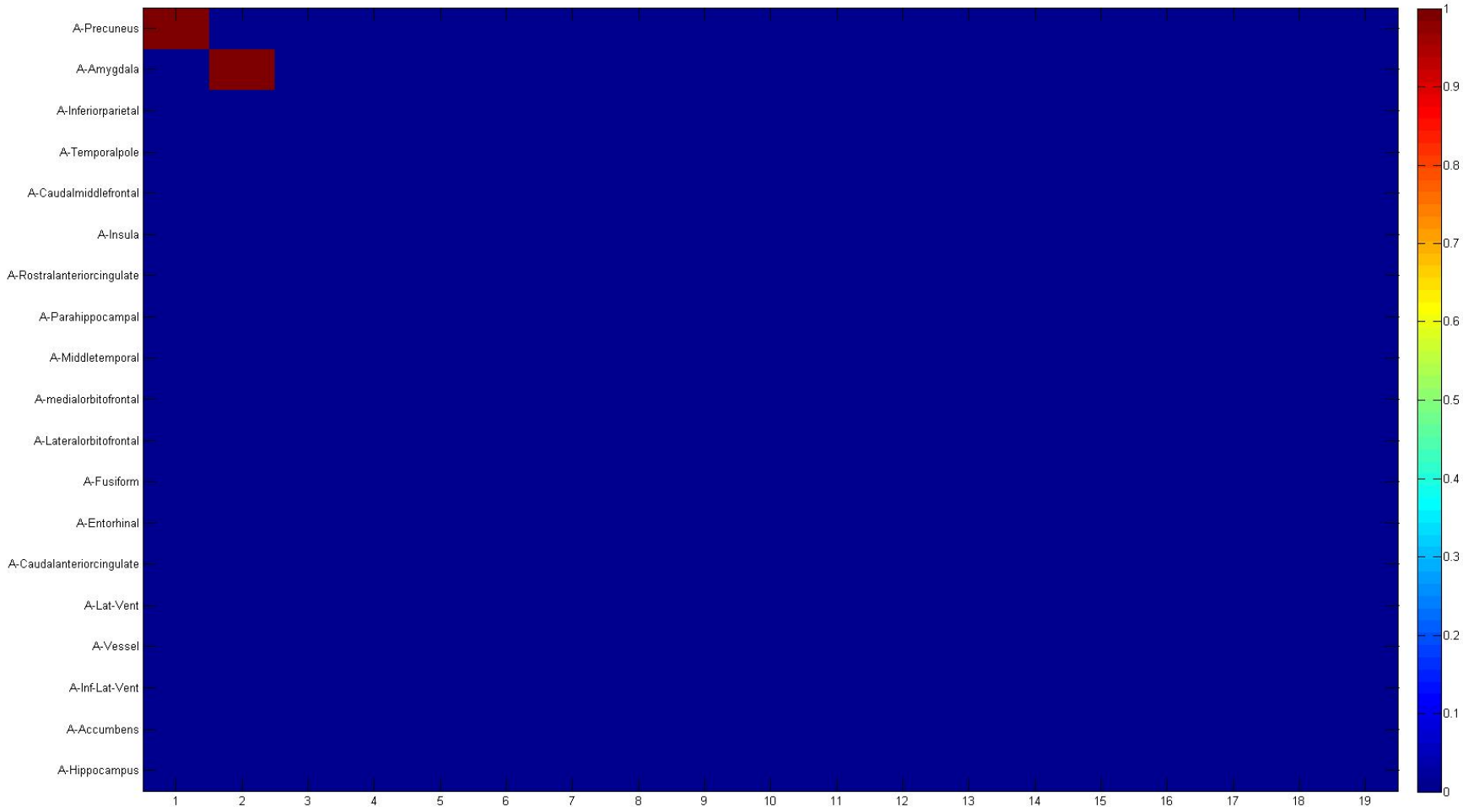


Figura B.36: Expanded Database CTL vs MCI pure most discriminative Features: Asymmetry coefficient.

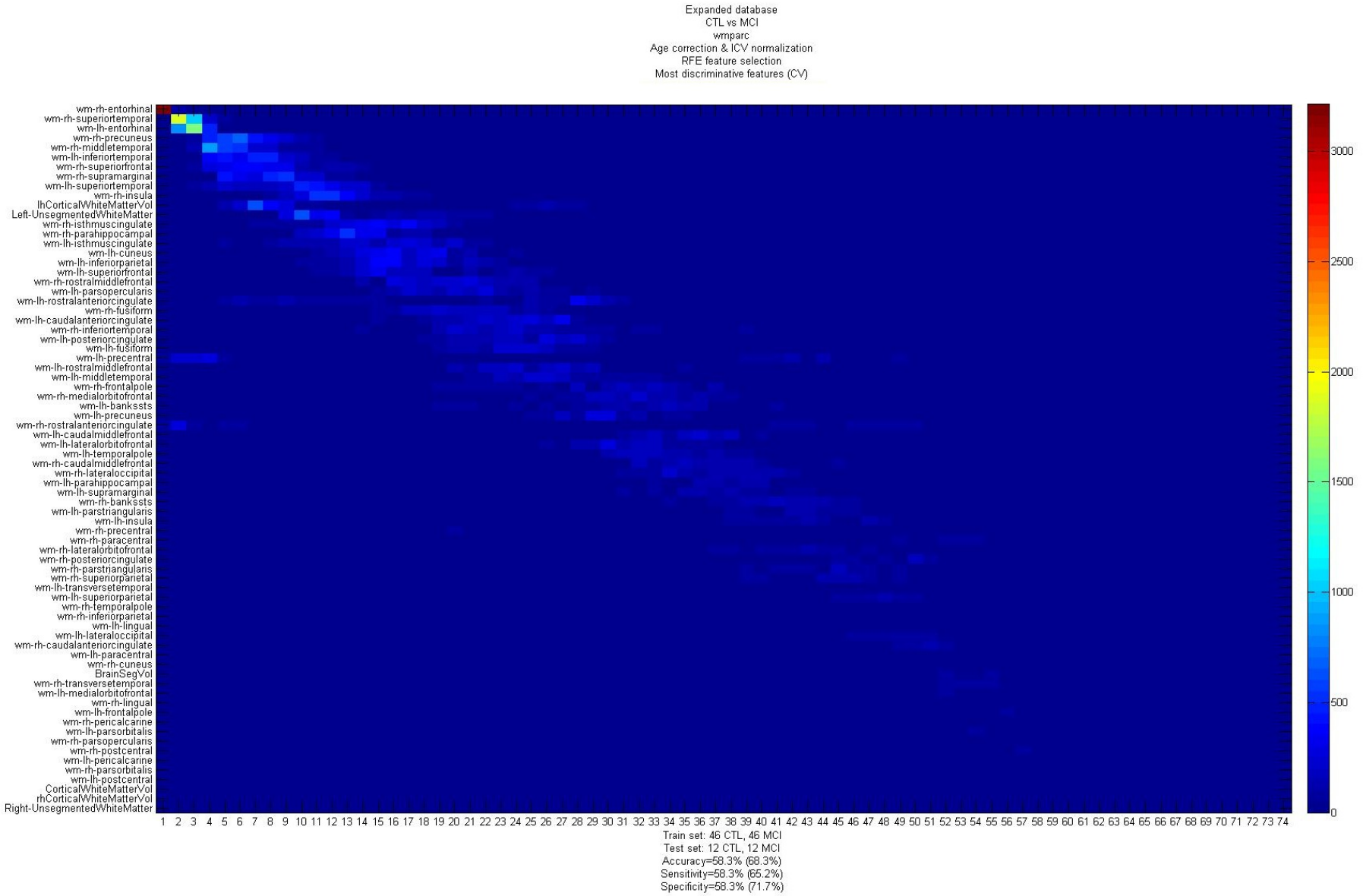


Figura B.37: Expanded Database CTL vs MCI CV most discriminative Features: WM, ICV normalization.

Figura B.38: Expanded Database CTL vs MCI pure most discriminative Features: WM, ICV normalization.

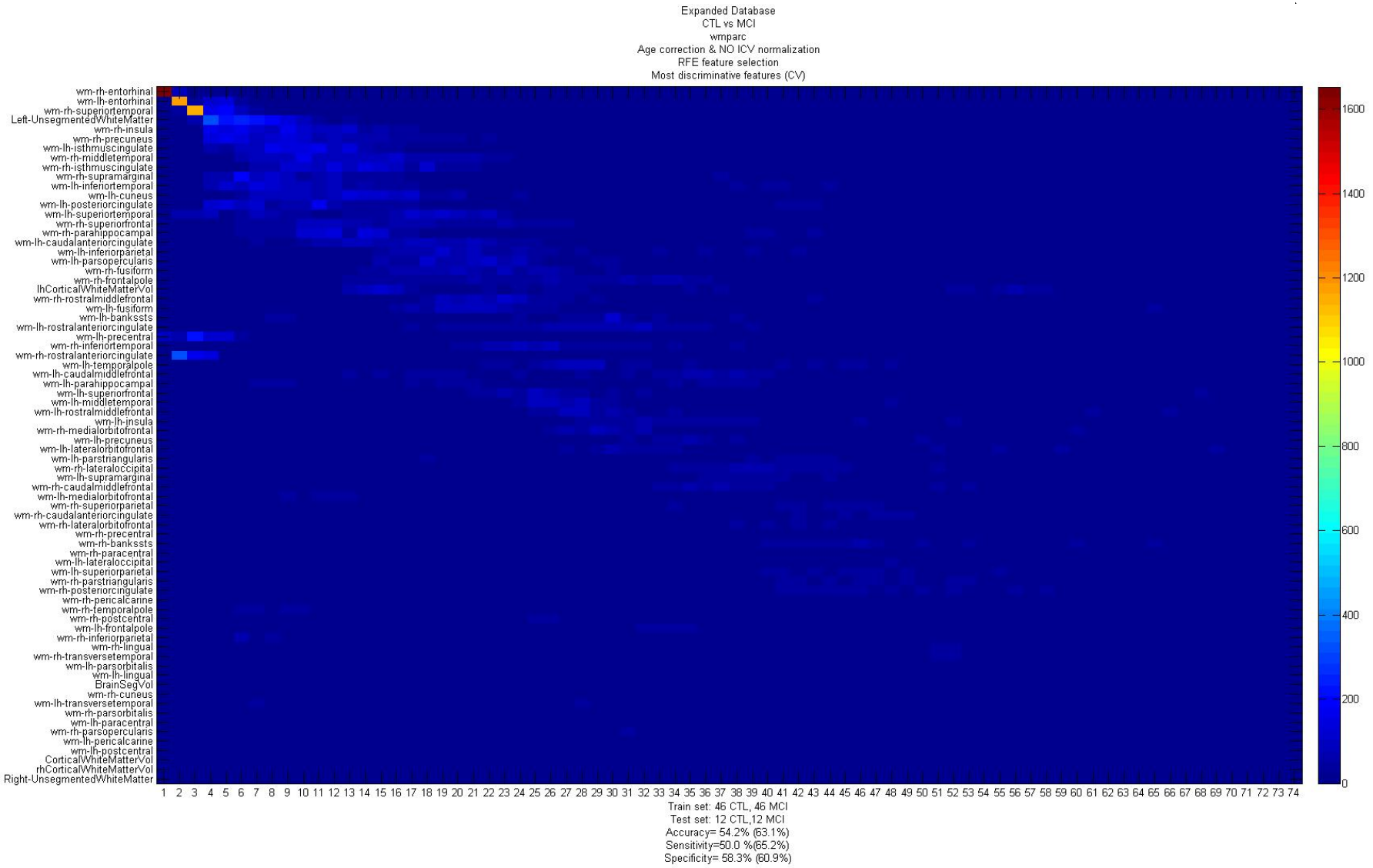


Figura B.39: Expanded Database CTL vs MCI CV most discriminative Features: WM, no ICV normalization.

Figura B.40: Expanded Database CTL vs MCI pure most discriminative Features: WM, no ICV normalization.

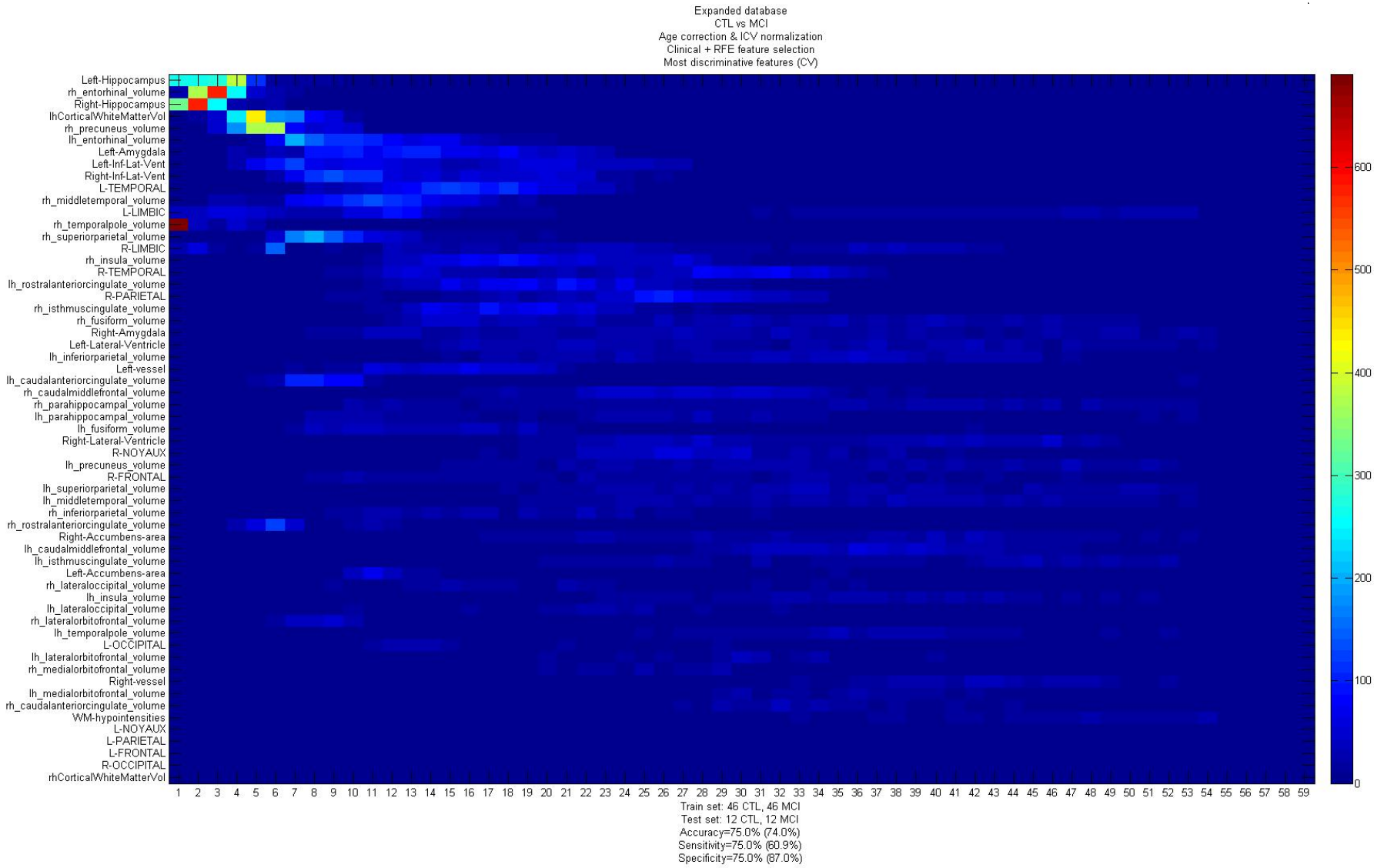


Figura B.41: Expanded Database CTL vs MCI CV most discriminative Features: GM regional volumes+ Lobe, ICV normalization.

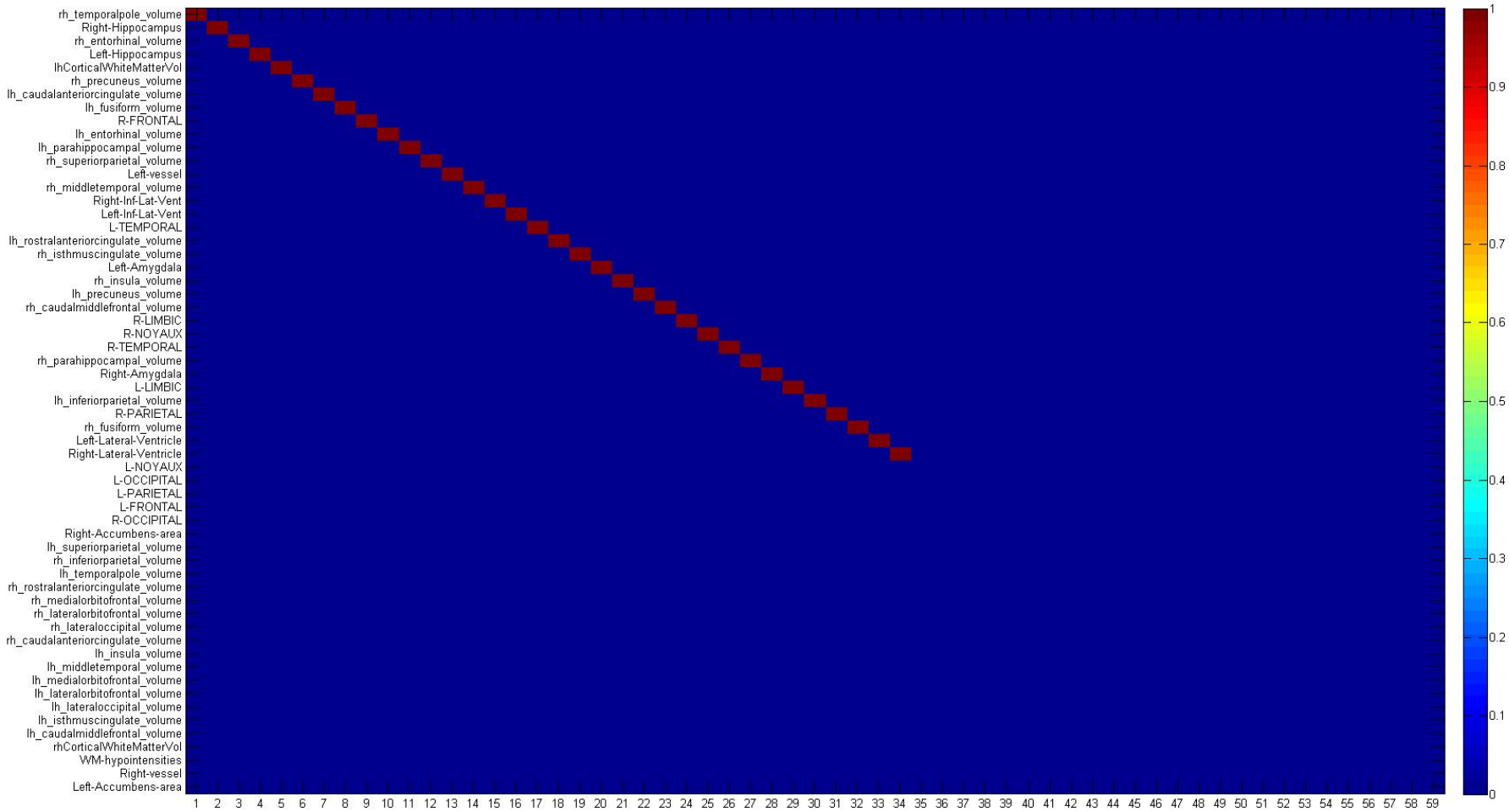


Figura B.42: Expanded Database CTL vs MCI pure most discriminative Features: GM regional volumes+ Lobe, ICV normalization.

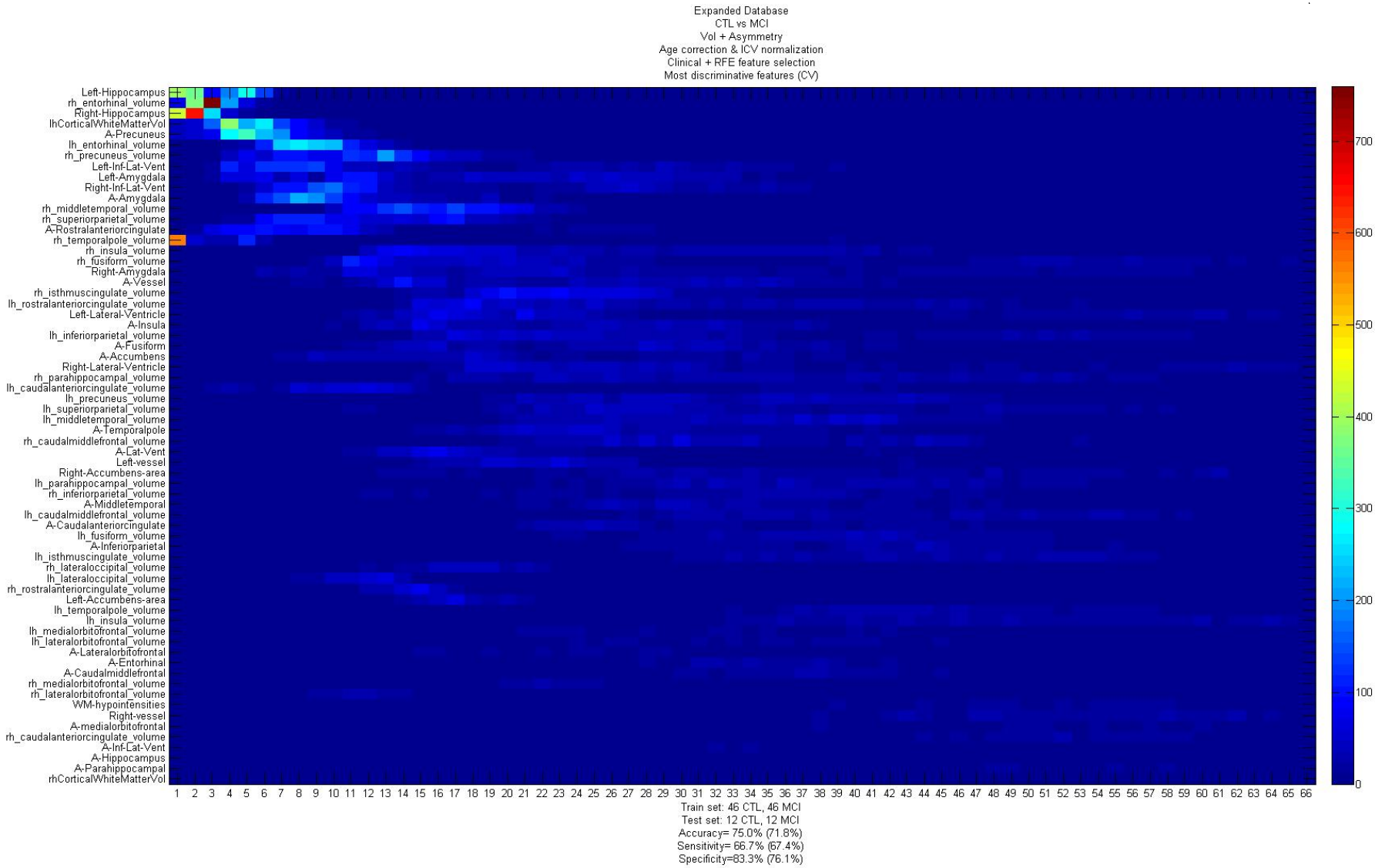


Figura B.43: Expanded Database CTL vs MCI CV most discriminative Features: GM regional volumes+ ASY, ICV normalization.

Figura B.44: Expanded Database CTL vs MCI pure most discriminative Features: GM regional volumes+ ASY, ICV normalization.

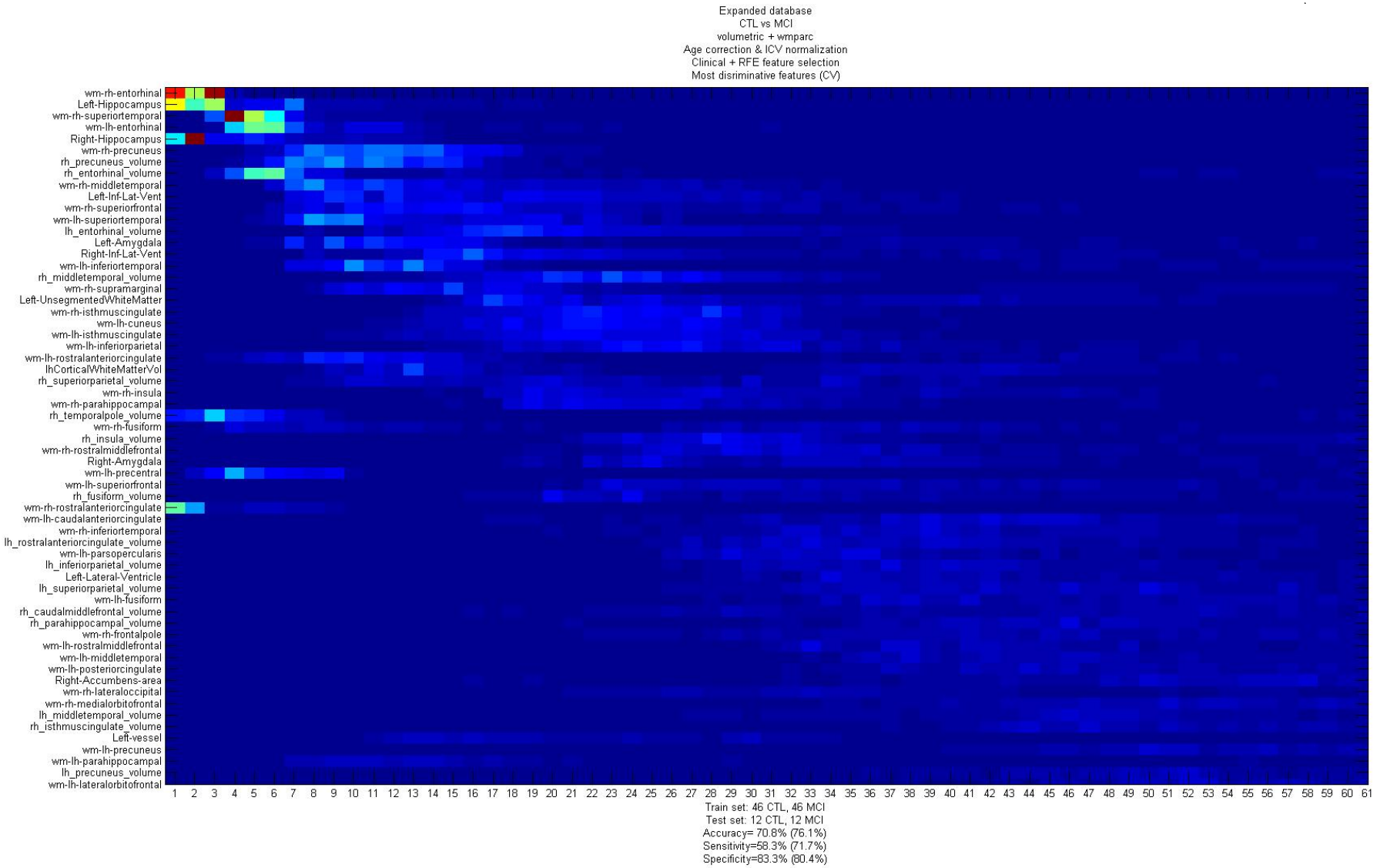


Figura B.45: Expanded Database CTL vs MCI CV most discriminative Features: GM regional volumes+ WM, ICV normalization.

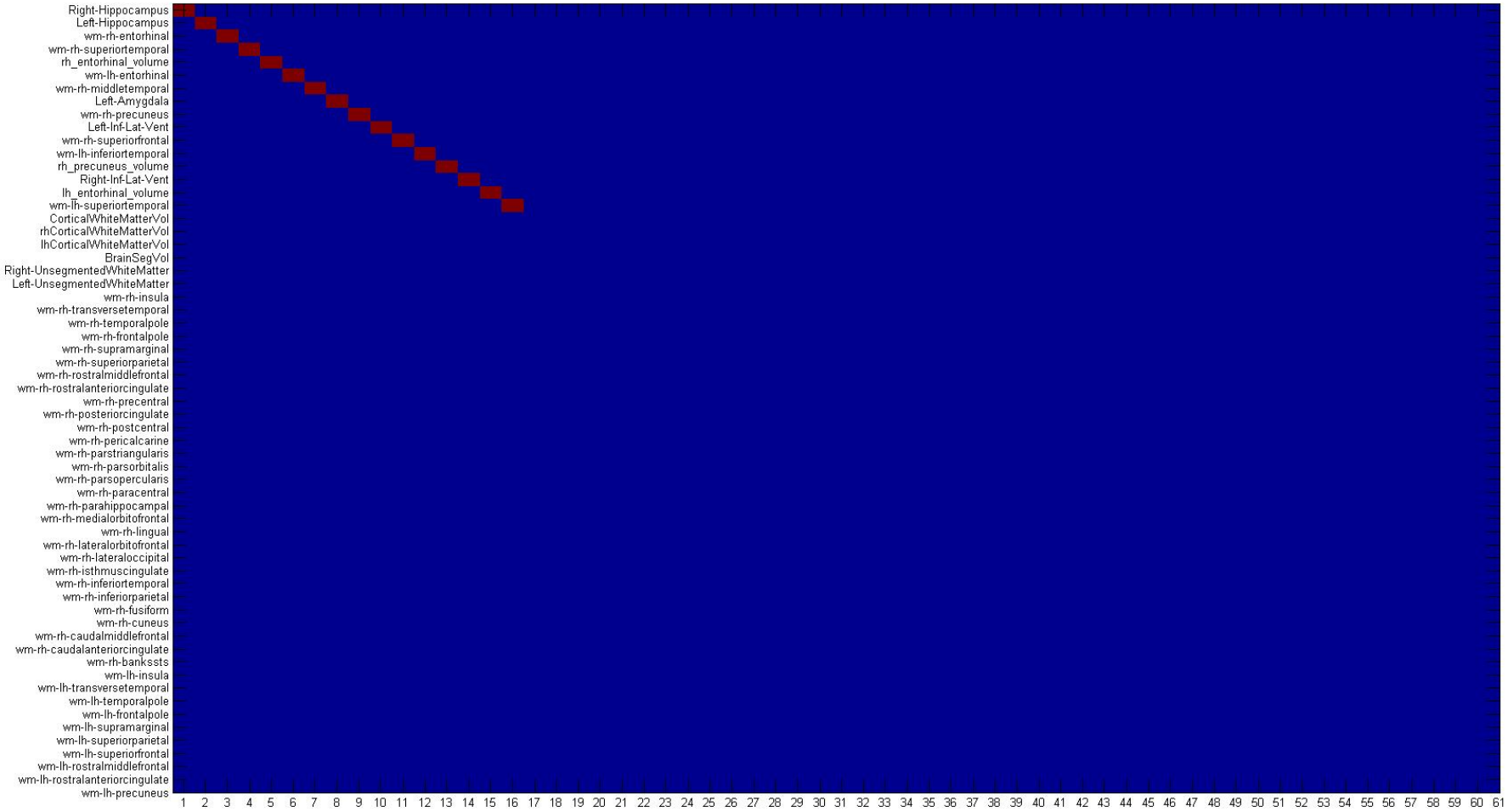


Figura B.46: Expanded Database CTL vs MCI pure most discriminative Features: GM regional volumes+ WM, ICV normalization.

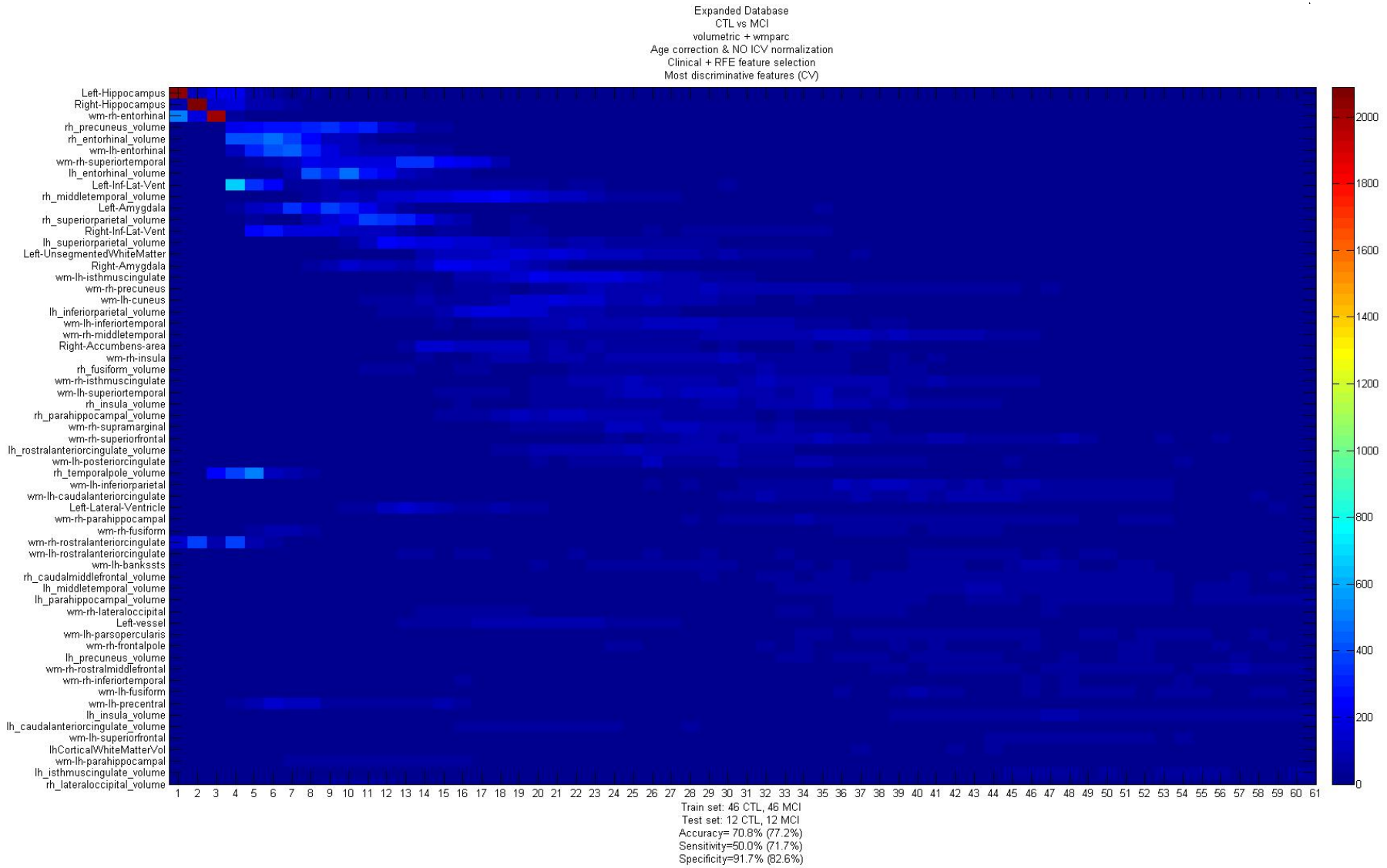


Figura B.47: Expanded Database CTL vs MCI CV most discriminative Features: GM regional volumes+ WM, no ICV normalization.

Figura B.48: Expanded Database CTL vs MCI pure most discriminative Features: GM regional volumes+ WM, ICV normalization.

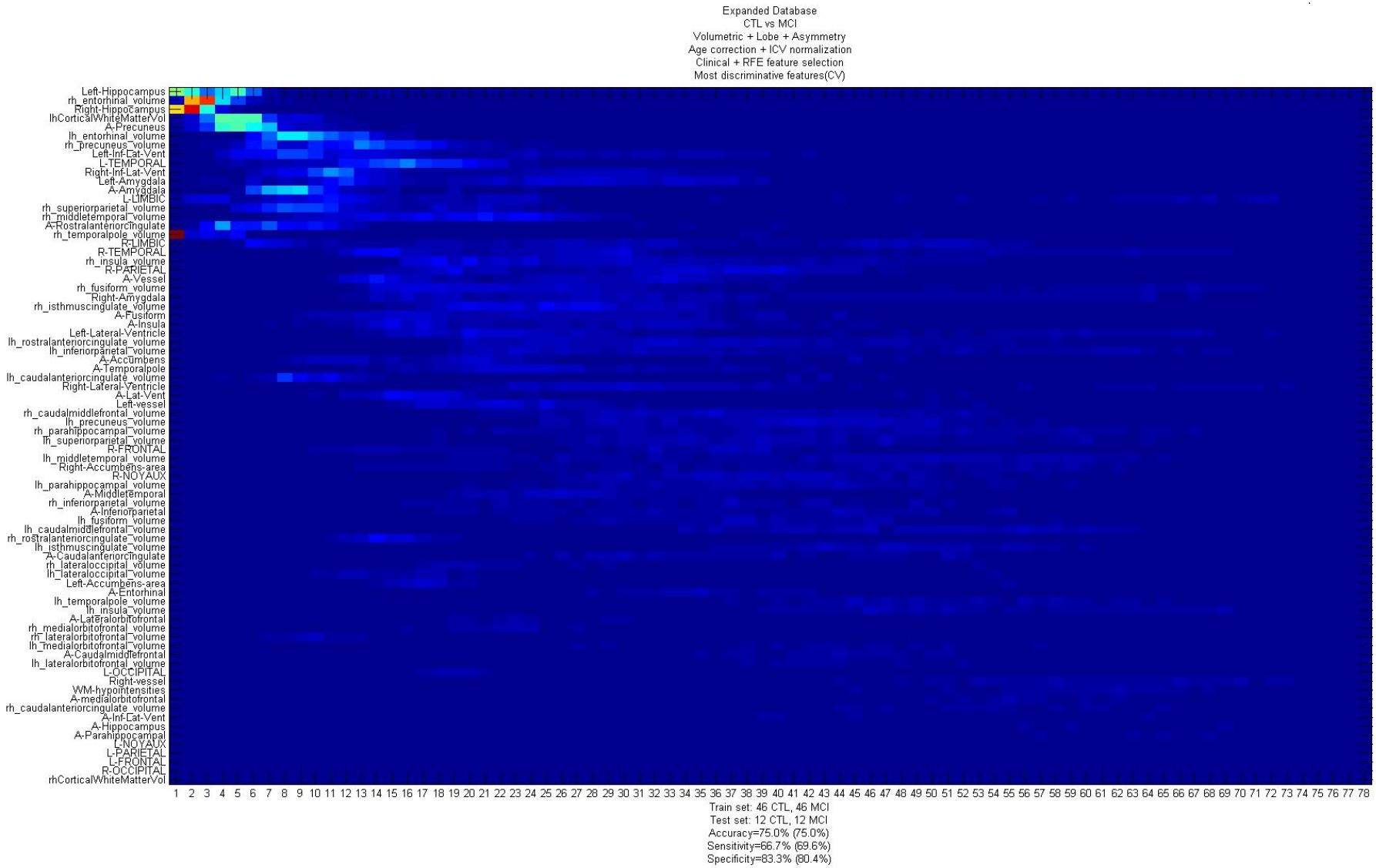


Figura B.49: Expanded Database CTL vs MCI CV most discriminative Features: GM regional volumes+ ASY+ Lobe, ICV normalization.

Figura B.50: Expanded Database CTL vs MCI pure most discriminative Features: GM regional volumes+ ASY+ Lobe, ICV normalization.

B.4. Complete Results for the Lausanne Database Multivariate & Multiscale Analysis

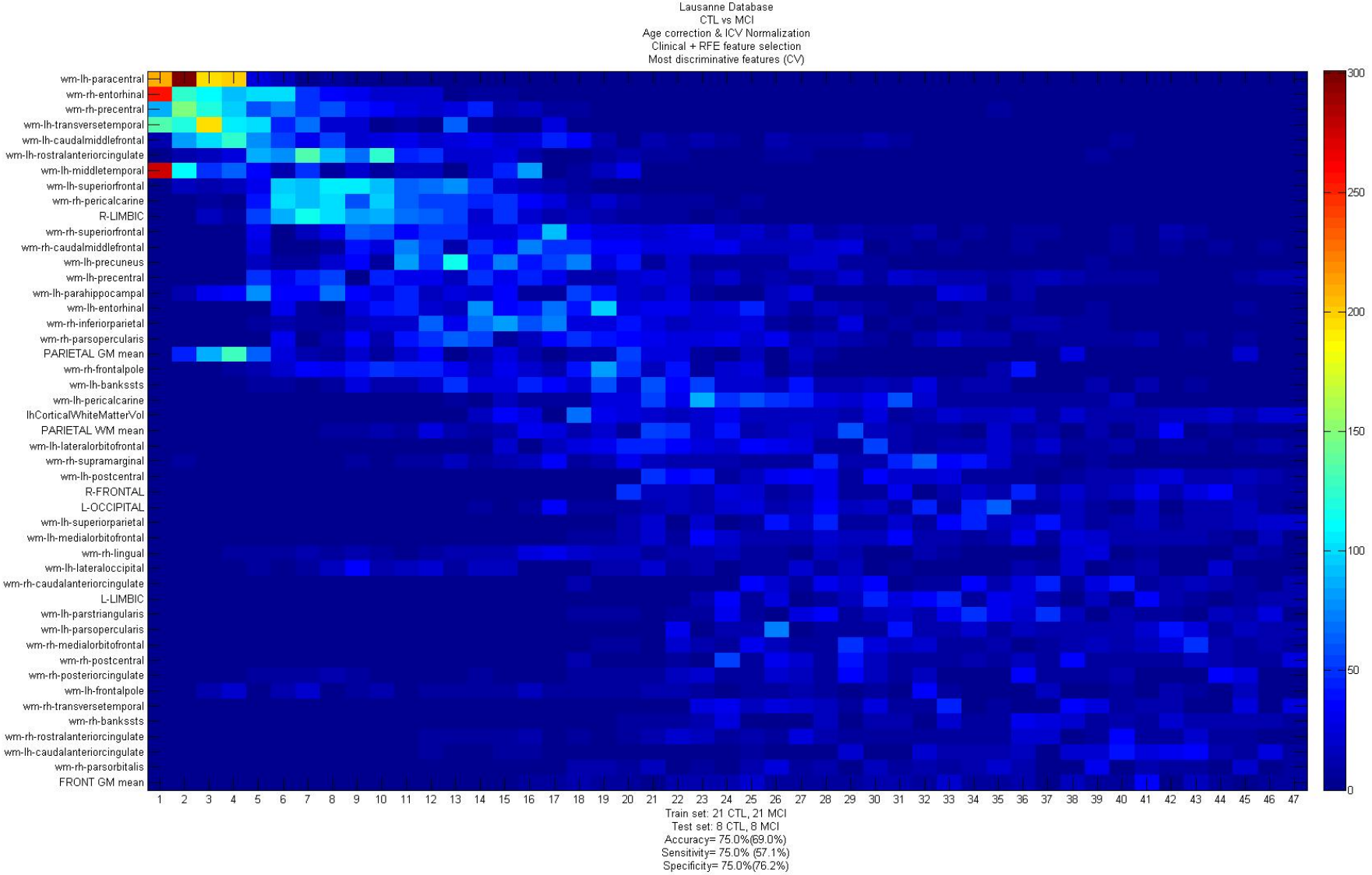


Figura B.51: Lausanne Database CTL vs MCI CV most discriminative Features: WM+ Lobe+ T_1-Lobe, ICV normalization.

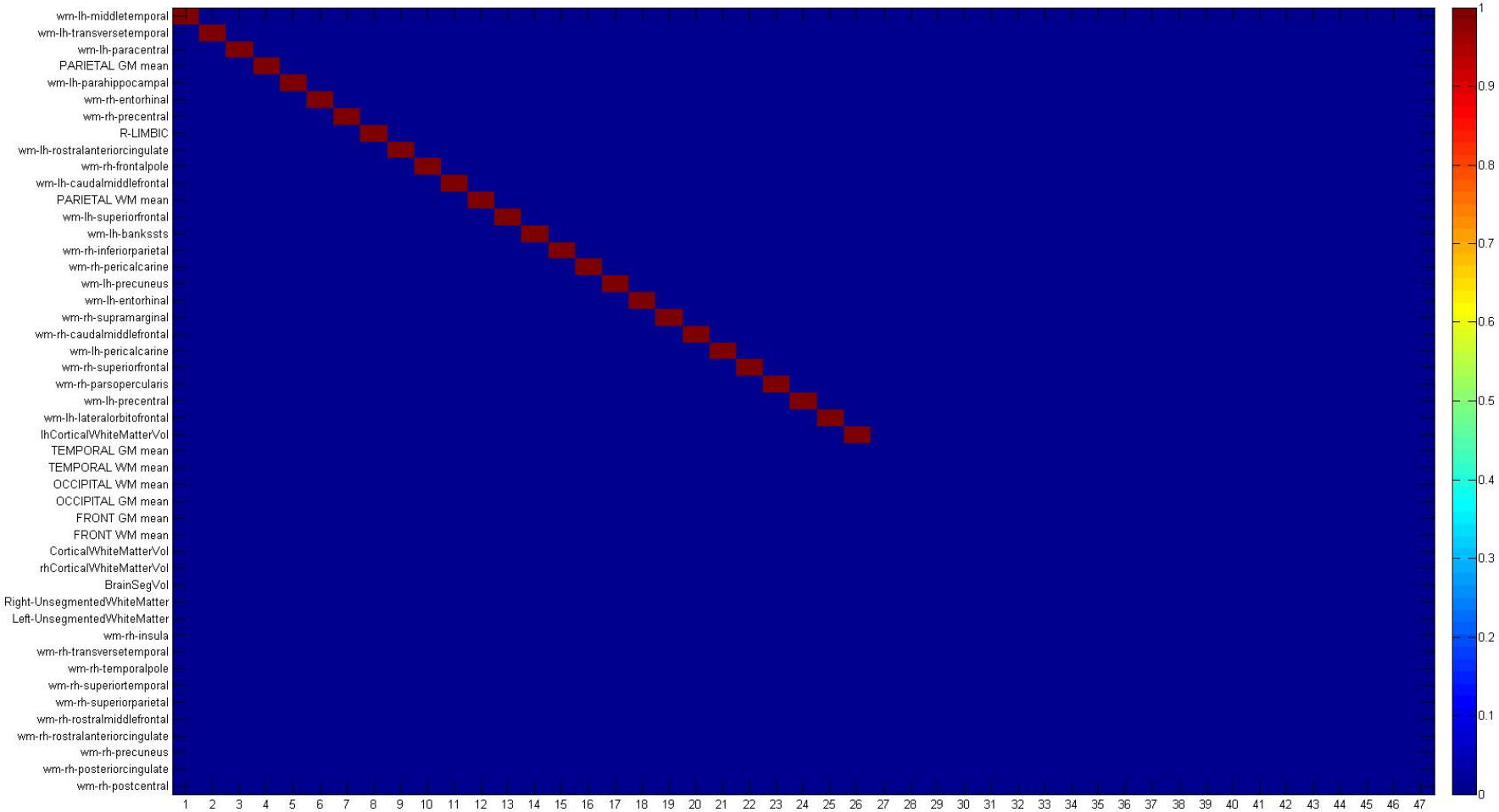


Figura B.52: Lausanne Database CTL vs MCI pure most discriminative Features: WM+ Lobe+ T_1-Lobe, ICV normalization.

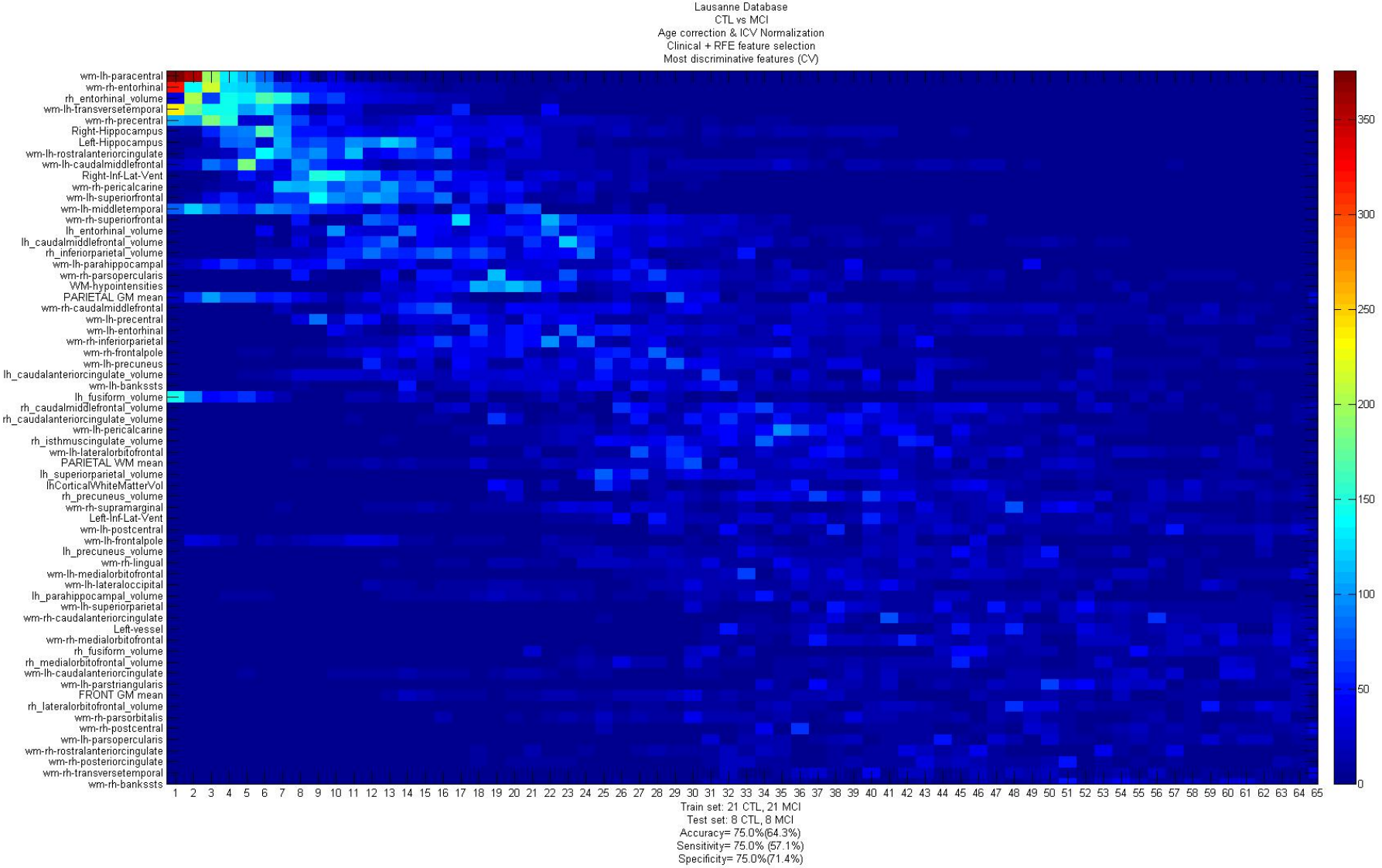


Figura B.53: Lausanne Database CTL vs MCI CV most discriminative Features: WM+ GM regional volumes+ T_1-Lobe, ICV normalization.

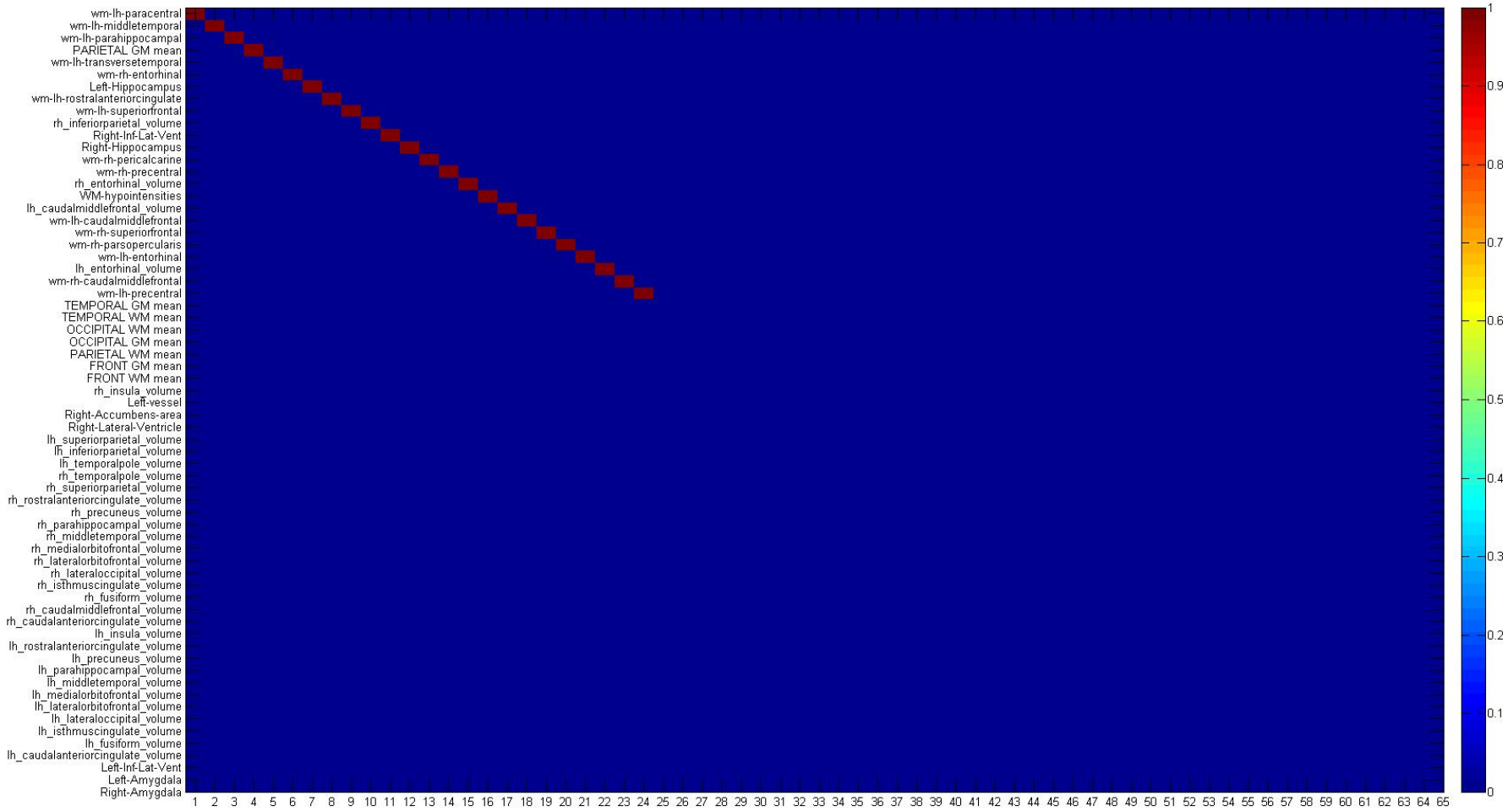


Figura B.54: Lausanne Database CTL vs MCI pure most discriminative Features: WM+ GM regional volumes+ T_1-Lobe, ICV normalization.

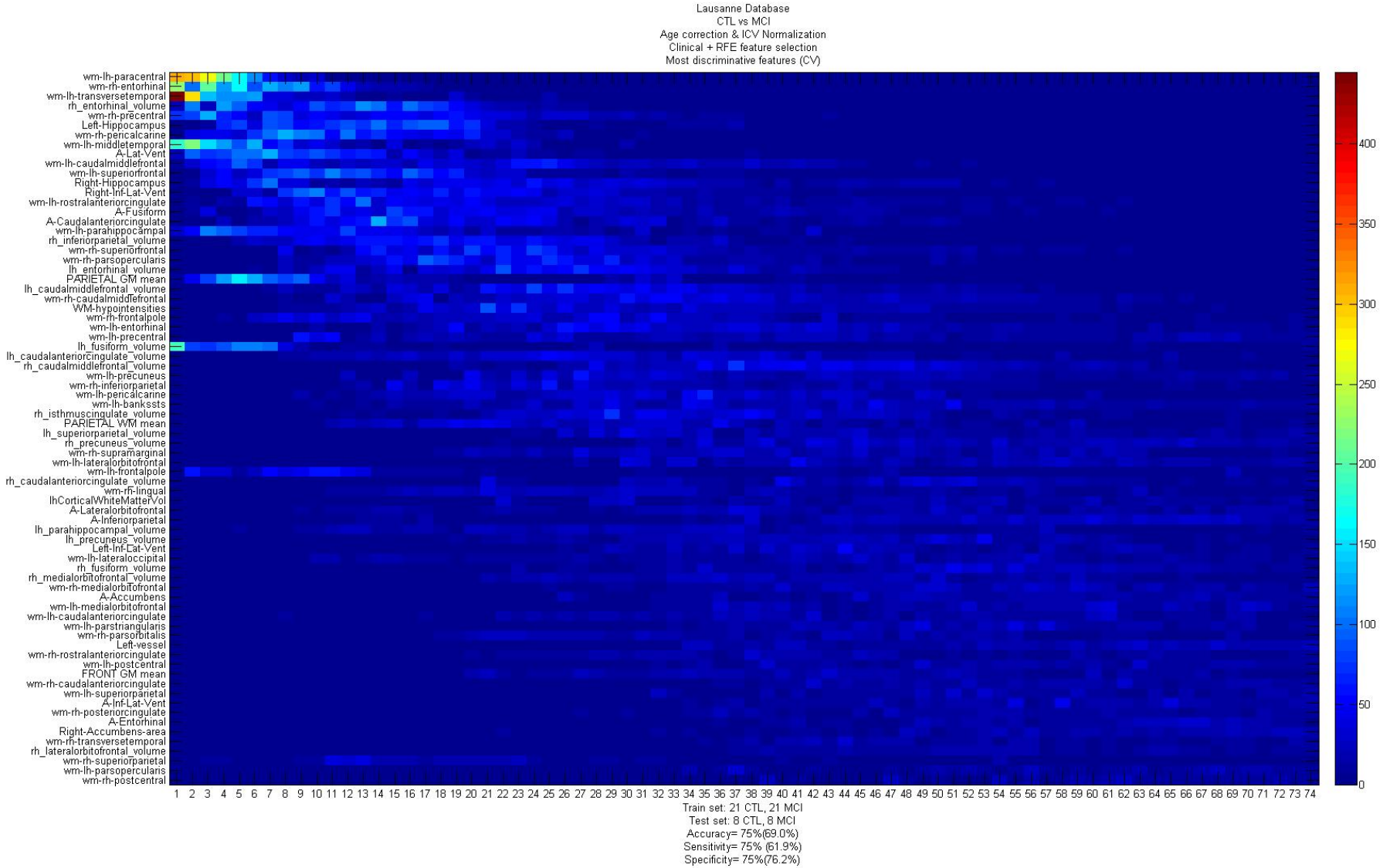


Figura B.55: Lausanne Database CTL vs MCI CV most discriminative Features: WM+ ASY+ GM regional volumes+ T_1-Lobe, ICV normalization.

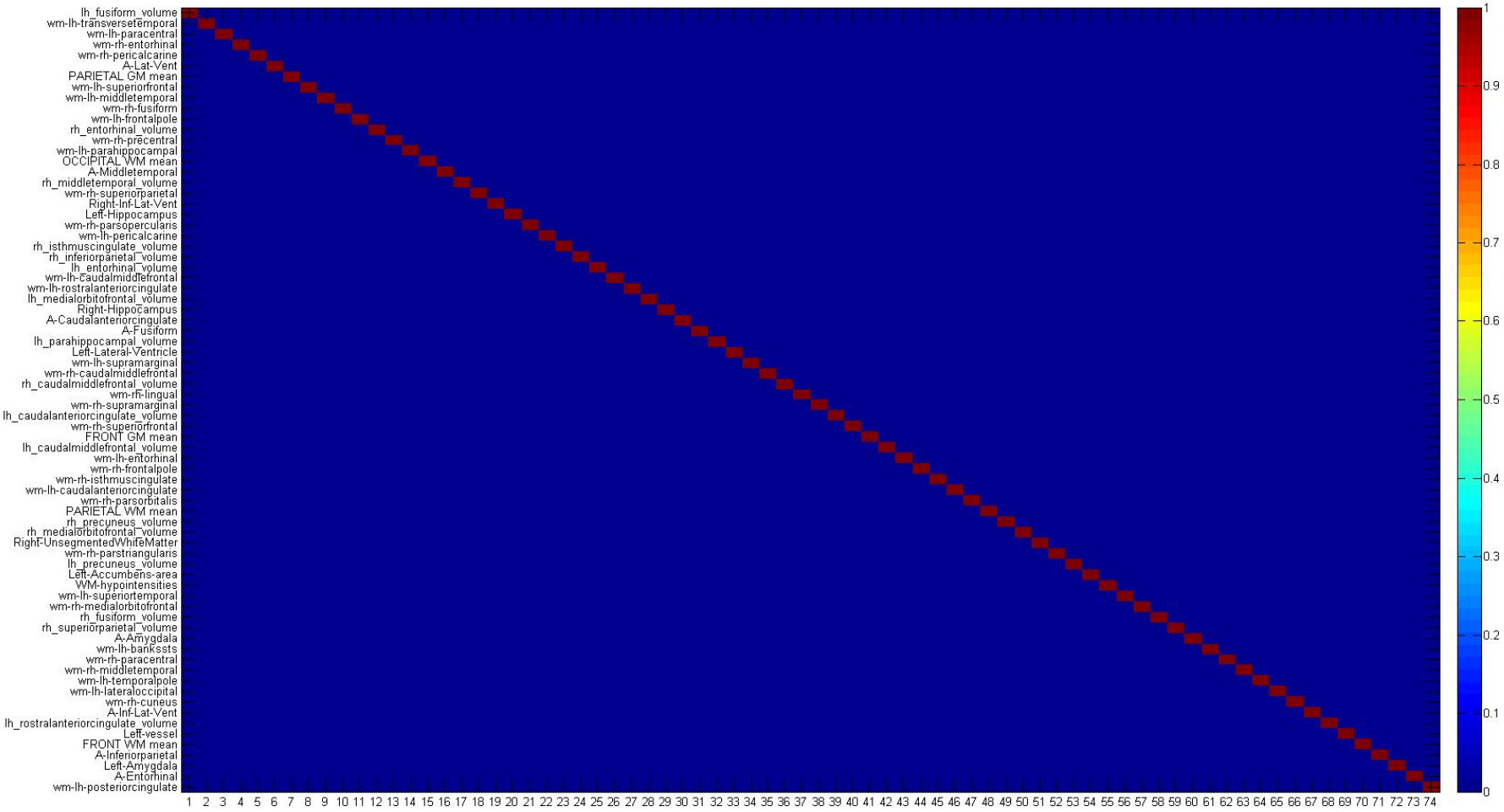


Figura B.56: Lausanne Database CTL vs MCI pure most discriminative Features: WM+ ASY+ GM regional volumes+ T_1 -Lobe, ICV normalization.

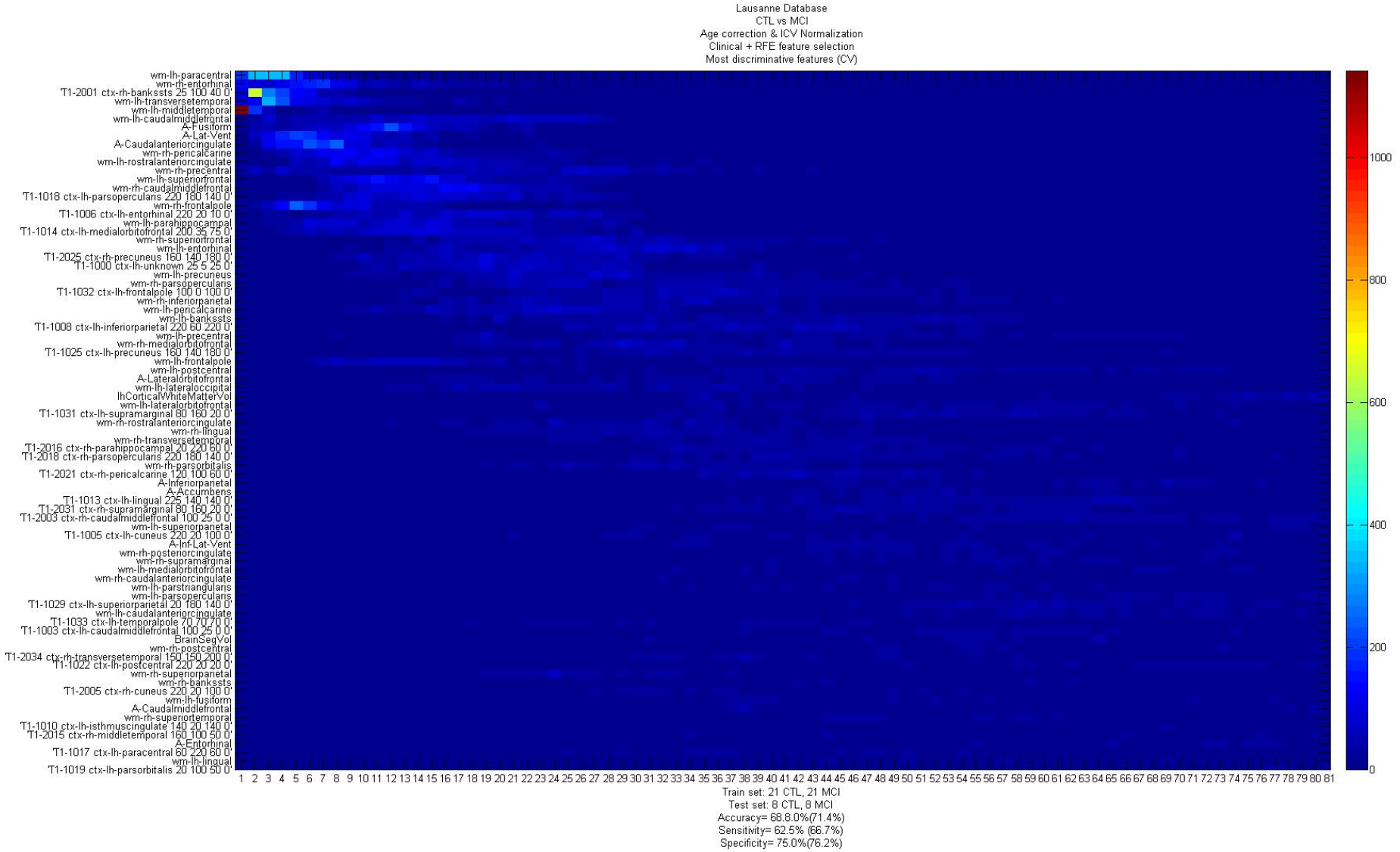


Figura B.57: Lausanne Database CTL vs MCI CV most discriminative Features: WM+ ASY+ T_1, ICV normalization.

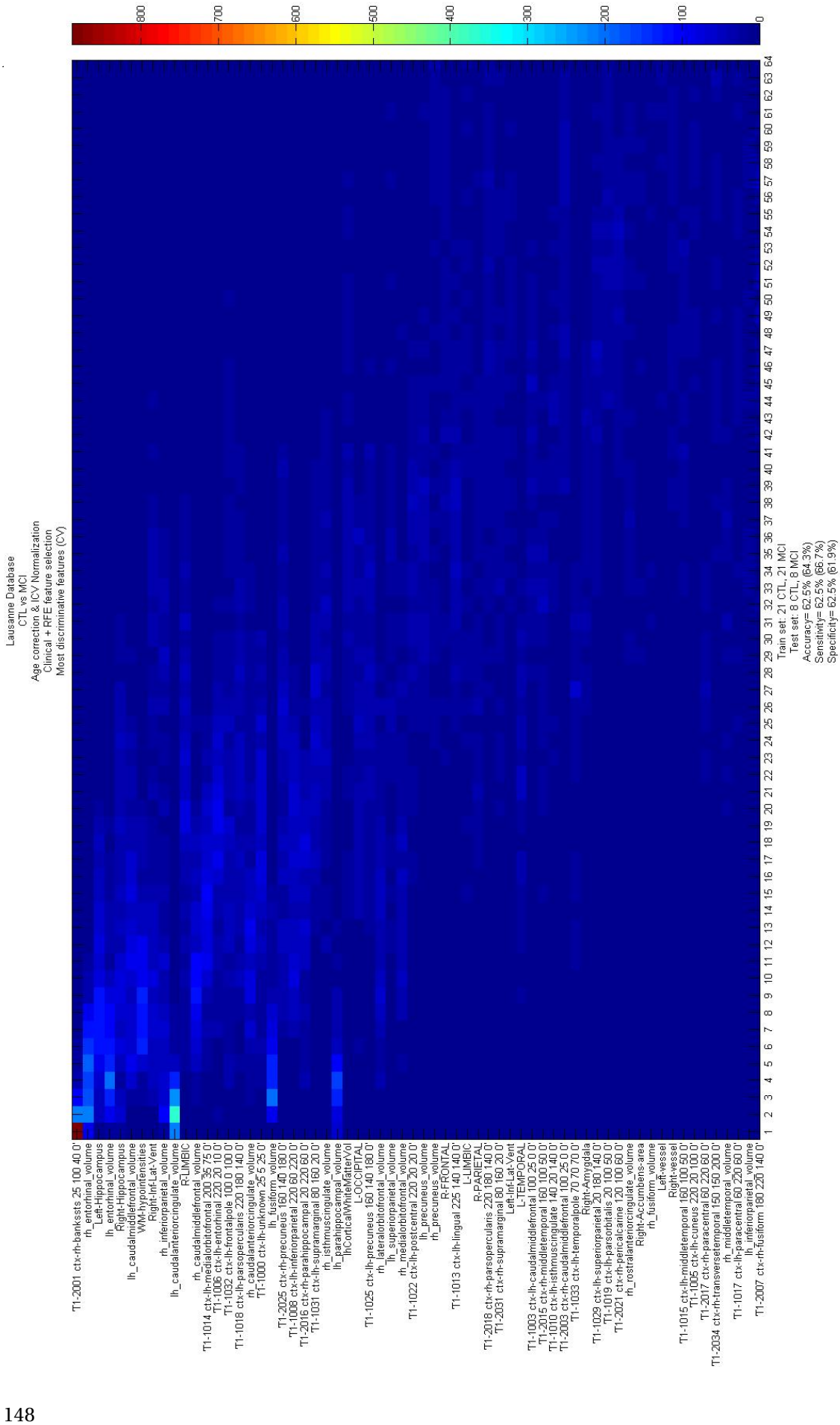


Figura B.58: Lausanne Database CTL vs MCI pure most discriminative Features: GM regional volumes+ Lobe+ T_1, ICV normalization.

C CV Accuracy- Confidence Interval Correspondance

In this Appendix, the CV Accuracy- Confidence Interval correspondance is given for the three performed analysis.

C.1. ADNI 1 Database

Cuadro C.1: ADNI 1 CV Accuracy- Confidance Intervals Correspondance.

CV Accuracy	Confidance Interval [Lower Bound- Upper Bound]
0 %	[0- 1.2240]
0.32258 %	[0.056966-1.8043]
0.64516 %	[0.17711-2.3214]
0.96774 %	[0.32965-2.8062]
1.2903 %	[0.50289-3.2702]
1.6129 %	[0.69085-3.7195]
1.9355 %	[0.88999-4.1576]
2.2581 %	[1.098-4.5868]
2.5806 %	[1.3133-5.0088]
2.9032 %	[1.5348-5.4246]
3.2258 %	[1.7615-5.8352]
3.5484 %	[1.9927-6.2412]
3.871 %	[2.228-6.6432]
4.1935 %	[2.4669-7.0416]
4.5161 %	[2.7089-7.4368]
4.8387 %	[2.9539-7.829]
5.1613 %	[3.2016-8.2186]
5.4839 %	[3.4517-8.6058]
5.8065 %	[3.7041-8.9907]
6.129 %	[3.9585-9.3735]
6.4516 %	[4.2149-9.7544]

6.7742 %	[4.4731-10.1335]
7.0968 %	[4.733-10.5108]
7.4194 %	[4.9945-10.8866]
7.7419 %	[5.2575-11.2609]
8.0645 %	[5.5219-11.6337]
8.3871 %	[5.7877-12.0052]
8.7097 %	[6.0547-12.3754]
9.0323 %	[6.323-12.7444]
9.3548 %	[6.5925-13.1122]
9.6774 %	[6.863-13.4789]
10 %	[7.1346-13.8446]
10.3226 %	[7.4073-14.2092]
10.6452 %	[7.6809-14.5729]
10.9677 %	[7.9554-14.9356]
11.2903 %	[8.2309-15.2974]
11.6129 %	[8.5072-15.6583]
11.9355 %	[8.7844-16.0184]
12.2581 %	[9.0624-16.3777]
12.5806 %	[9.3412-16.7361]
12.9032 %	[9.6207-17.0939]
13.2258 %	[9.901-17.4509]
13.5484 %	[10.182-17.8071]
13.871 %	[10.4637-18.1627]
14.1935 %	[10.746-18.5176]
14.5161 %	[11.0291-18.8719]
14.8387 %	[11.3127-19.2255]
15.1613 %	[11.597-19.5784]
15.4839 %	[11.8819-19.9308]
15.8065 %	[12.1674-20.2826]
16.129 %	[12.4534-20.6338]
16.4516 %	[12.74-20.9845]
16.7742 %	[13.0272-21.3346]
17.0968 %	[13.3149-21.6841]
17.4194 %	[13.6031-22.0332]
17.7419 %	[13.8919-22.3817]
18.0645 %	[14.1811-22.7297]
18.3871 %	[14.4708-23.0772]
18.7097 %	[14.7611-23.4243]
19.0323 %	[15.0518-23.7709]
19.3548 %	[15.3429-24.117]
19.6774 %	[15.6345-24.4626]
20 %	[15.9266-24.8078]

20.3226 %	[16.2191-25.1526]
20.6452 %	[16.512-25.4969]
20.9677 %	[16.8054-25.8408]
21.2903 %	[17.0992-26.1843]
21.6129 %	[17.3934-26.5274]
21.9355 %	[17.6879-26.8701]
22.2581 %	[17.9829-27.2123]
22.5806 %	[18.2783-27.5542]
22.9032 %	[18.5741-27.8957]
23.2258 %	[18.8702-28.2368]
23.5484 %	[19.1667-28.5776]
23.871 %	[19.4636-28.9179]
24.1935 %	[19.7609-29.2579]
24.5161 %	[20.0585-29.5976]
24.8387 %	[20.3565-29.9369]
25.1613 %	[20.6548-30.2758]
25.4839 %	[20.9535-30.6144]
25.8065 %	[21.2525-30.9527]
26.129 %	[21.5518-31.2906]
26.4516 %	[21.8515-31.6282]
26.7742 %	[22.1515-31.9654]
27.0968 %	[22.4518-32.3024]
27.4194 %	[22.7525-32.639]
27.7419 %	[23.0535-32.9753]
28.0645 %	[23.3548-33.3112]
28.3871 %	[23.6564-33.6469]
28.7097 %	[23.9583-33.9823]
29.0323 %	[24.2605-34.3173]
29.3548 %	[24.563-34.6521]
29.6774 %	[24.8658-34.9865]
30 %	[25.169-35.3206]
30.3226 %	[25.4724-35.6545]
30.6452 %	[25.7761-35.9881]
30.9677 %	[26.0801-36.3213]
31.2903 %	[26.3843-36.6543]
31.6129 %	[26.6889-36.987]
31.9355 %	[26.9938-37.3194]
32.2581 %	[27.2989-37.6516]
32.5806 %	[27.6043-37.9834]
32.9032 %	[27.91-38.315]
33.2258 %	[28.2159-38.6463]
33.5484 %	[28.5221-38.9774]

33.871 %	[28.8286-39.3081]
34.1935 %	[29.1354-39.6387]
34.5161 %	[29.4424-39.9689]
34.8387 %	[29.7497-40.2989]
35.1613 %	[30.0573-40.6286]
35.4839 %	[30.3651-40.958]
35.8065 %	[30.6731-41.2872]
36.129 %	[30.9815-41.6162]
36.4516 %	[31.2901-41.9448]
36.7742 %	[31.5989-42.2733]
37.0968 %	[31.908-42.6014]
37.4194 %	[32.2173-42.9293]
37.7419 %	[32.5269-43.257]
38.0645 %	[32.8368-43.5844]
38.3871 %	[33.1469-43.9116]
38.7097 %	[33.4572-44.2385]
39.0323 %	[33.7678-44.5652]
39.3548 %	[34.0786-44.8916]
39.6774 %	[34.3897-45.2178]
40 %	[34.701-45.5438]
40.3226 %	[35.0126-45.8695]
40.6452 %	[35.3244-46.1949]
40.9677 %	[35.6364-46.5202]
41.2903 %	[35.9487-46.8452]
41.6129 %	[36.2612-47.1699]
41.9355 %	[36.574-47.4944]
42.2581 %	[36.887-47.8187]
42.5806 %	[37.2002-48.1427]
42.9032 %	[37.5137-48.4665]
43.2258 %	[37.8273-48.7901]
43.5484 %	[38.1413-49.1134]
43.871 %	[38.4554-49.4365]
44.1935 %	[38.7698-49.7594]
44.5161 %	[39.0845-50.082]
44.8387 %	[39.3993-50.4045]
45.1613 %	[39.7144-50.7266]
45.4839 %	[40.0297-51.0486]
45.8065 %	[40.3453-51.3703]
46.129 %	[40.6611-51.6918]
46.4516 %	[40.9771-52.013]
46.7742 %	[41.2933-52.334]
47.0968 %	[41.6098-52.6548]

47.4194 %	[41.9265-52.9754]
47.7419 %	[42.2434-53.2958]
48.0645 %	[42.5605-53.6159]
48.3871 %	[42.8779-53.9358]
48.7097 %	[43.1955-54.2554]
49.0323 %	[43.5134-54.5748]
49.3548 %	[43.8314-54.894]
49.6774 %	[44.1497-55.213]
50 %	[44.4682-55.5318]
50.3226 %	[44.787-55.8503]
50.6452 %	[45.106-56.1686]
50.9677 %	[45.4252-56.4866]
51.2903 %	[45.7446-56.8045]
51.6129 %	[46.0642-57.1221]
51.9355 %	[46.3841-57.4395]
52.2581 %	[46.7042-57.7566]
52.5806 %	[47.0246-58.0735]
52.9032 %	[47.3452-58.3902]
53.2258 %	[47.666-58.7067]
53.5484 %	[47.987-59.0229]
53.871 %	[48.3082-59.3389]
54.1935 %	[48.6297-59.6547]
54.5161 %	[48.9514-59.9703]
54.8387 %	[49.2734-60.2856]
55.1613 %	[49.5955-60.6007]
55.4839 %	[49.918-60.9155]
55.8065 %	[50.2406-61.2302]
56.129 %	[50.5635-61.5446]
56.4516 %	[50.8866-61.8587]
56.7742 %	[51.2099-62.1727]
57.0968 %	[51.5335-62.4863]
57.4194 %	[51.8573-62.7998]
57.7419 %	[52.1813-63.113]
58.0645 %	[52.5056-63.426]
58.3871 %	[52.8301-63.7388]
58.7097 %	[53.1548-64.0513]
59.0323 %	[53.4798-64.3636]
59.3548 %	[53.8051-64.6756]
59.6774 %	[54.1305-64.9874]
60 %	[54.4562-65.299]
60.3226 %	[54.7822-65.6103]
60.6452 %	[55.1084-65.9214]

60.9677 %	[55.4348-66.2322]
61.2903 %	[55.7615-66.5428]
61.6129 %	[56.0884-66.8531]
61.9355 %	[56.4156-67.1632]
62.2581 %	[56.743-67.4731]
62.5806 %	[57.0707-67.7827]
62.9032 %	[57.3986-68.092]
63.2258 %	[57.7267-68.4011]
63.5484 %	[58.0552-68.7099]
63.871 %	[58.3838-69.0185]
64.1935 %	[58.7128-69.3269]
64.5161 %	[59.042-69.6349]
64.8387 %	[59.3714-69.9427]
65.1613 %	[59.7011-70.2503]
65.4839 %	[60.0311-70.5576]
65.8065 %	[60.3613-70.8646]
66.129 %	[60.6919-71.1714]
66.4516 %	[61.0226-71.4779]
66.7742 %	[61.3537-71.7841]
67.0968 %	[61.685-72.09]
67.4194 %	[62.0166-72.3957]
67.7419 %	[62.3484-72.7011]
68.0645 %	[62.6806-73.0062]
68.3871 %	[63.013-73.3111]
68.7097 %	[63.3457-73.6157]
69.0323 %	[63.6787-73.9199]
69.3548 %	[64.0119-74.2239]
69.6774 %	[64.3455-74.5276]
70 %	[64.6794-74.831]
70.3226 %	[65.0135-75.1342]
70.6452 %	[65.3479-75.437]
70.9677 %	[65.6827-75.7395]
71.2903 %	[66.0177-76.0417]
71.6129 %	[66.3531-76.3436]
71.9355 %	[66.6888-76.6452]
72.2581 %	[67.0247-76.9465]
72.5806 %	[67.361-77.2475]
72.9032 %	[67.6976-77.5482]
73.2258 %	[68.0346-77.8485]
73.5484 %	[68.3718-78.1485]
73.871 %	[68.7094-78.4482]
74.1935 %	[69.0473-78.7475]

74.5161 %	[69.3856-79.0465]
74.8387 %	[69.7242-79.3452]
75.1613 %	[70.0631-79.6435]
75.4839 %	[70.4024-79.9415]
75.8065 %	[70.7421-80.2391]
76.129 %	[71.0821-80.5364]
76.4516 %	[71.4224-80.8333]
76.7742 %	[71.7632-81.1298]
77.0968 %	[72.1043-81.4259]
77.4194 %	[72.4458-81.7217]
77.7419 %	[72.7877-82.0171]
78.0645 %	[73.1299-82.3121]
78.3871 %	[73.4726-82.6066]
78.7097 %	[73.8157-82.9008]
79.0323 %	[74.1592-83.1946]
79.3548 %	[74.5031-83.488]
79.6774 %	[74.8474-83.7809]
80 %	[75.1922-84.0734]
80.3226 %	[75.5374-84.3655]
80.6452 %	[75.883-84.6571]
80.9677 %	[76.2291-84.9482]
81.2903 %	[76.5757-85.2389]
81.6129 %	[76.9228-85.5292]
81.9355 %	[77.2703-85.8189]
82.2581 %	[77.6183-86.1081]
82.5806 %	[77.9668-86.3969]
82.9032 %	[78.3159-86.6851]
83.2258 %	[78.6654-86.9728]
83.5484 %	[79.0155-87.26]
83.871 %	[79.3662-87.5466]
84.1935 %	[79.7174-87.8326]
84.5161 %	[80.0692-88.1181]
84.8387 %	[80.4216-88.403]
85.1613 %	[80.7745-88.6873]
85.4839 %	[81.1281-88.9709]
85.8065 %	[81.4824-89.254]
86.129 %	[81.8373-89.5363]
86.4516 %	[82.1929-89.818]
86.7742 %	[82.5491-90.099]
87.0968 %	[82.9061-90.3793]
87.4194 %	[83.2639-90.6588]
87.7419 %	[83.6223-90.9376]

88.0645 %	[83.9816-91.2156]
88.3871 %	[84.3417-91.4928]
88.7097 %	[84.7026-91.7691]
89.0323 %	[85.0644-92.0446]
89.3548 %	[85.4271-92.3191]
89.6774 %	[85.7908-92.5927]
90 %	[86.1554-92.8654]
90.3226 %	[86.5211-93.137]
90.6452 %	[86.8878-93.4075]
90.9677 %	[87.2556-93.677]
91.2903 %	[87.6246-93.9453]
91.6129 %	[87.9948-94.2123]
91.9355 %	[88.3663-94.4781]
92.2581 %	[88.7391-94.7425]
92.5806 %	[89.1134-95.0055]
92.9032 %	[89.4892-95.267]
93.2258 %	[89.8665-95.5269]
93.5484 %	[90.2456-95.7851]
93.871 %	[90.6265-96.0415]
94.1935 %	[91.0093-96.2959]
94.5161 %	[91.3942-96.5483]
94.8387 %	[91.7814-96.7984]
95.1613 %	[92.171-97.0461]
95.4839 %	[92.5632-97.2911]
95.8065 %	[92.9584-97.5331]
96.129 %	[93.3568-97.772]
96.4516 %	[93.7588-98.0073]
96.7742 %	[94.1648-98.2385]
97.0968 %	[94.5754-98.4652]
97.4194 %	[94.9912-98.6867]
97.7419 %	[95.4132-98.902]
98.0645 %	[95.8424-99.11]
98.3871 %	[96.2805-99.3091]
98.7097 %	[96.7298-99.4971]
99.0323 %	[97.1938-99.6703]
99.3548 %	[97.6786-99.8229]
99.6774 %	[98.1957-99.943]
100 %	[98.776-100]

C.2. Lausanne Database

Cuadro C.2: Lausanne Database CV Accuracy- Confidance Interval Correspondance.

CV Accuracy	Confidance Interval [Lower Bound- Upper Bound]
0%	[0-8.3800]
2.381%	[0.42154-12.3212]
4.7619%	[1.3158-15.7899]
7.1429%	[2.459-19.0094]
9.5238%	[3.7662-22.0651]
11.9048%	[5.1938-25.0004]
14.2857%	[6.716-27.8411]
16.6667%	[8.3159-30.604]
19.0476%	[9.982-33.3008]
21.4286%	[11.7058-35.9398]
23.8095%	[13.481-38.5275]
26.1905%	[15.3026-41.0687]
28.5714%	[17.167-43.5672]
30.9524%	[19.0711-46.026]
33.3333%	[21.0125-48.4475]
35.7143%	[22.9892-50.8336]
38.0952%	[24.9997-53.186]
40.4762%	[27.0429-55.5057]
42.8571%	[29.1177-57.7938]
45.2381%	[31.2234-60.0509]
47.619%	[33.3596-62.2775]
50%	[35.526-64.474]
52.381%	[37.7225-66.6404]
54.7619%	[39.9491-68.7766]
57.1429%	[42.2062-70.8823]
59.5238%	[44.4943-72.9571]
61.9048%	[46.814-75.0003]
64.2857%	[49.1664-77.0108]
66.6667%	[51.5525-78.9875]
69.0476%	[53.974-80.9289]
71.4286%	[56.4328-82.833]
73.8095%	[58.9313-84.6974]
76.1905%	[61.4725-86.519]
78.5714%	[64.0602-88.2942]
80.9524%	[66.6992-90.018]
83.3333%	[69.396-91.6841]
85.7143%	[72.1589-93.284]
88.0952%	[74.9996-94.8062]
90.4762%	[77.9349-96.2338]

92.8571 %	[80.9906-97.541]
95.2381 %	[84.2101-98.6842]
97.619 %	[87.6788-99.5785]
100 %	[91.6201-100]

C.3. Expanded Database

Cuadro C.3: Expanded Database CV Accuracy- Confidance Interval Correspondance.

CV Accuracy	Confidance Interval [Lower Bound- Upper Bound]
0 %	[0-4.0100]
1.087 %	[0.19213-5.9028]
2.1739 %	[0.5982-7.5835]
3.2609 %	[1.1151-9.1534]
4.3478 %	[1.7036-10.6517]
5.4348 %	[2.3435-12.0985]
6.5217 %	[3.023-13.5058]
7.6087 %	[3.7344-14.8812]
8.6957 %	[4.4721-16.2303]
9.7826 %	[5.2324-17.5568]
10.8696 %	[6.0122-18.8638]
11.9565 %	[6.8091-20.1536]
13.0435 %	[7.6215-21.428]
14.1304 %	[8.4477-22.6886]
15.2174 %	[9.2865-23.9365]
16.3043 %	[10.137-25.1728]
17.3913 %	[10.9982-26.3984]
18.4783 %	[11.8693-27.6141]
19.5652 %	[12.7498-28.8203]
20.6522 %	[13.6391-30.0178]
21.7391 %	[14.5367-31.207]
22.8261 %	[15.4422-32.3883]
23.913 %	[16.3551-33.5621]
25 %	[17.2753-34.7288]
26.087 %	[18.2022-35.8886]
27.1739 %	[19.1358-37.0419]
28.2609 %	[20.0757-38.1887]
29.3478 %	[21.0217-39.3295]
30.4348 %	[21.9737-40.4643]
31.5217 %	[22.9314-41.5933]
32.6087 %	[23.8948-42.7167]
33.6957 %	[24.8636-43.8347]

34.7826 %	[25.8378-44.9473]
35.8696 %	[26.8171-46.0547]
36.9565 %	[27.8016-47.157]
38.0435 %	[28.7912-48.2543]
39.1304 %	[29.7856-49.3466]
40.2174 %	[30.785-50.434]
41.3043 %	[31.7891-51.5167]
42.3913 %	[32.798-52.5946]
43.4783 %	[33.8116-53.6678]
44.5652 %	[34.8298-54.7363]
45.6522 %	[35.8527-55.8002]
46.7391 %	[36.8801-56.8595]
47.8261 %	[37.9121-57.9143]
48.913 %	[38.9487-58.9645]
50 %	[39.9898-60.0102]
51.087 %	[41.0355-61.0513]
52.1739 %	[42.0857-62.0879]
53.2609 %	[43.1405-63.1199]
54.3478 %	[44.1998-64.1473]
55.4348 %	[45.2637-65.1702]
56.5217 %	[46.3322-66.1884]
57.6087 %	[47.4054-67.202]
58.6957 %	[48.4833-68.2109]
59.7826 %	[49.566-69.215]
60.8696 %	[50.6534-70.2144]
61.9565 %	[51.7457-71.2088]
63.0435 %	[52.843-72.1984]
64.1304 %	[53.9453-73.1829]
65.2174 %	[55.0527-74.1622]
66.3043 %	[56.1653-75.1364]
67.3913 %	[57.2833-76.1052]
68.4783 %	[58.4067-77.0686]
69.5652 %	[59.5357-78.0263]
70.6522 %	[60.6705-78.9783]
71.7391 %	[61.8113-79.9243]
72.8261 %	[62.9581-80.8642]
73.913 %	[64.1114-81.7978]
75 %	[65.2712-82.7247]
76.087 %	[66.4379-83.6449]
77.1739 %	[67.6117-84.5578]
78.2609 %	[68.793-85.4633]
79.3478 %	[69.9822-86.3609]

80.4348 %	[71.1797-87.2502]
81.5217 %	[72.3859-88.1307]
82.6087 %	[73.6016-89.0018]
83.6957 %	[74.8272-89.863]
84.7826 %	[76.0635-90.7135]
85.8696 %	[77.3114-91.5523]
86.9565 %	[78.572-92.3785]
88.0435 %	[79.8464-93.1909]
89.1304 %	[81.1362-93.9878]
90.2174 %	[82.4432-94.7676]
91.3043 %	[83.7697-95.5279]
92.3913 %	[85.1188-96.2656]
93.4783 %	[86.4942-96.977]
94.5652 %	[87.9015-97.6565]
95.6522 %	[89.3483-98.2964]
96.7391 %	[90.8466-98.8849]
97.8261 %	[92.4165-99.4018]
98.913 %	[94.0972-99.8079]
100 %	[95.9919-100]

Bibliografía

[1] Westman, E., Simmons, A., Zhang, Y., Muehlboeck, J., Tunnard, C., Liu, Y., Collins, L., Evans, A., Mecocci, P., Vellas, B., Tsolaki, M., Kłoszewska, I., Soininen, H., Lovestone, S., Spenger, C., Wahlund, L., 2010. Multivariate analysis of MRI data for Alzheimer's disease, mild cognitive impairment and healthy controls. NeuroImage 54 (2011), 1178-1187.

[2] Heckemann, R.A., Keihaninejad, S., Aljabar, P., Gray, K.R., Nielsen, C., Rueckert, D., Hajnal, J.V., Hammers, A., and The Alzheimer's Disease Neuroimaging Initiative, 2011. Automatic morphometry in Alzheimer's disease and mild cognitive impariment. NeuroImage 56 (2011), 2024-2037.

[3] Fennema-Notestine, C., Hagler Jr., D.J., McEvoy, L.K., Fleisher, A.S., Wu, E.H., Karow, D.S., Dale, A.M., and the Alzheimer's Disease Neuroimaging Initiative, 2009. Structural MRI Biomarkers for Preclinical and Mild Alzheimer's Disease. Human Brain Mapping (2009)

[4] Klöppel, S., Stonnington, C.M., Chu, C., Draganski, B., Scahill, R.I., Rohrer, J.D., Fox, N.C, Jack Jr, C.R., Ashburner, J., Frackowiak, S.J., 2008. Automatic classification of MR scans in Alzheimer's disease. Brain (2008), 131, 681-689.

[5] Cuingnet, R., Gerardin, E., Tessieras, J., Auzias, G., Lehéricy, S., Habert, M-O. Chupin, M., Benali, H., Colliot, O., and The Alzheimer's Disease Neuroimaging Initiative, 2010. Automatic classification of patients with Alzheimer's disease from structural MRI: A comparision of ten methods using the ADNI database. NeuroImage (2010).

[6] Fan, Y., Shen, D., Davatzikos, C., 2005. Classification of Structural Images via High-Dimensional Image Warping, Robust Feature Extraction, and SVM. J. Duncan and G. Gerig (Eds.): MICCAI 2005, LNCS 3749, pp. 1 – 8, 2005.

[7] Davatzikos, C., Xu, F., An, Y., Fan, Y., Resnick, S.M., 2009. Longitudinal progression of Alzheimer's-like patterns of atrophy in normal older adults: the SPARE-AD index. Brain 2009: 132 2026–2035.

[8] Klöppel, S., Stonnington, C.M., Barnes, J., Chen, F., Chu, C., Good, C.D., Mader, I., Mitchell, L.A., Patel, A.C., Roberts, C.C., Fox, N.C., Jack Jr, C.R., Ashburner, J., Fracowiak, R.S.J, 2008. Accuracy of dementia diagnosis—a direct comparison between radiologists and a computerized method. Brain 2008, 131, 2969–2974.

[9] Vemuri, P., Gunter, J.L., Sanjem,M.L., Whitwell, J.L., Kantarci, K., Knopman, D.S., Boeve, B.F., Petersen, R.C., Jack Jr., C.R, 2008. Alzheimer's disease diagnosis in individual subjects using structural MR images: Validation studies. 39 (2008) 1186-1197.

[10] Vemuri, P., Whitwell, J.L., Kantarci, K., Josephs, K.A., Parisi, J.E., Shiung, M.S., Knopman, D.S., Boeve, B.F., Petersen, R.C., Dickson, D.W., Jack Jr, C.R., 2008. Antemortem MRI based STructural Abnormality iNDex (STAND)-scores correlate with postmortem Braak neurofibrillary tangle stage. NeuroImage (2008).

[11] Juszczak, P., Tax, D.M.J., Duin, R.P.W., 2002. Feature Scaling in support vector data description. E.F. Deprettere, A. Belloum, J.W.J. Heijnsdijk, F. van der Stappen (eds.), Proc. ASCI 2002, 8th Annual Conf. of the Advanced School for Computing and Imaging, ASCI, Delft, 2002, 95-102.

[12] Fan, Y., Batmanghelich, N., Clark, C.M., Davatzikos, C., the Alzheimer's Disease Neuroimaging Initiative, 2008. Spatial patterns of brain atrophy in MCI patients, identified via high-dimensional pattern classification, predict subsequent cognitive decline. NeuroImage 39 (2008) 1731-1743.

[13] Desikan,R.S., Cabral, H.J., Hess, C.P., Dillon, W.P., Glastonbury, C.M., Weiner, M.W., Schamansky, N.J., Greve, D.N., Salat, D.H., Buckner, R.L., Fischl, B.; Alzheimer's Disease Neuroimaging Initiative, 2009. Automated MRI measures identify individuals with mild cognitive impairment and Alzheimer's disease. Brain 2009: 132; 2048–2057.

[14] Pereira,F., Mitchell, T., Botvinick, M., 2009. Machine learning classifiers and fMRI: A tutorial overview. NeuroImage 45 (2009) S199–S209.

[15] Pereira,F., Botvinick,M., 2011. Information mapping with pattern classifiers: A comparative study. NeuroImage 56 (2011) 476–496.

[16] Davatzikos, C., Bhatt, P., Shaw, L.M., Batmanghelich, K.N., Trojanowski, J.Q., 2011. Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification. Neurobiology of Aging 32 (2011) 2322.e19 –2322.e27.

[17] Dukart, J., Schroeter, M.L., Mueller, K., The Alzheimer's Disease Neuroimaging Initiative, 2011. Age Correction in Dementia – Matching to a Healthy Brain. PLoS ONE 6(7): e22193. doi:10.1371/journal.pone.0022193.

[18] Good, C.D., Johnsrude, I.S., Ashburner, J., Henson, R.N.A., Friston, J.K., Frackowiak, R.S.J., 2001. A Voxel-Based Morphometric Study of Ageing in 465 Normal Adult Human Brains. NeuroImage 14, 21–36 (2001) doi:10.1006/nimg.2001.0786

[19] Dukart, J., Mueller, K., Horstmann, A., Barthel, H., Möeller, H.E., Villringer, A., Sabri, O., Schroeter, M.L., 2011. Combined Evaluation of FDG-PET and MRI Improves Detection and Differentiation of Dementia. PLoS ONE 6(3): e18111. doi:10.1371/journal.pone.0018111

[20] Franke, K., Ziegler, G., Klöppel, S., Gaser, C., and the Alzheimer's Disease Neuroimaging Initiative, 2010. Estimating the age of healthy subjects from T1 -weighted MRI scans using kernel methods: Exploring the influence of various parameters. NeuroImage 50 (2010) 883–892.

[21] Guyon, I., Weston, J., Barnhill, S., Vapnik, V., 2002. Gene Selection for Cancer Classification using Support Vector Machines. Machine Learning, 46, 389–422, 2002.

[22] Pereira, F., 2007. Beyond Brain Blobs: Machine Learning Classifiers as Instruments for Analyzing Functional Magnetic Resonance Imaging Data. Carnegie Mellon University,School of Computer Science,Pittsburgh,PA,15213.

[23] Fung, G., Stoeckel, J., 2007. SVM feature selection for classification of SPECT images of Alzheimer's disease using spatial information. Knowl Inf Syst (2007) 11(2): 243–258. DOI 10.1007/s10115-006-0043-5.

[24] Chaves,R., Ramírez, J., Górriz, J.M., López, M., Salas-Gonzalez, D., Álvarez, I., Segovia, F., 2009. SVM-based computer-aided diagnosis of the Alzheimer's disease using *t*-test NMSE feature selection with feature correlation weighting. Neuroscience Letters 461 (2009) 293–297.

[25] Liang, P., Wang, Z., Yang, Y., Li, K., 2012. Three Subsystems of the Inferior Parietal Cortex are Differently Affected in Mild Cognitive Impariment. Journal of Alzheimer's Disease 29 (2012) 1–13. DOI 10.3233/JAD-2012-111721

[26] Davatzikos, C., Fan, Y., Wu, X., Shen, D., Resnick, S.M., 2008. Detection of Prodromal Alzheimer's Disease via Pattern Classification of MRI. Neurobiol Aging. 2008 April ; 29(4): 514–523. doi:10.1016/j.neurobiolaging.2006.11.010.

[27] Lao, Z., Shen, D., Xue, Z., Karacali, B., Resnick, S.M., Davatzikos, C., 2004. Morphological classification of brains via high-dimensional shape transformations and machine learning methods. NeuroImage 21 (2004) 46– 57.

[28] Hall, M.A., 2009. Correlation-based Feature Selection for Machine Learning. Deparment of Computer Science, The Universitiy of Waikato, Hamilton, New Zeland.

[29] Yu, L., Liu, H., 2003. Feature Selection for High-Dimensional Data: A Fast Correlation-Based Filter Solution. Proceedings of the Twentieth International Conference on Machine Learning (ICML-2003), Washington DC, 2003.

[30] Guyon, I., Elisseeff, A., 2003. An Introduction to Variable and Feature Selection. Journal of Machine Learning Research 3 (2003) 1157-1182.

[31] Chen, Y.W., Lin, C.J., 2003. Combining SVMs with Various Feature Selection Strategies. NIPS 2003 Feature Selection Challenge.

[32] Wu, Y., Zhang, A., 2004. Feature Selection for Classifying High-Dimensional Numerical Data. CVPR'04 Proceedings of the 2004 IEEE computer society conference on Computer vision and pattern recognition.

[33] Rakotomamonjy, A., 2003. Variable Selection Using SVM-based Criteria. Journal of Machine Learning Research 3 (2003) 1357-1370.

[34] Desikan, R.S., Ségonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L., Dale, A.M., Maguire, R.P., Hyman, B.T., Albert, M.S., Killiany, R.J., 2006. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. NeuroImage 31 (2006) 968 – 980.

[35] Fischl, B., van der Kouwe, A., Destrieux, C., Halgren, E., Ségonne, F., Salat, D.H., Busa, E., Seidman, L.J., Goldstein, J., Kennedy, D., Caviness, V., Makris, N., Rosen, B., Dale, A.M., 2004. Cerebral Cortex January 2004;14:11–22; DOI: 10.1093/cercor/bhg087.

[36] Fischl, B., Salat, D.H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, ,D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., Dale, A.M, 2002. Whole Brain Segmentation: Automated Labeling of Neuroanatomical Structures in the Human Brain. Neuron, Vol. 33, 341–355.

[37] Hsu, C-W., Chang, C-C., Lin, C-J., 2010. A Practical Guide to Support Vector Classification. URL: http://www.csie.ntu.edu.tw/~cjlin

[38] Chang, C-C., Lin, C-J., 2011. LIBSVM: A Library for Support Vector Machines. URL: http://140.112.30.28/~cjlin/papers/libsvm.pdf

[39] Fan, R-E., Chang, K-W., Hsieh, C-J., Wang, X-R., Lin, C-J., 2012. LIBLINEAR: A Library for Large Linear Classification. URL: http://www.csie.ntu.edu.tw/~cjlin/papers/liblinear.pdf

[40] Yu, Y., 2007. SVM-RFE Algorithm for Gene Feature Selection. University of Delaware. Electrical& Computer Engineering. Computer &Information Science. URL: http://www.eecis.udel.edu/~yuy/report0531.pdf

[41] Alzheimer's Association, 2012 Alzheimer's Disease Facts and Figures, Alzheimer's & Dementia, Volume 8, Issue 2.

[42] Hänninen, T., Hallikainen, M., Tuomainen, S., Vanhanen, M., Soininen, H., 2002. Prevalence of mild cognitive impairment: a population-based study in elderly subjects. Acta Neurologica Scandinavica, 106: 148–154. doi: 10.1034/j.1600-0404.2002.01225.x

[43] Lopez, L.O., Jagust, W.J., DeKosky, S.T., Becker, J.T., Fitzpatrick, A., Dulberg, C., Breitner, J., Lyketsos, C., Jones, B., Kawas, C., Carlson, M., Kuller, L., 2003. Prevalence and Classification of Mild Cognitive Impairment in the Cardiovascular Health Study Cognition Study. Archives of Neurology 2003; 60:1385-1389.

[44] Petersen, R.C., Smith, G.E., Warning, S.C., Ivnik, R.J., Tangalos, E.G., Kokmen, E., 1999. Mild Cognitive Impariment. Clinical Characterization and Outcome. Archives of Neurology, 1999; 56:303-308.

[45] Davies, P.L., 1993. Aspects of robust linear regression. The annals of statistics, 1993; Vol. 21, No. 4, 1843-1899.

[46] Maronna, R., Martin, D., Yohai, V., 2006. Robust Statistics: Theory and Methods. Wiley series in probability and statistics. John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, 2006.

[47] Braak, H., Braak, E., 1991. Neuropathological stageing of Alzheimer- releated changes. Acta Neuropathol, 1991. 82:239 - 259

[48] O'brien, J.T., 2007. Role of imaging techniques in the diagnosis of dementia. The British Journal of Radiology, 80,2007, S71–S77.

[49] Ries, M.L., Carlsson, C.M., Rowley, H.A., Sager, M.A., Gleason, C.E., Asthana, S., Johnson, S.C., 2008. Magnetic resonance imaging characterization of brain structure and function in mild cognitive impairment: a review. J. Am. Geriatr. Soc. 56, 920–934.

[50] Du, A.T., Schuff, N., Amend, D., Laakso, M.P., Hsu, Y.Y., Jagust, W.J., yaffe, K., Kramer, J.H., Reed, B., Norman, D., Chui, H.C., Weiner, M.W.,Magnetic resonance imaging of the entorhinal cortex and hippocampus in mild cognitive impairment and Alzheimer's disease, 2001. J Neurol Neurosurg Psychiatry 2001;71:441–447.

[51] Jack Jr., C.R., Petersen, R.C., O'Brien, P.C., Tangalos, E.G., 1992. MR-based hippocampal volumetry in the diagnosis of Alzheimer's disease. Neurology 42, 183–188.

[52] Laakso, M.P., Soininen, H., Partanen, K., Lehtovirta, M., Hallikainen, M., Hanninen, T., Helkala, E.L., Vainio, P., Riekkinen Sr., P.J., 1998. MRI of the hippocampus in Alzheimer's disease: sensitivity, specificity, and analysis of the incorrectly classified subjects. Neurobiol. Aging 19, 23–31.

[53] Juottonen, K., Laakso, M.P., Partanen, K., Soininen, H., 1999. Comparative MR analysis of the entorhinal cortex and hippocampus in diagnosing Alzheimer disease. AJNR Am. J. Neuroradiol. 20, 139–144.

[54] Xu, Y., Jack Jr., C.R., O'Brien, P.C., Kokmen, E., Smith, G.E., Ivnik, R.J., Boeve, B.F., Tangalos, R.G., Petersen, R.C., 2000. Usefulness of MRI measures of entorhinal cortex versus hippocampus in AD. Neurology 54, 1760–1767.

[55] Edelman, R.R., Warach, S., 1993. Magnetic Resonance Image. The New England Journal of Medicine. Medical Progress, 708- 716.

[56] New Zeland Brain Research Institute, 2012. Magnetic Resonance Imaging at 3 Tesla. http://www.nzbri.org/research/labs/mri.php

[57] Magnetic Resonance - Technology Information Portal, 2012. MRI Imageshttp://www.mr-tip.com/serv1.php?type=img&img=Brain%20MRI%20Images%20Axial%20T2

[58] Laboratory of Functional and Molecular Imaging, National Institute of Neurological disorders and Stroke. http://www.lfmi.ninds.nih.gov/gallery.php

[59] Kloppel, S., Stonnington, C.M., Chu, C., Draganski, B., Scahill, R.I., Rohrer, J.D. et al., 2008. A plea for confidence intervals and consideration of generalizability in diagnostic studies. Brain 2008a (in press).

[60] Guyon, I., Makhoul, J., and Vapnik, V., 1998. What size test set gives good error rate estimates? IEEE Pattern Analysis and Machine Intelligence, Vol. 20, January 1998, 52-64.

[61] Sherrod, P.H., 2012. DTREG, Software For Predicting Modeling and Forecasting, 2012. http://www.dtreg.com/svm.htm

[62] Jakkula, V., 2012. Tutorial on Support Vector Machine (SVM). School of EECS, Washington State University, 2012.

[63] United States Census Bureau, 2012. School Districts.http://www.census.gov/did/www/schooldistricts/index.html

[64] Xie, Z.X., Hu ,Q.H., Yu, D.R, 2006. Improved Feature Selection Algorithm Based on SVM and Correlation. J. Wang et al. (Eds.): ISNN 2006, LNCS 3971, pp. 1373–1380, 2006.

[65] Chang, K.J, Jara, H., 2005. Applications of quantitative T1, T2, and proton density to diagnosis. Journals, Volume 34, Number 1, January 2005, http://www.appliedradiology.com/Issues/2005/01/Supplements/Applications-of-quantitative-T1,-T2,-and-proton-density-to-diagnosis.aspx

[66] Dugdale, D.C., Hoch, D.B., Zieve, D., 2008. Medical Encyclopedia, Lobes of the brain. Universitiy of Maryland. Medical center, 2008, http://www.umm.edu/imagepages/9549.htm

[67] Greve, D., 2012. Working with FreeSurfer ROIs. FreeSurfer Course, April 2-4, 2012, https://surfer.nmr.mgh.harvard.edu/

[68] Salat, D.H., Greve, D.N., Pacheco, J.L., Quinn., B.T., Helmer, K.G., Buckner, R.L., Fischl, B., 2009. Regional white matter volume differences in nondemented aging and Alzheimer's disease. NeuroImage 44, 1247–1258, 2009.

[69] R.E. Fan, K.W. Chang, C.J. Hsieh, X.R. Wang, C.J. Lin, 2009. LIBLINEAR: A Library for Large Linear Classification. Journal of Machine Learning Research 9, 2008, 1871-1874. Software available at http://www.csie.ntu.edu.tw/~cjlin/liblinear

[70] Wallis, S., 2009. Binomial distributions, probability and Wilson's confidence interval. London: Survey of English Usage, University College of London, 20 December 2009.

[71] Brown, L.D., Cai, T.T., DasGupta A., 2001. Interval Estimation for a Binomial Proportion. Statistical Science 2001, Vol. 16, No. 2, 101–133.