Acknowledgements

To Jonas, my project supervisor, for all the time he has devoted me and the patience he has had with me. Thanks a lot.

To Dimitri, my project director, for giving me the chance to work at EPFL.

To Iñaki, for his assistance and the good advices given.

To Juan Carlos, Vanesa, Elie y Jesús for the good moments spent in Zaragoza during the master and for being there whenever I needed them.

To Fernando, Toño, Edu, Pilar,Eva,Pili, and all the other Master mates. You are great.

To Domingo, for our mutual support in the difficult beginning in Lausanne,for the time we spent looking for accommodation, for bearing me and for all the good moments in Switzerland.

To Elena, Mireia and Diego for supporting me during these months and for "burning" my laboratory on my birthday and give me a break in the last sprint.

To Maria, Alberto, Maria J.,Juan and all my Erasmus mates for the luck we had being together.

To **my parents**, my brothers and sisters and my brother-in-law for trusting me. And specially, it dedicated to the youngest in my family, Marcelo.

Abstract

Functional magnetic resonance imaging (fMRI) opens unprecedented possibilities to study the human brain at work in a non-invasive and non-radiating way. Using the hemodynamic response as a proxy for neuronal activation, this imaging modality is able to generate whole-brain volumes about every 1-4 seconds, depending on the desired spatial resolution (about 1-5 mm in each dimension).

Conventional fMRI data analysis relies on a voxel-wise linear regression followed by statistical hypothesis testing. Another interesting measure that can be extracted from fMRI data is functional connectivity; i.e., for two given brain regions, the temporal correlation of their timecourses indicates their degree of "interaction". The advantage of functional connectivity is that it does not necessarily require a task and can thus be computed for resting-state scanning sessions (without any stimulation paradigm).

Recently, the application of pattern recognition to neuro-imaging has gained significant interest in the field. In particular, this methodology can be applied to "decode" the subject's status (e.g., control versus disease/disorder) from functional connectivity measurements during resting state. The functional connectivity matrix is built from 90 regions as defined by a commonly used atlas. Here, we want to study the confound of the scanner on the classification result, which is an important topic to allow the generalization of this method for multicentric studies in the future.

Our approach is to see whether confounds such as scanner, but also age, sex, influence the classification task. For this purpose, we have two sets of resting-state fMRI data from healthy subjects acquired using two identical MR scanners (Siemens 3T), and we have one set of resting-state fMRI from age- and gender-matched healthy subjects and patients.

In the first part, we have implemented a script that automatically scans the fields in the DICOM header to check if any differences occurred. This script allows to easily screen datasets for subtle differences in the scanning protocol (e.g., echo time, repetition time) and to generate histogram of any other parameters that are stored in the data (e.g., age). In the second part, we implemented a classification pipeline with the task to decode the scanner from the data. Specifically, we showed that the scanner can be identified from the mean fMRI intensity images only, which clearly indicates that scanner-specific confounds do impact the image data. The spatial patterns that distinguish both scanners are reminiscent for subtle changes in field homogeneity of the head coils.

Finally, we explore how the scanner confound influence functional-connectivity decoding, and we propose strategies to diminish this effect such as projecting the connectivity matrix in a subspace of components that do not allow to decode the scanner effect.

Resumen

Functionalmagneticresonanceimaging (fMRI) es una reciente técnica no invasiva ni radiactiva que permite estudiar la actividad funcional del cerebro. Esta técnica consiste en la adquisición de una secuencia de volúmenes cerebrales y basándose en la respuesta hemodinámica del mismo permite mapear las regiones donde hay actividad.

Una de las posibilidades que permite fMRI es analizar las conexiones funcionales entre las distintas regiones del cerebro tanto cuando el sujeto está en estado de reposo como durante la ejecución de una tarea. A partir de la correlación de los diferentes volúmenes adquiridos en una prueba fMRI se puede averiguar el grado de interacción que existe entre dos regiones cerebrales. Recientemente, la teoría de reconocimiento de patrones está ganado un notable interés en el campo de la neuroimagen y permite diferenciar entre estado funcional de sujetos enfermos y controles. Algunos atlas catalogan el cerebro en 90 regiones por hemisferio y con fMRI se está estudiando cómo el patrón de conectividad que posee un cerebro enfermo, en reposo o durante cualquier actividad para desarrollar su matriz de conectividad cerebral.

El objetivo del proyecto fue estudiar si existen factores de confusión dificultan la tarea del conocer adecuadamente como es la matriz de conectividad durante el estado de reposo. Esto se fundamente en que previamente fue creada una matriz de conectividad para el estado de reposo pero los resultados conseguidos no fueron los esperados. Para ello se ha analizado si estos factores de confusión son debidos al escáner, la edad, el género del sujeto o simplemente a los métodos usados en la clasificación de las matrices de conectividad.

Para buscar la existencia de estos factores de confusión se ejecutado dos procedimientos de estudio. Por un lado, se ha implementado un método que permite analizar las cabeceras de los archivos DICOM. Estas contienen información (tiempo de eco, tiempo de repetición, etc.) acerca de la configuración de los diferentes parámetros del escáner. Así, el objetivo fue determinar si los parámetros de dos escáneres diferentes estaban configurados de manera similar o si por el contrario había grandes diferencias entre ellos.

El segundo procedimiento trata de analizar la información de intensidad sobre las imágenes de fMRI para determinar si existen cambios significativos entres éstas tanto espacial como temporalmente. Para ello se han implementado varios algoritmos basados en *machine learning* con el objetivo de chequear si realmente existían diferencias notables de intensidad entre las distintas imágenes.

Indice

1. Introduction

1. 1. Functional Magnetic Resonance

The discovery magnetic resonance imaging is revolutionizing the neuroimaging science since this can be used to map changes in brain hemodynamics. Knowing the brain areas which participate in specific function is used to build map of human brain function. The technique for studying the functional cerebral activity is called functional magnetic resonance imaging (fMRI).This method is based on non-invasive recordings of associated with changes in the blood oxygen, offering a high spatial resolution. The new ability to directly observe brain function let assess neurological status and neurosurgical risk.

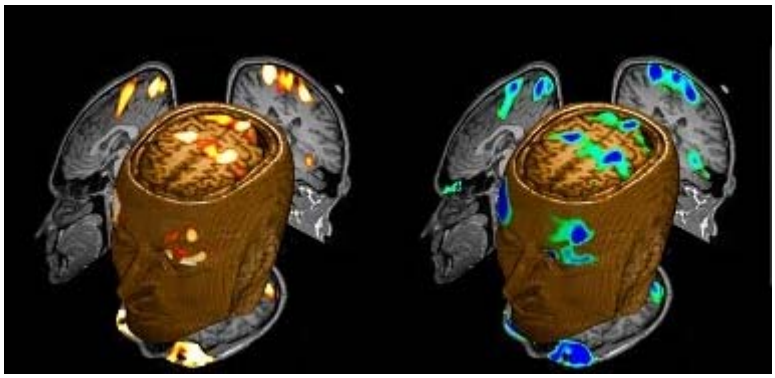


Figure 1.1.fMRI [1].

Specifically, fMRI is based on the increase in blood flow to the local vasculature that accompanies neural activity in the brain. This effect is due to reduction in deoxyhemoglobin because the increase in blood flow occurs without an increase of similar magnitude in oxygen extraction.Thedeoxyhemoglobin has a paramagnetic quality which acts as contrast enhancing agent and serves as the source of the signal for fMRI.

fMRI allow the study of the brain areas associated with specific functions.For clinical use, this is very important then allow to build functional markers, i.e., mapof cerebral activity on both healthy subjects and subjects suffering from different diseases. Thefinal objective will be to develop afunctional diseasesbrains atlas which helps to medical teams to diagnose cerebral illness by imaging techniques.

1. 2. Functional Connectivity

The activated brain areas are identified by Magnetic Resonance (MR) signal changes that occur in response to a specific stimulus or task or as a result of some other change in brain state. Actually the neuroimaging allows studying the connection between the different brain regions using multiples methods. Two regions are functionally connected when exists some correlation in their activities.

In order to obtain a good knowledge of the functional connectivity, the subjects are studied performing different activities. Moreover, it has been observed the existence of a temporal correlation between both brain hemispheresin resting state and this generates a large interest for researchers.

The fMRI signal varies within a small intensity range which is not easily appreciated by the human eye. For this reason, machine learning algorithms are used to search the activity pattern existent in the brain and decoding stimuli, mental behavior, etc. For analyzing fMRI data is necessary the use of classifiers which learn said patterns and it can help to recognize new population characterized by the same mental behavior.

1. 3. Brain decoding

So as todevelop functional market is a good knowledge of the brain connections and areas active associated to aspecific function. Thus, it is required to acquire a huge quantity of functional images to be able to correctly understand these brain connections. For each subject, atest of fMRI generates multiple volumes of the brain acquired temporally, each volume is constituted byhundreds of voxels (i.e., 3D pixels). Elaborating a good predictive marker requires to have functional information from many different subjects thus it is necessary to manipulate millions of voxels. Pattern recognition is a statistical powerful tool that allows analyzing a large amount of data and developing markers.

1. 4. Functional connectivity decoding

After knowing howthe different brain regions are correlated, it is possible to build a connectivity matrix for aspecific mental behavior. The connectivity matrix is based on graph theory. Vertices are identified as brain regions and these are connected by edges whose weights represent the relationship between different brain areas. So,this

allowsknowing brainstates from functional connectivity regions, instead of the brain voxel activation values.

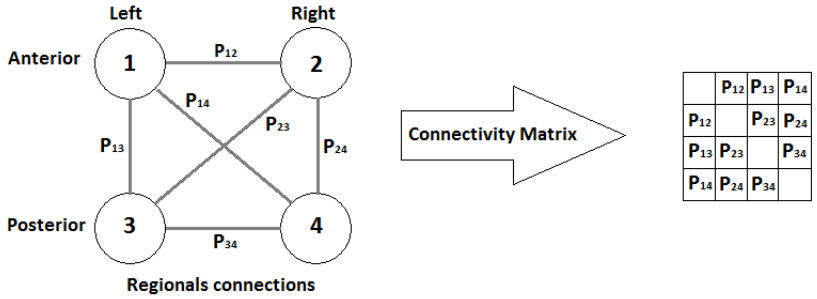


Figure 1.2. Regionals connections and symmetric connectivity matrix for four brain regions modeled.

The main idea is to analyze the brain connectivity as a complete graph. E.g. if the brain is segmented in 4 regions (Figure 1.2), the number of connections is equal to six. Thus, the row and column of the connectivity matrix are brain regions and their edges the connectivity level.

This technique may be interesting to develop functional markers easily because it allows synthesizing several activated voxels set as a single brain region.

1. 5. Confounding factor

The motivation of this project is to learn more about functional connectivity in healthy and sick brains, and apply this knowledge for the developing of predictive markers. As a blood analysis can determine severalillness, functional markers may diagnose different brains diseases.

fMRI is an expensive technique and it is difficult to find many subjects scanned in the same machine for a specific function. For this reason, the markers are made by functional images obtained from different scanners. It could then induce some confounding factors which make difficult to build a reliable marker.

When in a given activity state,different scanners never give identical results for the same subject. This is the reason why developing functional markers remains a difficult task. Therefore, depending on the scanners used, some differences are found both on software and hardware modules:

- The acquisition time can be tuned by the operator.
- The magnetic field strength can be slightly different. So, the average intensity pictured on images might not be similar.
- The use of a different pillow changes the distance between patient's head and antennas of the scanner.
- Some antennas might be out of order.
- Etc.

All these parameters might consequently induce differences or confounding factors on fMRI analysis.

1. 6. Goal of the project

Thus, the main goal of this project is to know whether scanners generate confounding factors which influence negatively on functional connectivity markers.

For this purpose, the possible existence of confounding factors will be studied onDICOM headers present in functional images.The technical and acquisition parameters will be analyzed with an algorithm allowing theanalysis of the DICOM header. Indeed, this algorithm is able to check if a relationship exists between these parameters for different scanned groups.

Additionally, pattern recognition techniques can help to estimate spatial or temporal negative effects on information obtained from the fMRI images.By analyzing the spatio-temporal stability of the fMRI signal on the functional images, it could aid to kwon whether there is any anomalous effect.

2. Effects of changes in acquisition protocol

MRI data can be collected from two different protocols: Structural or functional acquisition. According tothese protocols,therealized studies aim to check if the confounding factors depend onscanning site, group scanned, or changes in the acquisition protocol over time.

2. 1. Databases

The subjects obtained to analyze the changes in the protocol acquisition come from two different databases.The first it is a group of subjects given by the "Center HospitalierUniversitaireVaudois" (CHUV) from Lausanne while the second group is constituted by patients scanned in the Brain Behavior Laboratory (BBL) from Geneva and the "HôspitauxUniversitaires de Genève" (HUG).

2. 1. 1. CHUVrest database

The CHUVrest database consists in 36 subjects, 22 with multiple sclerosis and 14 healthy controls, scanned in resting-state for two different protocols: structural (MPRAGE) and functional (fMRI). By subject, a time-series of 450 images (temporal files) for fMRI protocol and 190 temporal files for MPRAGE protocol are acquired.For each subject, each time-series is the temporal information acquired during one scanned session.

2. 1. 2. CMSTrest database

The CMSTrest (Cognitive Memory Stress Test) database is a group of 59 subjects come from the BBL and HUGscanning sites, but only 38 of them are used in this study.The BBLandthe HUG sites use a Siemens 3T TIM Trio scanner with the same software revision (VB15), but operators and acquisition traditions may differ. The CMSTrest database is divided in three categories of subjects: Younger Controls (YC), Elder Controls (EC) and Mild Cognitive Impairment (MCI). These patients have been scanned in resting-state for fMRI protocol and the number of temporal volumes

acquired is of 450 by subject. CMSTrest database has been organizedinto two different datasets:

- *Version A*: 20 patients from CMSTrest database distributed according to gender, acquisition scanner and division (EC, YC and MCI), seeTable 2.1. (a).
- *Version B*: 24 subjects grouped by gender and group (Table 2.1. (b).).

	HUG		**BBL**	
	Male	*Female*	*Male*	*Female*
YC	1	0	1	0
EC	2	4	2	4
MCI	2	1	2	1

(a)

	Male	**Female**
YC	4	4
EC	4	4
MCI	4	4

(b)

Table 2.1. Distribution of CMSTrest database.**(a)**Version A. **(b)**Version B.

2. 2. The DICOM standard

The Digital Imaging and Communications in Medicine (DICOM) is a format the interchange of medical imaging.It is mainly used formanipulating, storing, picturing and transmitting these one on TCP/IP protocol.

One DICOM file is structured in a header and the image acquired. The header iscomposed of severalstandard fields set which specify both administrative data (patients data, recording place, etc.) and data about the images. Some parameters or DICOM fields like the TR (Time of Repetition), TE (Time of Echo), number of slices, the proof of slices, etc. could be changed by the operator and induce confounding factors.

2. 3. Design of the DICOM checker

For the CHUVrest dataset, the goal isto ascertain whether there is relationship between the successful classification of subjects and changes in the acquisition protocol fields. For the CMSTrest dataset, the goal is to study whether acquisition protocol parameters change between sites (Version A), and whether the acquisition protocol parameters change between different subject groups (Version B).

In order to perform the analysis, all DICOM fields have been read temporally by subject (time-series by patient) and then this will be comparedon all patients from the dataset. Finally,thefields whose information is variable will be analyzed to determine his relationship with the classification of subjects (CHUVrest) or scanning site (CMSTrest version A) or subject group (CMSTrest version B).

The DICOM checker will receive one database of patients. At the beginning,the temporal DICOM information has been analyzed over each subject. The fields, whose values are different in each volume from the time-series, will be deleted. Normally, these are fields such as*Filename*, *FileModDate* (hour and date of the acquisition), etc. which do not give relevant information. The rest of fields by subject are stored into a table called *FieldsSubj*.This table has by row the different fields and by column the values for each temporal field.*FieldSubj* can have fields which take one value alone and another fields which take several values.

The *FieldsSubj* tables, from the different subject, are compared and thus it can search the fieldswhose values are different among all patients. It is possible that in the *FieldsSubj* tables do not appear the same fields, for this, only the common fields will be compared. The criterion for filtering the non-interesting fields among subject is:

- The fields whose values are different between all subjects, are not taken in account because it is not possible to establish one correlation with these ones.
- The fields whose values are similar between all subjects, are not taken in account.

Then overthe rest of fields a new table is built, *CommonFields*, where the interesting fields among all subjects from the database are stored. *CommonFields* has by row the interesting fields and by column the different values of this field between all subjects.

*FieldsSubj*and*CommonFields*are used to build an integer matrix, called *IntMat*. This matrix will have as row the interesting fields found previously and as column the subjects that are in the database. The field *i* from the *CommonFields* takes *n* different values, thus an integer from 1 to *n*will be assigned for this field. The process is repeated for all the fields present in*CommonFields*. Then the interesting fields are read over the *FieldsSubj* tables. The integer matrix will be built according to the values taken in the

interesting fields by subject and the integer assigned to *CommonFields* table for the same field.

Fields	Values fields		
'ImplementationClassUID'	'1.2.840.114089.1.0.0.3.1.8'	'1.2.840.114089.1.0.0.3.2.2'	
'ImplementationVersionName'	'DCF 3.1.8b'	'DCF 3.2.2c'	
'InstanceCreationDate'	'20081023'	'20081124'	'20081119'
'StudyDate'	'20081023'	'20081124'	'20081119'

(a)

1	1	2	1
2	1	*2*	1
1	3	3	2
2	2	1	3

(b)

Table 1.1. (a) *CommonFields* table.**(b)***IntMat* matrix. This matrix has a size of [number-fields-interest, number-subject]. If itis taken the value *IntMat(2,3)*, this is equal to *2*. That value means that the field 'ImplementationVersionName' for the patients 3 is equivalent to 'DCF 3.2.2c' from the CommonFields table because it is in the column 2 and row from specific field. It is same for other fields of interest.

Finally, the correlation is performed between the integer matrix and a vector of classification, called *Labels* vector. This vector will have a size of [1,*N*], being *N* the number of subjects from the database, and it will take the value 0 or 1 according to a concrete criterion desired. For example, it could get assigned the value 0 for those subjects scanned in BBL and 1 for scanned them in HUG.

The correlation between the matrix and the vector described, indicates whether there is some important relationship between the interesting fields and a subject classification type.

2. 4. Analysis and Results

CHUVrest and CMSTrest have been analyzed with the DICOM checker. So for each one of them has been assessed hisIntMatmatrix. Then, the Pearson Correlation Coefficient (*r*) between IntMat (X) and Labels (Y) vector is computed.

$$r = \frac{\sigma_{xy}}{\sigma_x \cdot \sigma_y} \qquad [2.1]$$

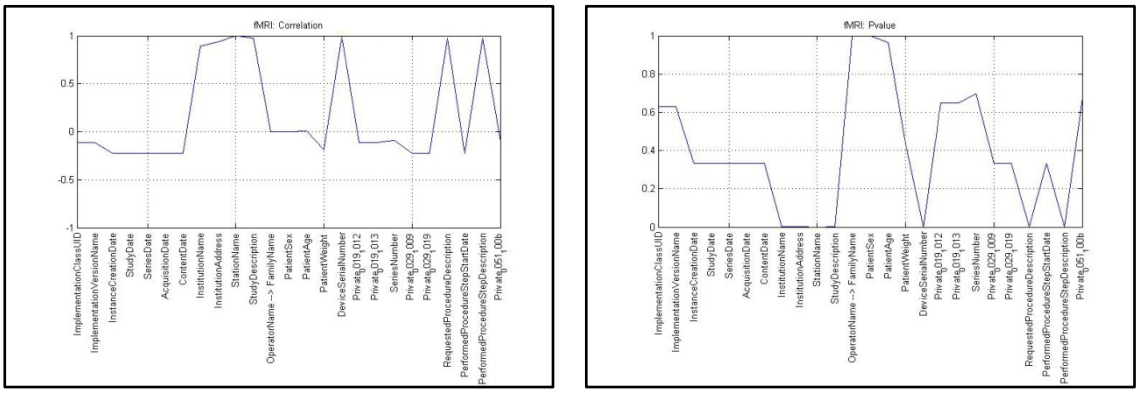
whereσ_{xy} is (X,Y)'s covariance and σ_X y σ_Y theirstandard deviations.

If correlation coefficients are less than -0.4 for a significant field and his p-value is lower than 0.05, it can be rejected that the field has a important significance.

2.4. 1.Site-specific effects

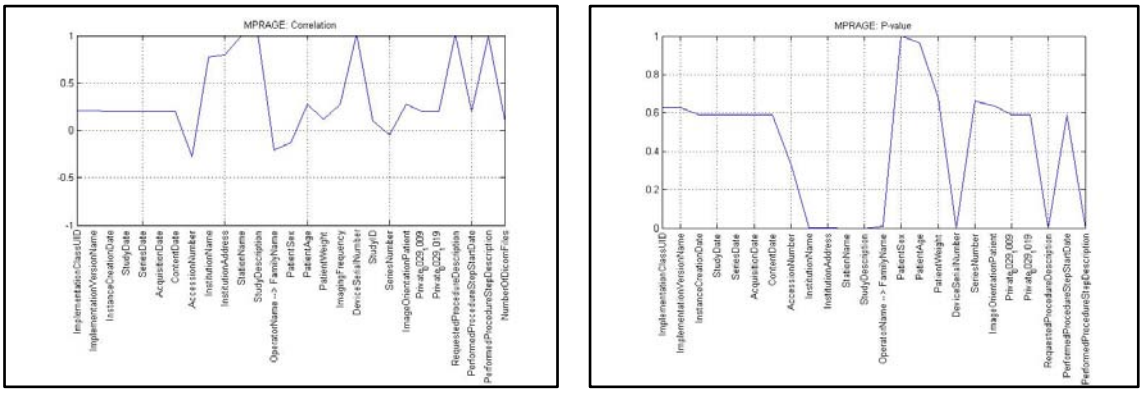
The version A from the CMSTrest database will give us information about possible site-specific effects, i.e., whether asignificant relationship is present in this database and the values taken by their parameters.

The Pearson Correlation has been computed for the *IntMat*, calculated using the DICOM checker, from the extracteddataset and his *Labels* vector. This one takes the value 0, if the subject were scanned in HUG, and 1, if the subject were scanned in BBL. The process is repeated for both protocols.



(a) **(b)**

Figure 2.1. Correlation on acquisition site for fMRI protocol. **(a)** Correlation. **(b)** P-value.



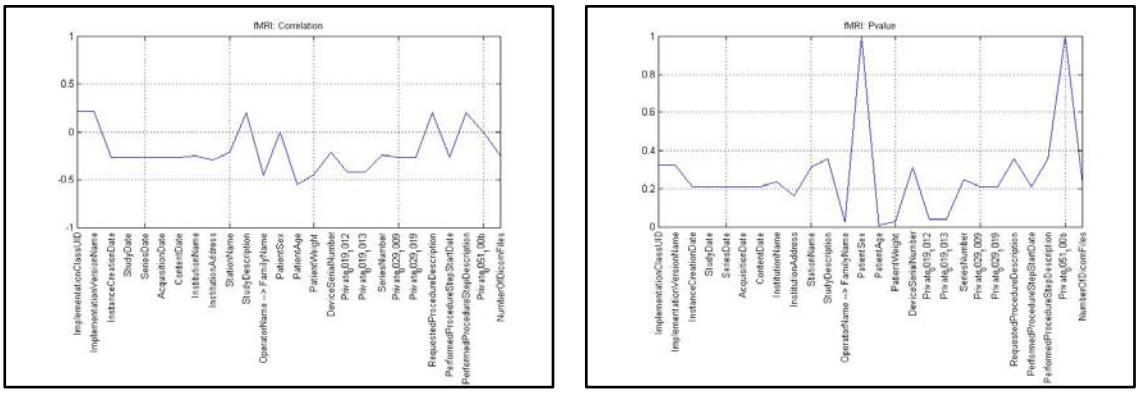
(a) **(b)**

Figure 2.2. Correlation on acquisition site for MPRAGE protocol. **(a)** Correlation. **(b)** P-value.

The correlation and p-value on fMRI and MPRAGE are very similar (Figure 2.1 and Figure 2.2), i.e., it only changes the protocol. The number of significant fields is 25 for fMRI and 27 for MPRAGE. These fields are more or less similar in both protocol but neither of them indicates a notable correlation according to acquisition-site.

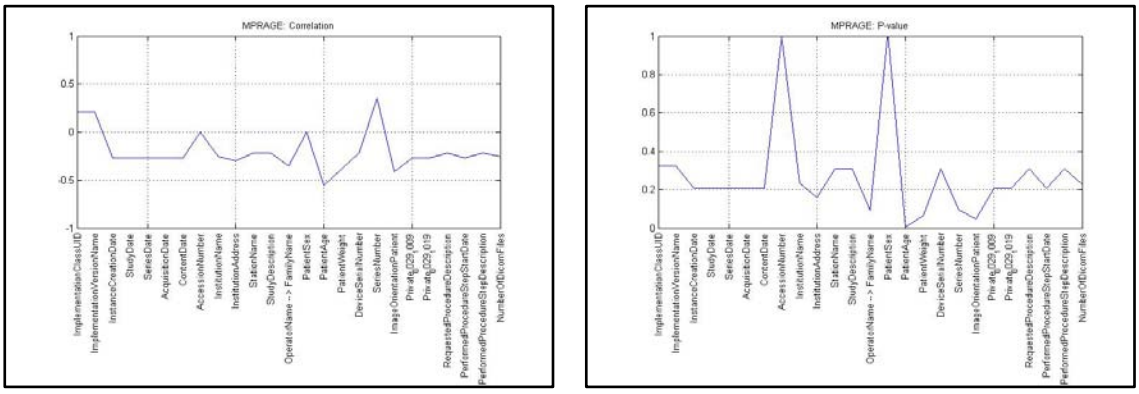
2.4. 2. Group-specific changes

In this case, we want to study the scanner's influence in the YC, EC or MCI groups. For this, we have 8 subjects per group classified in 4 men and 4 women. Labels vector takes the value 1 if the patient belongs to YC group, the value 2 if he is a member of EC group and 3 for MCI.



(a) **(b)**

Figure 2.3. Correlation on group-specific for fMRI protocol. **(a)** Correlation. **(b)** P-value.



(a) **(b)**

Figure 2.4. Correlation on group-specific for MPRAGE protocol. **(a)** Correlation. **(b)** P-value.

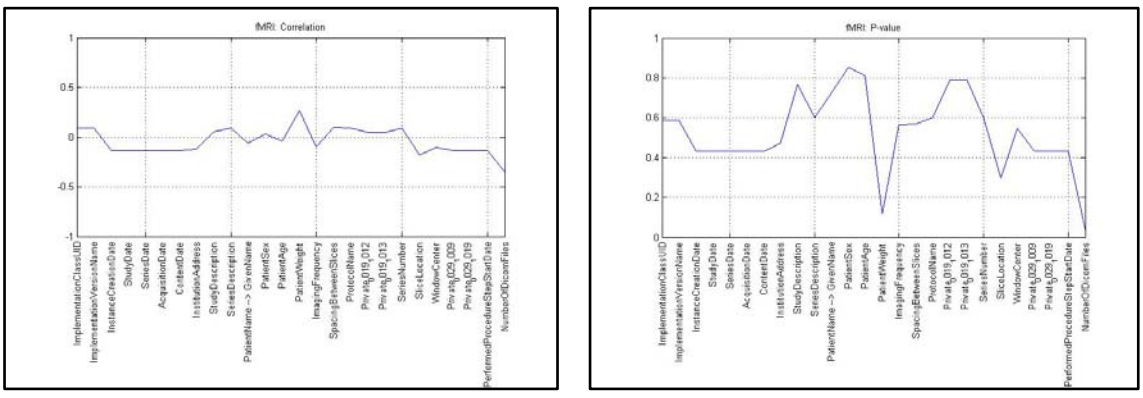
DICOM checker gives us 26 significant fields for fMRI protocol and 25 for MPRAGE. The *OperatorName* and *PatientAge* fields have an important correlation and this is rejected by their p-values (Figure 2.3 (b)). *OperartorName* field does not give a relevant information while the correlation obtained on *PatientAge* is normal because the group is constituted mainly by subjects with an age range from 55 to 85 years old. In CHUVrest version A there are only two subjects with 25 and 29 years old.

By MPRAGE protocol (Figure 2.4), *PatientAge* has again an important correlation but, moreover, *ImageOrientationPatien*and *Patientweight*seem to be significant too.*ImageOrientationPatien* refers to the orientation of image plane acquired with respect to the patient so it does not induce influence on magnetic resonance imaging (MRI) test. *Patientweight* has a significant correlation by the same reason that *PatientAge*.

Finally, for the group-specific information, no existence of confounding factors was found.

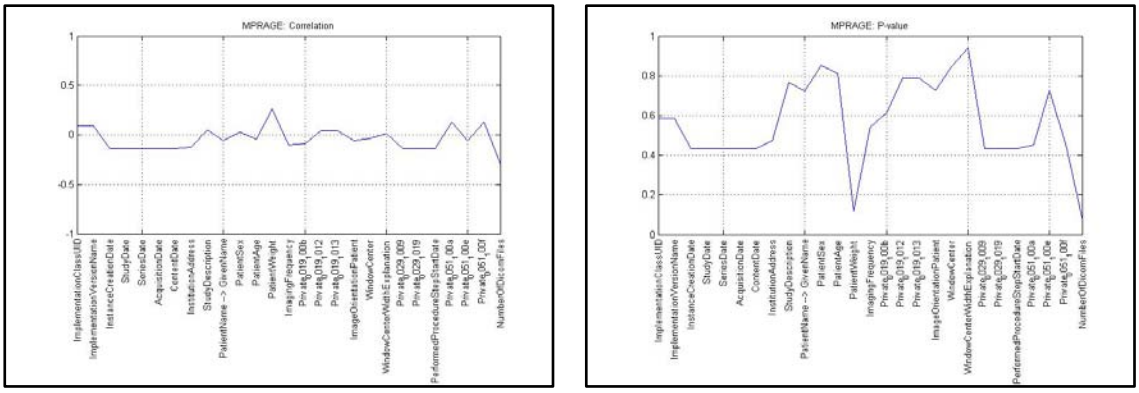
2.4. 3. Effects on classification accuracy

The CHUVrestdatabase allowsto know if there is a relationship between the connectivity matrix and the scanner parameters.The Labels takes the value 0, in the patients wrong classified, or 1, in otherwise. A subject is wrong classified when a patient is a control and the classification says that he is a sick, or vice versa. The Pearson Correlation has been computed for each protocol (MPRAGE and EPI).



(a) **(b)**

Figure 2.5. Correlation on accuracy classification for fMRI protocol. **(a)** Correlation. **(b)** P-value.



(a) **(b)**

Figure 2.6. Correlation on accuracy classification for MPRAGE protocol. **(a)** Correlation. **(b)** P-value.

According to the accuracy results obtainedof accuracy classification, neither fMRI (Figure 2.5) nor MPRAGE (Figure 2.6) protocol correlation give significant information. In both protocols,*NumberOfDicomFile* shows a significant correlation but this field only says thenumber of temporal filestaken by subject. Some temporal-volumescan be missed when scanners converted the image acquired into DICOM format. This field is non-significant in based on connectivity matrix.

2. 5. Conclusions

Firstly, we can observe that the fields of interest in different studies are similar. The private data elements are additional data defined by an implementer, to communicate information that is not contained in Standard Data Elements.

Here, we comment somefields between different studies. From among them all it is worth the trouble to emphasize a priori:

- 'ImplementationClassUID': Uniquely identifies the implementation which wrote this file and its content.
- 'ImplementationVersionName': Identifies a version for an Implementation Class UID using up to 16 characters.
- 'StudyDescription': Typically this field is used for a short description of the medical procedure such as you would see in CPT codes. In fact Study Description is a non-coded analogy for Procedure Code Sequence (0008,1032) where the code meaning element of the sequence item (0008,0104) is also 64 characters.

- ‘StationName’: User defined name identifying themachine that produced the compositeinstances.
- 'ImagingFrequency': Precision frequency in MHz of the nucleus being addressed.
- 'Private_0019_1012' and 'Private_0019_1013': These fields reference to series contract in the acquisition.
- 'SpacingBetweenSlices': Defines the position of the following slices (frames) in the pixel element. Each subsequent (frame) slice from the pixel element is parallel to the first and positioned along a normal to the first slice. The normal (vector) is defined as a cross product of the row vector with the column vector of the image slice.
- 'SliceLocation': Relative position of the image plane expressed in mm. See C.7.6.2.1.2 for further explanation.
- 'SpacingBetweenSlices': Distance between slices.
- 'ImageOrientationPatient': The image plane shall be specified by two values that designate the anatomical direction of the positive row axis (left to right) and the positive column axis (top to bottom). The first entry is the direction of the rows, given by the direction of the last pixel in the first row from the first pixel in that row.Image Orientation specifies the direction cosines of the first row and the first column with respect to the patient.
- 'Private_0019_100b': It is referred to cell spacing in the acquisition.
- 'RequestedProcedureDescription': Identifier that identifies the RequestedProcedure in the Imaging Service Request.Required if Sequence Item is present.
- 'DeviceSerialNumber': Manufacturer’s serial number of the equipment that produced the compositeinstances.

In view of all results we cannot say that there is some influence between scanners and different classifications. In each study is appreciated that the graphic results for both protocols are similar. This means the fields of interest and their values are usually the same regardless of protocols for one study.

In the first study (CMSTrest version A)occurs the same than in the second study, it means there are not any significant values which give relevant information. We only

find some values with a high level correlation but this is not important for us. This means that the acquisition parameters stored in DICOM files are not significantly different between sites, and suggests that multi-centric studies are possible.

In the second study (CMSTrest version B), where we seek the relationship between acquisition parameters and subject groups, we find interesting information a priori in:‘PatientAge’, 'ImageOrientationPatient', ‘Private_0029_1009’, ‘Private_0029_0019’, ‘PatientWeight’. The private fields are fields used by designers and it does not give us relevant information. In general, we did not find a significant influence.

For the first study (CHUVrest dataset) we have a regular correlation’s graph around zero so I do not have any significant value indicating a negative effect from acquisition protocol changes in time to classify between controls and patients.

In short, at the moment we cannot assure that scanners affect to different subject’s classification (CHUVrest) or that differences in acquisition protocol parameters are different between sites (CMSTrest version A).

3.Changes in images

The idea is to ascertain whether fMRI images coming from different scanning sites are combined to obtain larger databases, by attempting to build a scanner classifier. If we can find features in the images that let us predict which scanner an image was acquired from, then it is possible that combining data from different sites will be difficult because significant differences in some image statistics exist. A machine learning process is used to check if negative influenceexisted in the functional images data.

The confounding factors have been searched spatially and temporally, i.e. thetime-series mean intensity indicates if a spatial effect is present while the variance of the temporal information searches the temporal effects. Furthermore, the confounding factors effect on temporal connectivity matrix classification will be studied.

Many techniques have been proposed for statistically analysing fMRI data, and a variety of these are of general use. The aim of such analysis is to produce an image identifying the regions which show significant signal changes in response to the task. Each voxel is assigned a value dependent on the likelihood of the null hypothesis, that the observed signal changes can be explained purely by random variation in the data consistent with its variance, being false.

3. 1. Method

In order to find a possible negative influence from the scanners, it is necessary to have a largedatabase of functional images. For each fMRI test, a sequence of brain volumes is acquired. This is known as the time-series of one subject. The spatial and temporal information of each subject is analyzed with a machine learning model.

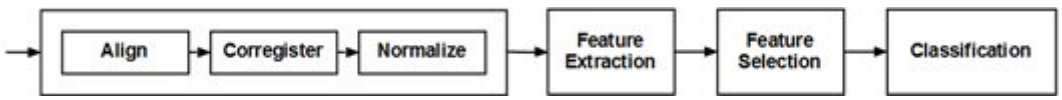


Figure 3.1. Machine learning blocks.

In medical processing image, beside sampling, denoising, etc., normally there are a number of different tools and methods for preprocessing; including: Realign, register and normalization.These are necessary steps to obtain voxels comparable among the different subjects. This will allow the executingof statistical analysis and knowing whether there are significant differences brain areas among subjects in resting-state.

3.1. 1. Databases

Two datasets have been used to study the existence of negative influence from different scanners on changes in intensity. The first of them is version A from theCMSTrest database for fMRI protocol.

The second dataset, *ADHDrest*, is extracted from "The ADHD – 200 Sample". *The ADHD-200 Sample is a grassroots initiative, dedicated to accelerating the scientific community's understanding of the neural basis of ADHD through the implementation of open data-sharing and discovery-based science. Towards this goal, we are pleased to announce the unrestricted public release of 776 resting-state fMRI and anatomical datasets aggregated across 8 independent imaging sites, 491 of which were obtained from typically developing individuals and 285 in children and adolescents with ADHD (ages: 7-21 years old). Accompanying phenotypic information includes: diagnostic status, dimensional ADHD symptom measures, age, sex, intelligence quotient (IQ) and lifetime medication status. Preliminary quality control assessments (usable vs. questionable) based upon visual time-series inspection are included for all resting state fMRI scans* [6].

ADHDrest database has been selected carefully. In the large database there are subjects scanned in various centers but only the patients scanned in "Oregon Health & Science University" (OHSU) and "Washington University in St. Louis" (WU)were taken. The final number of subjects taken was 48 where 24 (12 male and 12 female) are from OHSU and 24 (12 male and 12 female) from WU. The diagnostic for all of them is "Typically Developing Children" and the quality control of the session of scanning is cataloged as good. For each patients had around 200 volumes (size of 49·58·47 voxels by volume) of fMRI in resting-state.

All information by subject is processed and transformed to obtain oneoptimal distribution. So a big matrix wheresubjects are by columns andvoxels information by

rows will be built. This matrix willcontain the time-course average intensity or variance intensity information.

3.1. 2. Preprocesing

To study the spatial and temporal influence from scanners on CMSTrest version Adataset, the images set have been processed and transformed. For this, Statistical Parametric Mapping (SPM) has been used.The SPM is a powerful toolbox designed for analysis of brain imaging data sequences. It refers to the construction and assessment of spatially extended statistical processes used to test hypotheses about functional imaging data. The SPM is used to the analysis of fMRI, PET, SPECT, EEG and MEG.

SPM let realign, coregister and normalize easily. The realign process deals to remove movement-related artefacts in fMRI time-series. The headers are modified for each of the input images, such that they reflect the relative orientation of the data. The first image in the list specified by the user is used as a reference to which all subsequent scans are realigned. After realignment, the images are resliced such that they match the first image voxel for voxel. Apart from the images aligned this process creates one file with the mean of the resliced images [7].

Finally, the normalize module usually used forMRI, PET or SPECT images. The idea is normalize these images into a standard space defined by some ideal model or template image. If different subjects have different craniums sizes, the normalization will correct it and all brain structures of each patient practically represent the same voxels.

3.1. 3. Feature extraction

Feature extraction in pattern recognition is a special form of dimensionality reduction. If the input data is too large andredundant (much data, but not much information) thisis transformed into a reduced representation set of features. Transforming the input data into the set of features is called feature extraction. If the features extracted are carefully chosen, the relevant information from input data will be extracted in order to perform the desired task. Thus, the size of input data will be decreased.

On the other hand, feature patterns means characteristics measurable set like colour, texture, shape, etc., which can represent an image. A priori, a functional image is characterised for the intensity level of their voxels. So, the intensity level has been decided as our feature pattern.

For example, theCMSTrest dataset is formed by a time-series of 450 images and each volume has 64·64·21 voxels (38707200 voxels by subject in total over time). This is too much information by patient. It is necessary to reduce itbecause it is possible that many voxels have redundant information.

If the temporal information is reduced, the number of voxels could be optimally decreased by subjects. For this the time-series information could be calculated performing the average intensity, or computing the temporal variance. So, for CMSTrest version Adataset, the number of voxels could be decreased to 86016 voxelsby subject.

By subject, SPMconverts DICOM format into NIFI format and it gives us one 3D volume (64·64·21 voxels) with the temporal mean intensity. Then the volume is reshaped to vector column for building a matrix where their columns are the observations and their rows are subjects. This is called *IntensityMean* matrix.

The temporal information is analysed calculating the variance of the time-series. First,the temporal mean intensity volumeis assessed for one subject and then the subtraction between each time-series volume and the mean intensity volume is performed. So, a new intensity time-series centred in zero is obtained for each subject. Finally, the variance of the new time-series is computed by subject and a temporal variance volume is obtained. This volume is reshaped as a column vector (86016x1). This process is repeated for all subjects and the *TimeVariance* matrix is built.*TimeVariance* matrix and *IntensityMean* matrix will be analysed carefully applying pattern recognition techniques.

3.1. 4. Feature selection

In feature selection, the idea is to find an attribute selection relevant for building robust learning models. The main goal is to decrease the size of data for the machine learning process and finding the data which represent better each group. Given a feature set $X=\{x_i \mid i=1..N\}$, the feature selection process tries to find a subset $X'=\{x_{i1},x_{i2},\ldots,x_{iM}\}$, with $M<N$, which represent better to set and let a correct classification.

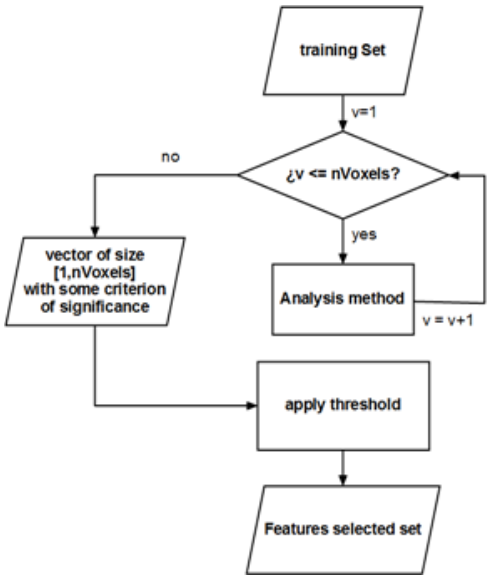


Diagram 3.1. Feature Selection.

Feature Selection requires an objective function to search and evaluating candidate subset. The objective functions are divided in two groups: Filters and wrapper.

- Filter objective function selects the subset using an independent criterion which the classifier. The objective function evaluates features sets by their information contented like interclass distance, statistical dependence or information theoretic measures.
- Wrapper objective function uses the same criterion in feature selection and classify of patterns. It means, the function objective is a pattern classifier which evaluates features sets by their predictive accuracy by statistical resampling or cross validation.

The features sets have been checked with three methods to analyse the relationship of voxels among different subjects: Kruskal-Wallis, Point-Biserial Correlation and Naïve Bayes. Naïve Bayes will be explained later in section 3.1. 5 but this is used as a wrapper model too.

- **Kruskal-Wallis (filter).**

Kruskal-Wallis test is a non-parametric method that it does not assume a normal population and it is one way analysis of variance (ANOVA) of input-data by ranking. It is used for testing equality of population median among groups. Fortwo populations

set,Kruskal-Wallis test gives us a p-values column vector, of size equal to number of voxels by subjects, which indicates if there are significant medians different at 5% for the voxels among patients. Thus, the voxels whose p-value is less than 5% will be the samples which represent the input data.

Lilliefors test was used for observing the probability density function (pdf)of the information from the voxels for subjects set. This test takes the default null hypothesis that the sample in vector of input comes from a distribution in the normal. The null hypothesis is reflected at the 5% significance level [2,3].On CMSTrest version Adataset, Lilliefors test indicated the 19.6% of sample had Gaussian distribution and 80.4% had a non-normal distribution. Thusto take Kruskal-Wallis test as one of the method applied in features selection process was decided.

- **Point-Biserial Correlation (filter).**

The point-biserial correlation coefficient (rpb) is a correlation coefficient used when one variable, X, is dichotomous. *A dichotomy is any splitting of a whole into exactly two non-overlapping parts, meaning it is a procedure in which a whole is divided into two parts or in half*[14].

The point-biserial correlation is equivalent to the Pearson correlation, that is, if we have one continuously measured variable X and a dichotomous variable Y, $r_{XY} = r_{pb}$. To calculate r_{pb}, assume that the dichotomous variable Y take the two values 0 and 1. Then the point-biserial correlation coefficient is calculated as follows [3, 5]:

$$R_{pb} = \frac{(M_a - M_b)}{S_n}\sqrt{\frac{N_a N_b}{n^2}}, \qquad \text{[ec. 3.1]}$$

$$S_n = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}, \qquad \text{[ec. 3.2]}$$

Where M_a is the arithmetic average from categories A, M_b is the arithmetic average from categories B, S_n is standard deviation used when you have data for every member of the population, N_a is the number of data from categories A, N_b is the number of data from categories B, n is the total samples size.

The features selection process assesses the point-biserial correlations coefficient of each feature for all subjects and the labels vector with two values (0 for subjects of one scanner and 1 for subjects of another scanner). That like, we know that connection there is between one feature and his categories.

In this method, the filter of features is not done with a significant threshold but we have chosen previously how many features are desired (e.g. D). The algorithms of feature selections gives the D best features, it means, the D features with higher Point-Biserial Correlations.

3.1. 5. Classification

Statistical classification has as goal to identify to which sub-population the new observation belongs to. The new sub-population is unknown and it is called testing set. For identify the origin of testing set data is had a training set data which contains observations of sub-population known.

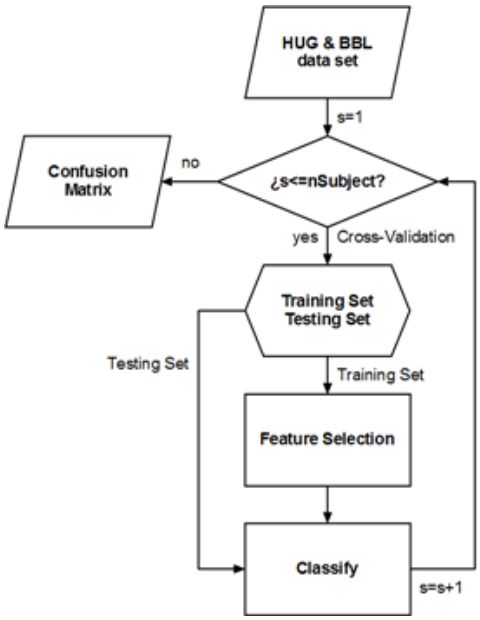


Diagram 3.2. Step to classify.

For assessing the origin of testing set data is used the cross-validation technique. This, in one round partitions the population in two sub-sets: training set and testing set. The idea is to analyse the training set and then validating the analysis on the testing set. The result of the classification will be showed in a confusion matrix which contains information about actual and predicted classifications done by a classification system.

Performance of such systems is commonly evaluated using the data in the matrix. The following table shows the confusion matrix for a two class classifier.

		Actual	
		Negative	**Positive**
Predicted	***Negative***	A	B
	Positive	C	D

Table 3.1. Confusion matrix.

a is the number of correct predictions that an instance is negative, ***b*** is the number of incorrect predictions that an instance is positive, ***c*** is the number of incorrect of predictions that an instance negative, and ***d*** is the number of correct predictions that an instance is positive.

The methods used to classify are Naïve Bayes, Support Vector Machine and Random Forest because they are models used for high dimensionality data, e.g., 86016 voxels are analyzed fromCMSTrest database version A or Bfor each subject. Naïve Bayes is an easy method of classification and useful if the population is small while Support Vector Machine is a model designed to avoid the overfitting. At last, Random Forest is a decision tree which tries to elude the variance of data.

Classifications in statistical population are evaluated according to a binomial confidence interval[10].This is a range of possible proportions which may contain the true proportion. If an experiment is repeated a fixed number of times and each trial has two possible answer, the true proportion have to be inside of the confidence interval. The range is based on the 95% of the population should be good validate.

- **Naïve Bayes.**

A Naïve Bayes (NB) classifier is a simple probabilistic classifier based on applying Bayes' theorem (from Bayesian statistics) with strong (naive) independence assumptions [11]. NB is a model trained for a supervised learning which receives a training population and it learns the model to validate the testing population.

An advantage of the naive Bayes classifier is that it requires a small amount of training data to estimate the parameters (means and variances of the variables) necessary for classification. Because independent variables are assumed, only the variances of the variables for each class need to be determined and not the entire covariance matrix.

NB is also used as wrapper technique, Diagram 3.3 shows the block included in "Analysis Method" from Diagram 3.2 to explain how are selected the features more representatives. Inside of feature selection block is applied cross-validation. The voxel*v* from allthe training set (19 voxels) has been takenfor creating a new sub-training set (18 voxels) and sub-testing set (1 voxel). The sub-testing set is validated with Naïve Bayes. Finally, for the feature *v* is got a value between 0 and 100 in percentage indicating how well has been classified the different sub-testing sets in each fold.A feature with value equal to 100 indicates which it classifies very well and 0 which it classifies very bad.

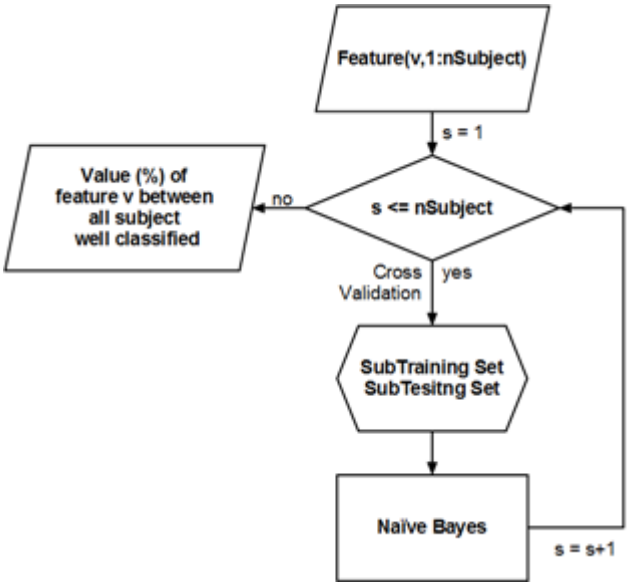


Diagram 3.3. Block Analysis Method for feature selection with Naïve Bayes.

- **Support Vector Machine.**

The Support Vector Machine (SVM) is a supervised learning methods used for classification and regression analysis and it is very useful on non-lineal problem.The SVM takes two different categories and training algorithm build a model that assigns a new sample into one category or the other. The learning statistical theory establishes two important characteristic which a learning algorithm must have:

- Ability to learn a rule which classifies rightly the highest number of observations in the training set.
- Ability to classify rightly the new observations with that rule.

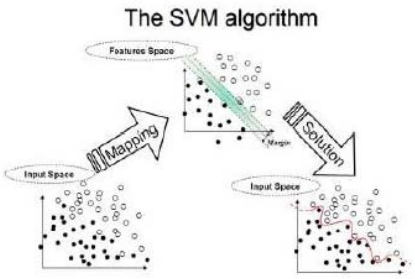


Figure 3.2. The SVM algorithm [12].

The classification consists in draw a lineal boundary which divides both samples categories. In non-lineal SVM, the idea is to define a new function which builds a dimensional space where the categories are separated by one lineal hyperplane.

The lineal hyperplane is chose to maximize the marginexistent between the samples of different categories. The *w* vector is a normal vector and it is perpendicular to the hyperplane.

$$w = \sum \alpha_n \omega_n x_n \text{ ,[ec. 3.3]}$$

whereα_n are the support vector coefficients, ω_n indicates the categories labels and x_n are the support vector.

- **Random Forest.**

Random forest (or random forests) is an ensemble classifier that consists of many decision trees and outputs the class that is the mode of the class's output by individual trees [3]. The basic idea is based on:

- A large trees number is built without beingpruned.
- Trees are built with a similar large dataset done through bootstrap on original sample. So two things are got: Correcting the error of prediction given by specific selection from the dataset. Second, having for each tree an independent sample to estimate the error of classification.

- In each division of one node is selected a random variables set with a size specified and the selection of variables is limited to this sub-set. So there is a higher variability of trees and the dependence of the result with the precedent divisions is decrease.
- Random Forest establishes a ranking based on the importance of the variables over the prediction of the output.

3.1. 6. Classification evaluation

In machine learning, the most popular form of feature selection is cross-validation. It is an algorithm that adds the best feature or deletes the worst feature at each fold. Cross-Validation is a statistical method of evaluating and comparing learning algorithms which segments data into two groups: one used to learn a model and the other used to validate the model. In typical cross-validation, the training and validation sets must cross-over in successive rounds such that each data point has a chance of being validated against. The basic form of cross-validation is leave-one-out or k-fold cross-validation. Leave-one-out cross-validation makes one population as testing set and the others samples as training set, while k-fold cross-validation applies k-times the validation dividing the dataset in k-samples.

The concept of overfitting is important in machine learning. This generally occurs when a model is excessively complex, such as having too many parameters relative to the number of observations. A learning algorithm is assumed to reach a state where it will also be able to predict the correct output for other examples, thus generalizing to situations not presented during training. However, especially in cases where learning was performed too long or where training examples are rare, the learner may adjust to very specific features of the training data, that have no causal relationship to the target function. In this process of overfitting, the performance on the training examples still increases while the performance on unseen data becomes worst.

3. 2. Results

Both databases explained in section 3.1. 1have been checked to study if there are significant changes in functional images. CMSTrest dataset allows studying the changes in images acquired in two scanners with the similar software version.

At the beginning, it could be supposed that theintensity mean intensity in two group of subject scanned in two different scanners is not similar. The coils from fMRI machines are not alike so the magnetic field will be different.Thus, each machine will draw with different mean intensity. If the effect on mean intensity is located spatially or temporally, it could be said that the effect could be generated by other elements like some of the antennas, etc.

It is important to know if there is both spatial influence and temporal influence. The spatial influence is analyzed computing the mean intensity from the time-series while the temporal influence will be analyzed with the variance from the time-series.

CMSTrest version Adataset was checked combining the different methods explained for feature selection and classification block. However, only the best features-classifier combinations are shown. After, some of them will be used to check the ADHDrestdataset.

3.2. 1. Result onCMSTrest version Adataset

TheCMSTrest version Adataset has been subjected to different models of machine learning based on the features selectors and classifiers explained earlier. This are combined to corroborate the results of classificationand assuring his non-dependency of the model used. It has been built both filters and one wrapper model.

The confidence interval forCMSTrest version Adataset says that the percentage of good decision has to be higher than 75%. If this percentage is lower, the lower bound on classification accuracy is random (below 50%) and it may not be said that any spatial or temporal effects exists.

- **Feature selection with Kruskal-Wallis and Naïve Bayes classifier (KW-NB).**

The *intensityMean* matrix fromCMSTrest version Adataset has been checked byKruskal-Wallis and it classified by Naïve Bayes. In order to select the best features, a cross-validation approach was used on the training set. Then, the testing set is classified using these features. The number of feature extracted from the training set depending on significant threshold used, named *sigThreshold*.

To compute the threshold it must be taken in account that there are 86.016 voxels by subject. The typical significant threshold for the p-values is 0.05, but it is necessary to apply Bonferroni Correction. This correction is based on the idea that the statistical significance level (threshold) must be testing for no dependent or independent hypotheses on each dataset. Being α significance level rejecting the null hypothesis, the significance threshold has to be α/n. So, in our case, n is equal to 86.016 and the significance threshold would have to be 0.05/86.016 what is very severe. If we take a value of n less severe like 1000, the threshold will be equal to $5 \cdot 10^{-5}$ and we can assess the voxel where there is a significant difference among subjects.

Several thresholds have been taken to filter the non-significantvoxels. The features selected on the p-value vector given by Kruskal-Wallis for training set has been filtered with a threshold from 0.0005 to 1 with a size of step equal to 0.008. Thus the testing set has been classified based on different number of features from around 100 to 62122 voxels in mean. The remaining features to 68016 do not take in account because they are black voxel outside brain.

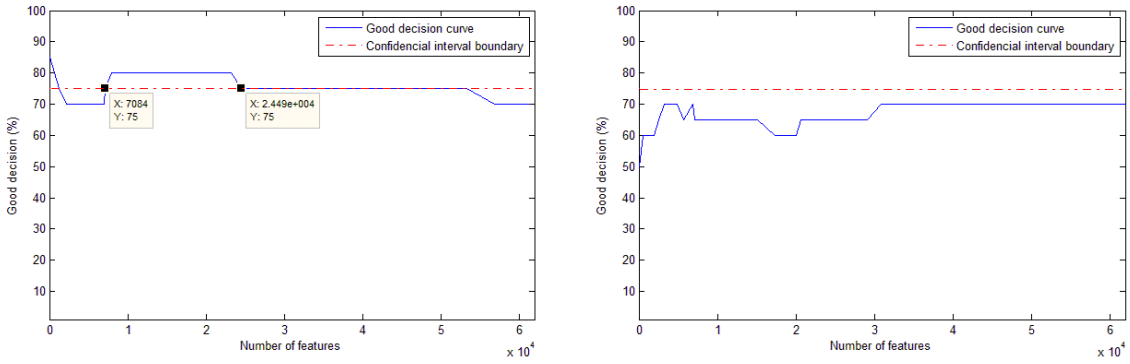


Figure 3.3. Good decision curve CMSTrest version A dataset according to KW-NB method. **(a).**On *IntensityMean* matrix. **(b)**On *TimeVariance* matrix.

The confidence interval is established in 75%. Thus, the spatial influence is guaranteed because the good decision curve on *IntensityMean* (Figure 3.3 (a)) is equal or higher than this interval for different number of features. On *TimeVariance* matrix (Figure 3.3 (b)), the classification is random for any quantity of features used, for this reason, temporal effects cannot be proved.

If a low threshold is taken like 0.0005 on *IntensityMean* matrix, an 85% of good decision is got obtainedCMSTrest version A dataset.

	BBL	**HUG**
BBL	8	1
HUG	2	9

(a)

	BBL										HUG									
Labels	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1
Predictions	1	0	0	0	0	0	0	0	1	0	1	1	1	1	0	1	1	1	1	1

(b)

Table 3.2.*IntensityMean* matrix CMSTrest-version Aclassification with KW-NB for a threshold of 0.0005. **(a)** Confusion matrix. **(b)**CMSTrest version A dataset labels and testing setvalidation.

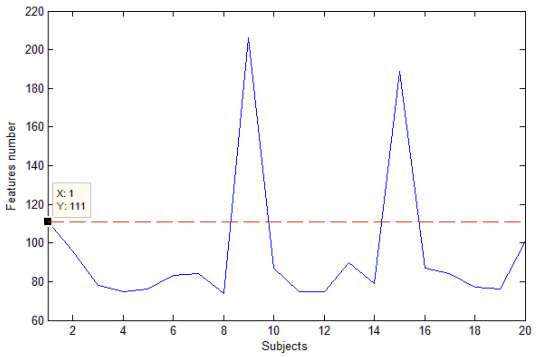


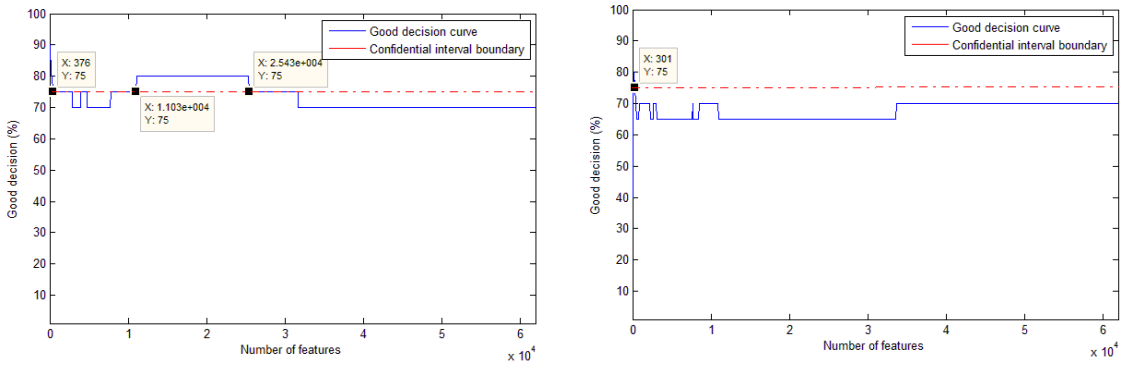
Figure 3.4. CMSTrest-version AIntesityMean features number over classification.Threshold equal to 0.0005.

Figure 3.4 shows the number of features used to validate the testing set. Subjects 1, 9 and 15 represented a substantially higher number of retained features. Interestingly, all these three subjects were uncorrectly classified (1, 9 and 15).This could be due to the high features number fed the classifier,the good decisionpercentage decreases(Figure 3.3 (a)).

- **Feature selection with Point Biserial Correlation Coefficient and Naïve Bayes classifier (PBC-NB).**

The *IntensityMean* matrix and *TimeVariance* matrix are checked withPBC to select the features on training set and then NB is used to classify the testing set. PBC let us select previously the number of features which are used to classify the testing set. The

algorithm has been tried with several thresholds varying from 1 to 62122 with a step size of 75.



(a) **(b)**

Figure 3.5. Good decision curve CMSTrest-version A datasetaccording to PBC-NB method.**(a).** On *IntensityMean* matrix. **(b)** On *TimeVariance* matrix.

PBC gives us the voxels ordered decreasingly according to their point-biserial coefficient (pbc). The temporal effects can be assured on 300 features (Figure 3.5 (b)) with the pbc higher, (D will be named for the number of features with pbc higher). In others cases the classification is random because it is lower than confidence interval.

There are two voxels set which guarantee the existence of a spatial effect (Figure 3.5 (a)). If the testing set is validated by the voxels included in [0, 376] or [11030, 20430] according to their pbc, the classification is higher than theconfidence interval. Thus, a spatial influence exists. If D equal to 100 features, an 85% and 80% good decision on *IntensityMean* and*TimeVariance* matrix is obtained respectively.

	BBL										HUG									
Labels	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1
Predictions (*IntensityMean*)	1	0	0	0	0	0	0	0	1	0	1	1	1	1	0	1	1	1	1	1
Prediction (*TimeVariance*)	0	0	0	0	0	0	0	0	0	1	1	1	0	1	1	1	1	1	0	1

(c)

Table 3.3. *IntensityMean*and*TimeVariance* matrixCMSTrest-version A classification. **(a)** Confusion matrix, D=100. **(b)**Labels and prediction classification.

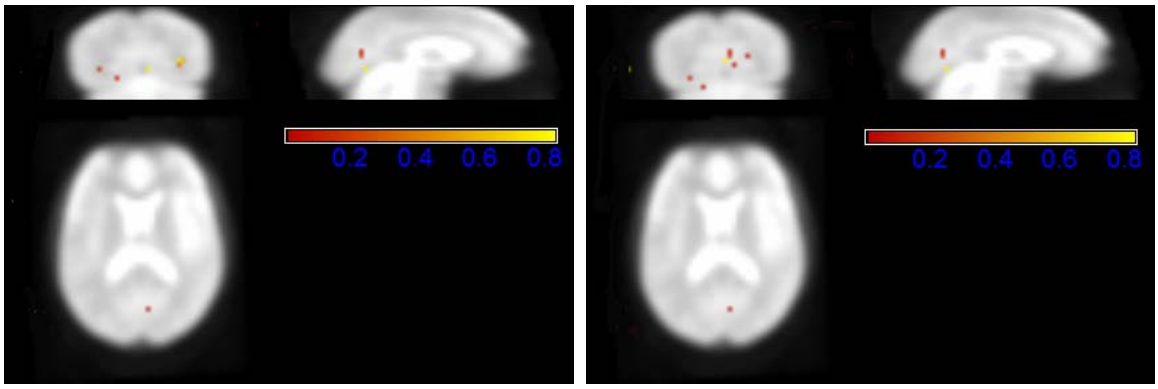
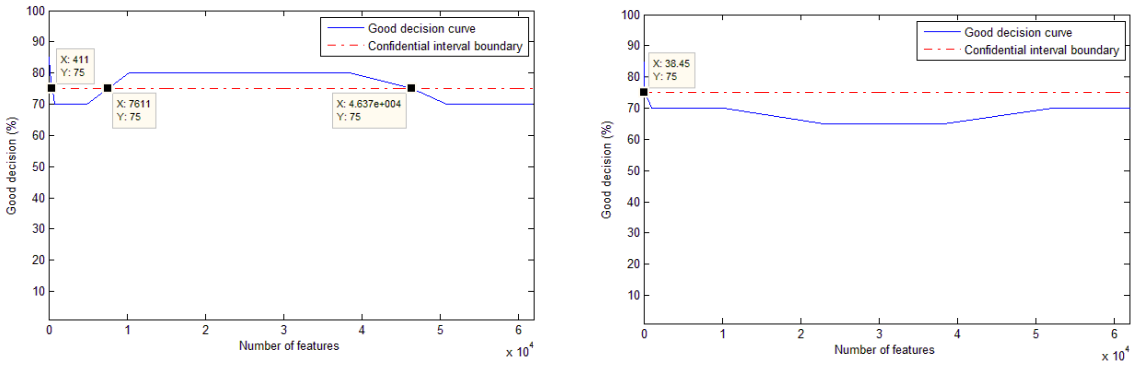


Image 3.1. Two sections for the stimulated voxels where is found a temporal influence. The color represented means the average pbc among the population sets. Yellow is pbc high and red low.

The temporal influence from scanners is important in a reduced voxels setbut these are distributed in small voxels sets inside the brain area (Image 3.1.).

- **Feature selection with Naïve Bayes and Naïve Bayes classifier (NB-NB).**

The wrapper machine learning method is characterized because the system of features selection and classification is the same. Naïve Bayes classifier used as features selector by feature among the training set. Then the mostrepresentativesvoxels are used to validate the testing set with Naïve Bayes again.



(a) **(b)**

Figure 3.6. Good decision CMSTrest-version curve according to NB-NB method. **(a)**On *IntensityMean* matrix. **(b)**On *TimeVariance* matrix.

IntensityMean and *TimeVariance* are checked for different thresholds (Figure 3.6). The spatial influence on intensity mean level is demonstrated if the n voxels which classify better are taken in the next ranges of features: [1, 411] and [7611, 46370]. The temporal influence cannot be guaranteed because the percentages of good decision only exceed the confidence boundary by the 43 voxels which represent better the training set.

If a threshold of 75 and 85 are taken on *IntensityMean*, a 70% and 80% of good decision is obtained.

	BBL	**HUG**
BBL	6	2
HUG	4	8

(a)

	BBL	**HUG**
BBL	7	1
HUG	3	9

(b)

	BBL										HUG									
Labels	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1
Pred. thresh = 75 (*IntensityMean*)	1	1	0	0	0	1	0	0	1	0	1	1	1	1	0	1	1	1	1	0
Pred. thresh = 80 (*IntensityMean*)	1	1	0	0	0	0	0	0	1	0	1	1	1	1	0	1	1	1	1	1

(b)

Table 3.4. *IntensityMean*matrix CMSTrest-version Aclassification for threshold of 75 and 80. **(a), (b)** Confusion matrix for the threshold of 75 and 80 respectively. **(b)** Labelsand prediction classification.

- **Feature selection with Point-Biserial correlation coefficient and Support Vector Machine classifier (PBC-SVM).**

The dataset has been analyzed with PBC as features selector and SVM as classifier. PBC-SVM also is used to check *IntensityMean* and *TimeVariance* matrix with several thresholds. In Figure 3.7, the spatial influence appears for a number of features low while the results for temporal information does not assure negative influence.

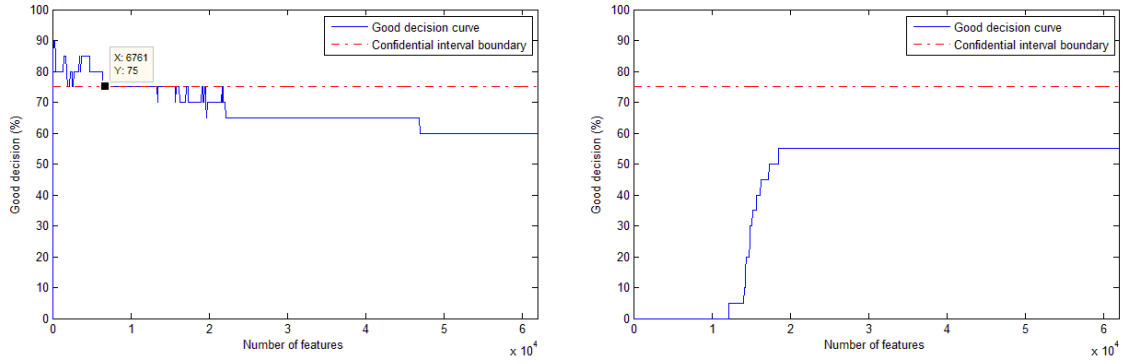


Figure 3.7. Good decision CMSTrest-version A curve according to PBC-SVM method. **(a)** On *IntensityMean* matrix. **(b)** On *TimeVariance* matrix.

- **Feature selection with Point-Biserial correlation coefficient and Random Forest classifier PBC-RF.**

PBC feature selector and Random Forest classifier are tried with various thresholds like the earlier models and the result are similar. Spatially (on *IntensityMean*, Figure 3.8) there is an effect from the scanner while the temporal influence is doubtful.

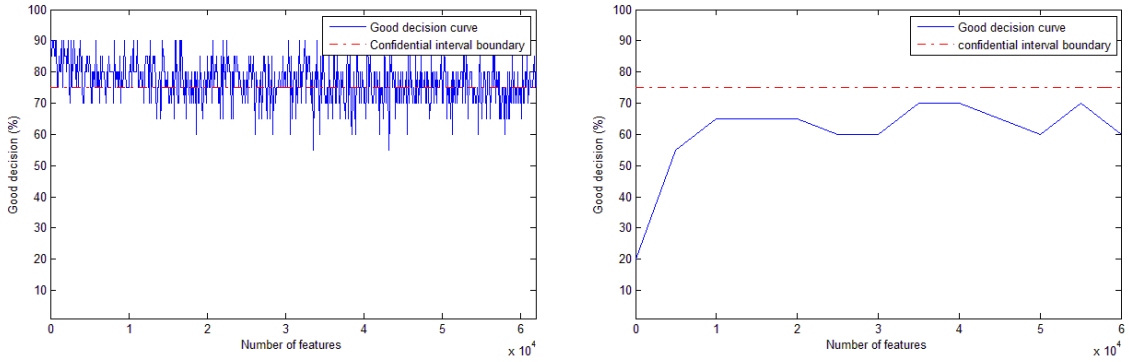
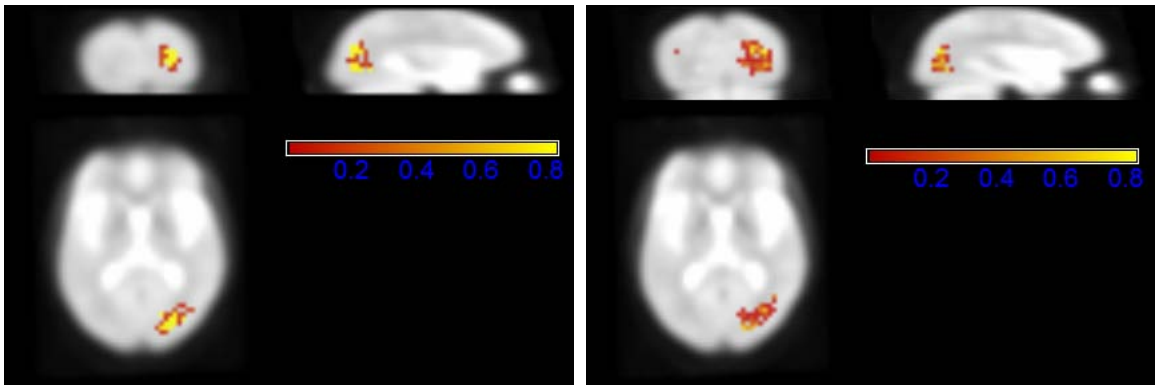


Figure 3.8. Good decision CMSTrest curve according to PBC-RF method. **(a)** On *IntensityMean* matrix. **(b)** On *TimeVariance* matrix.

- **Conclusions.**

For the different models of classification, the spatial effect has been demonstrated while the temporal effect cannot be assured. For all models, the voxels where there was a possible spatial effect were checked. These voxels were always located in Occipital, Parietal and a small part of the leftTemporal brainarea (Image 3.2.).

The analysis of the spatial effect is searched on the mean intensity of the time-series. Normally, the mean intensity from two scanners can be different. If the subjects from the CMSTrest version Adatabase are in resting state, the information in their voxel would have to be similar. Thus theaverage difference of intensity would also have to be similar in all voxels. So if there are voxels where the mean intensity is more significant than another, it can affirm that a negative effect exists. Besides, if this effect is located spatially, the influence from the scanners is clear.



(a) **(b)**

Image 3.2. Brain area where is found spatial effects from the scanners on CMSTrestversion A dataset.The yellow are voxels more different among subjects and lower difference in red.**(a)** Average weight vector, w (ec. 3.1), of decision boundary across folds of cross-validation by PBC-SVM. **(b)**Average intensity subtraction among CMSTrest according to PBC-SVM.

3.2. 2. Result on OHSU and WU dataset

For the ADHDrest database, all the previously explained preprocessing steps (realigning, corregistration, normalization) were not neccesary. All Nii files given by this web had been pre-processed and the noise of themhas been also removed.

TimeVariance and IntesityMean matrix were built like forCMSTrest version Adatabase. According to confidence interval, the percentage of good decision has to be higher than 67.7% for the lower bound of the confidence interval to be above 50%.

The spatial (on *IntensityMean*) and temporal influence (on *TimeVariance*) of the scanners ADHDrest dataset has been checked on PBC-SVM for multiples thresholds. In Figure 3.3, the percentages of good decision aresuperior toconfidence interval boundary, 67.7%, both spatial (Figure 3.3 (a)) and temporal (Figure 3.3 (b)) information. Thus, the classification is higher than random.

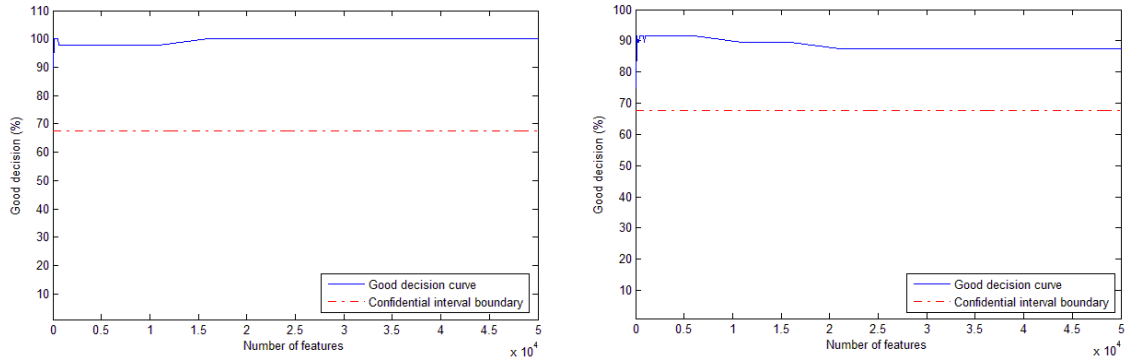


Figure 3.9. Good decision ADHDrest curve according to PBC-SVM method. **(a)** On *IntensityMean* matrix. **(b)** On *TimeVariance* matrix.

For a threshold of 100 the confusion matrix are shown in Table 3.5. The percentage of good decision for the mean intensity and temporal variance is 95% and 90% respectively.In Figure 3.10, it is pictured the mean intensity for one of significant voxels on *IntensityMean* matrix. As it can be seen, the average is different to both categories.

	WU	**OHSU**
WU	24	1
OHSB	0	23

(a)

	WU	**OHSU**
WU	21	1
OHSB	3	23

(b)

Table 3.5. ADHDrest dataset confusion matrix with PBC-SVM, D = 1000 features. **(a)** On*IntensistyMean*. **(b)** On *TimeVariance*.

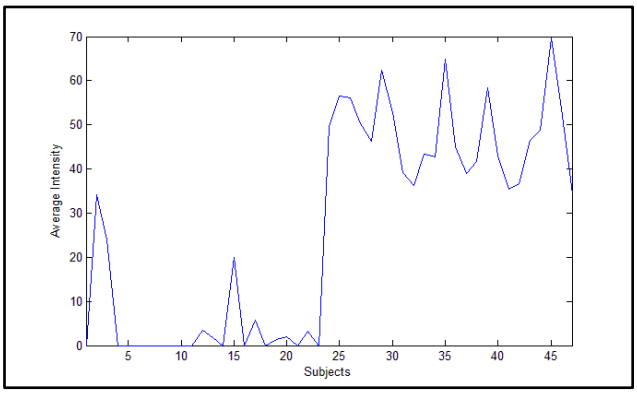
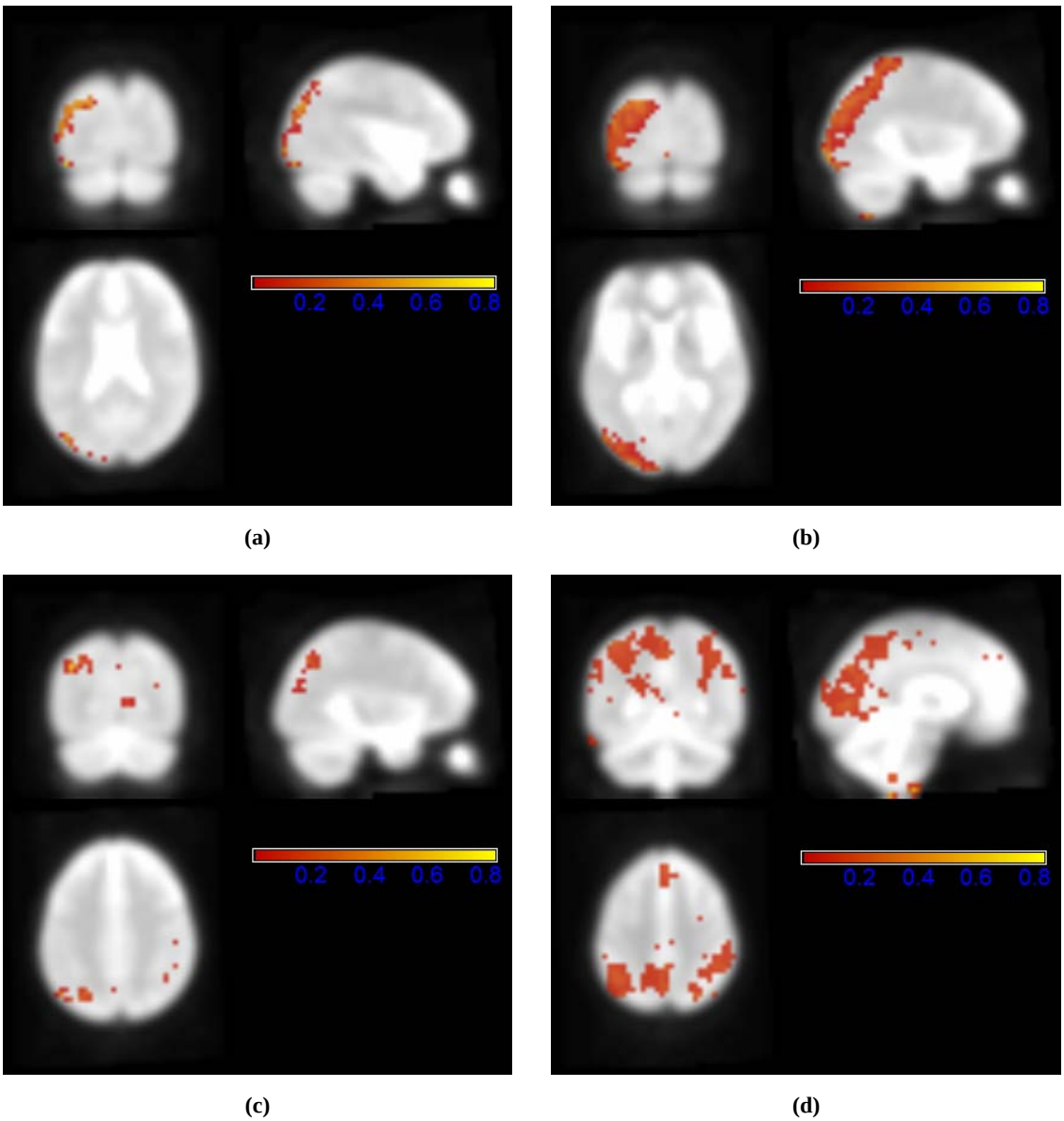


Figure 3.10. Mean intensity for one of significant voxels on *IntensityMean* matrix from the ADHDrest dataset.

The 100 voxels with higher pbc on *IntensityMean* (Image 3.3 (a))and *TimeVariance* (Image 3.3 (b)) matrix are pictured in Image 3.3. Again, the spatial and temporal effects are very located. The spatial effect is found on Parietal and Occipital from left brain area and the temporal effect is on both cerebral hemispheres.



(a) **(b)**

(c) **(d)**

Image 3.3. Significant voxels where there arespatial effects for the 100 features**(a)** with pbc higher and 1000 pbc higher **(b)**, and temporal influence for 100 **(c)** and 1000 **(d)** features with the pbc higher.In these images is pictured the mean weight vector (w) of decision boundary across folds of cross-validation.

The temporal gradient of temporal time-series is assessed.A gradient applied on temporalintensity indicates spatial effects regardless the average intensity from the scanners. PBC-SVM algorithm was applied to verify the existence of non-desired spatial effects.

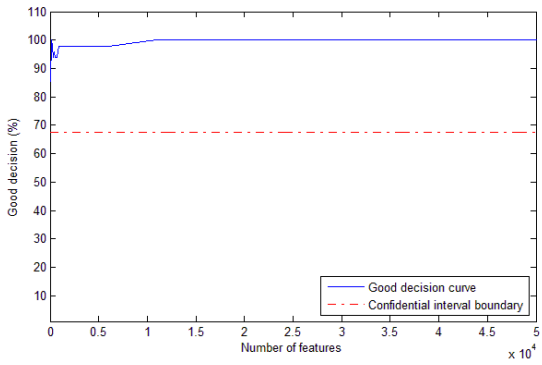


Figure 3.11. Good decision ADHDrest dataset curve according to temporal gradient if PBC-SVM algorithm is checked by different features number.

1000 voxels with the highest pbc are taken to validate the testing set (PBC-SVM method) and the good decision percentage is 97.9 % (upper than 67.7 %). These voxels appear distributed on the cerebral area but some of them are structured (Image 3.4). On the left brain hemisphere a small structure is located. Thus, a spatial effect is rejected.

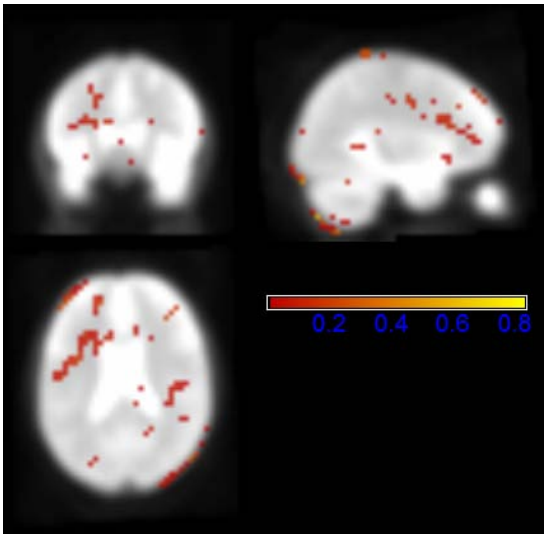


Image 3.4. Spatial effect on temporal gradient assessed by PBC-SVM for ADHDrest dataset. The red colour mean small differences whileyellow are the biggest differences among said dataset.

3.2. 3. Connectivity matrix

The connectivity matrix has been assessedfrom the regional average timecourses [5] on ADHDrest dataset. This is a symmetric matrix whose values represent the importance connections among brain regions ADHDrest.

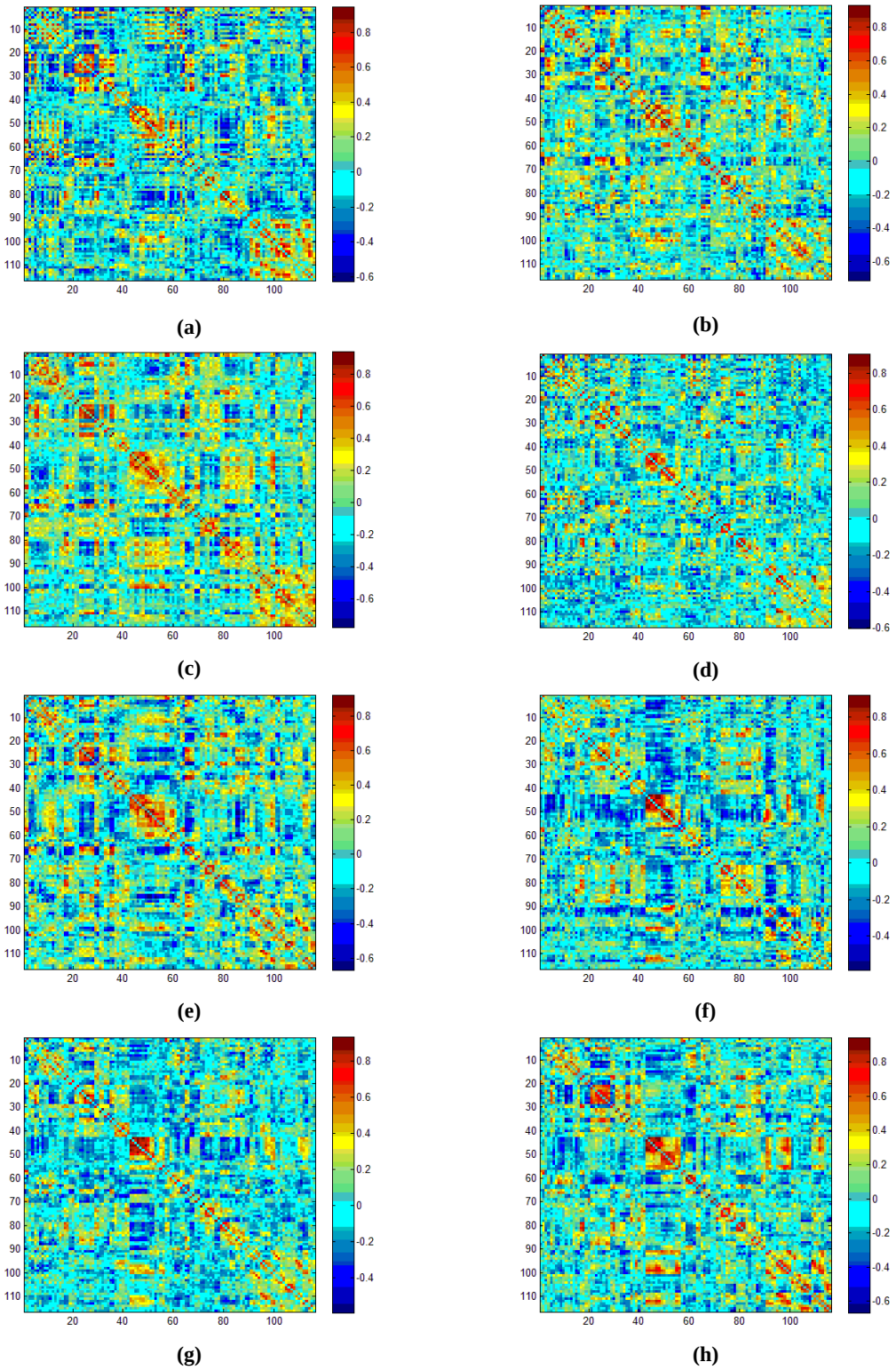


Figure 3.12. Connectivity matrix shape from eight ADHDrest subjects: 4 from the OHSU (**(a)**, **(b)**, **(c)** and **(d)**) and another from the WU (**(e)**, **(f)**, **(g)** and **(h)**).Each element of connectivity matrix indicates the connection level between two brain regions. Connection level is represented in red-blue range: strong in red and weak connectionsin blue.

The ADHDrest connectivity matrix has been checked by PBC-SVM (Figure 3.13). The good decision percentage to validate the testing set is always higher than confidence interval boundary (67.7%). So, the classification is not random. This means the connections among WU and OHSU regions subjects are not similar.

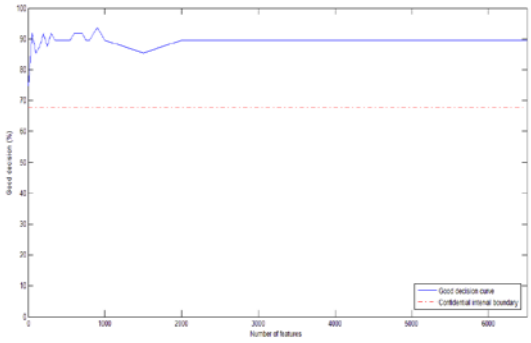


Figure 3.13. "good decision"-"Number features" curve on ADHDrest connectivity matrix and it analyzed with PBC-SVM method.

The regions with different connectivity among the ADHDrest subjects appear structured (Figure 3.14). Reddish regions structured (Figure 3.14 (a)) show a blue shade on Figure 3.14 (b), thus their p-values are lower than 0.05 and the ADHDrest connectivity significant differences are rejected.

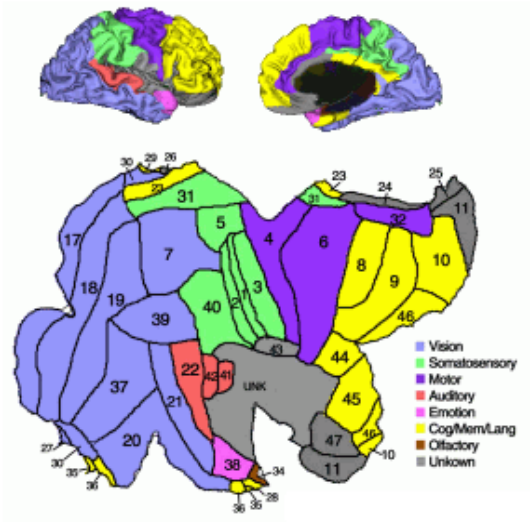


Image 3.5. Brodmann's Areas approximated in the human brain [20].

The 29 (Left insula), 39 (left parahippocampalgyrus), 73 (left putamen). regions taken the highest values (Figure 3. 15), so the connections on ADHDrest subjects brains regions change importantly according to the scanning site.The fact that the affected regions are always in the left hemisphere may indicate a possible scanner-related effect.

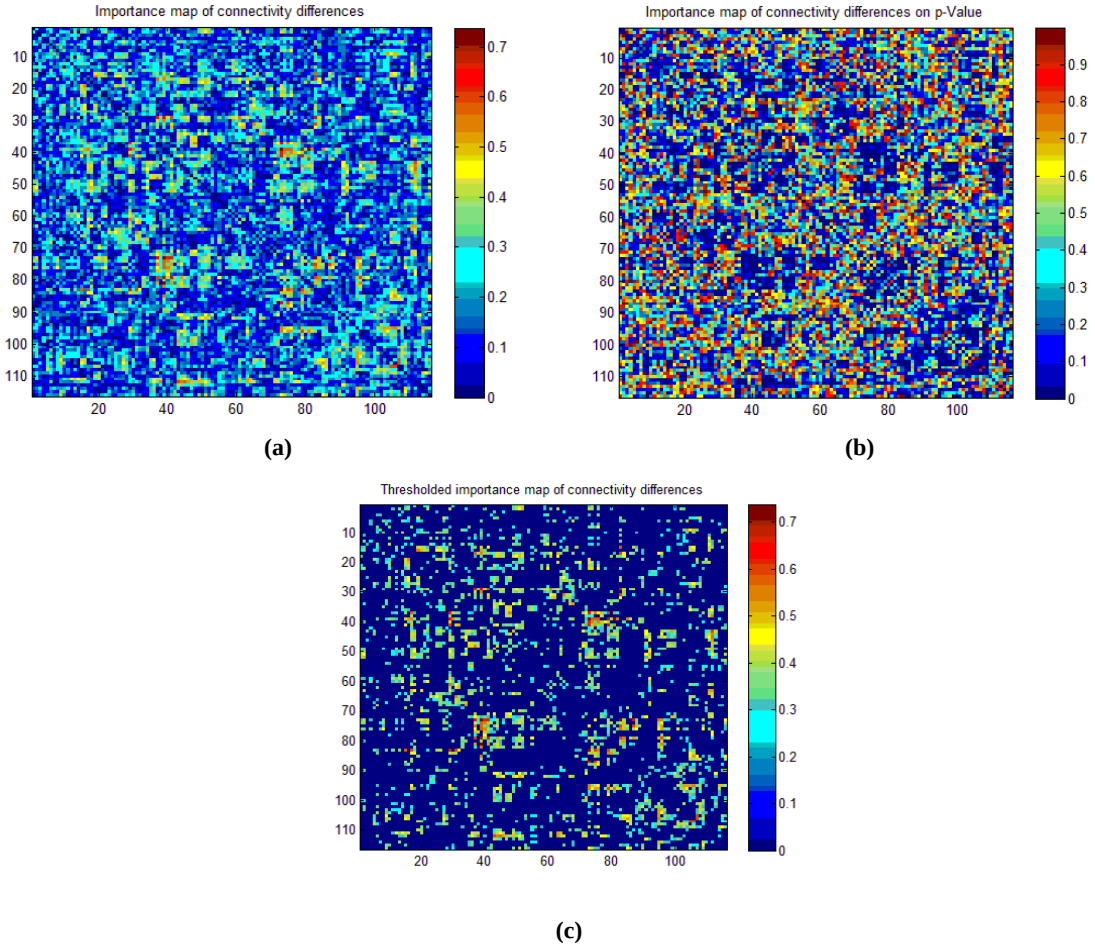


Figure 3.14. connectivity differences on ADHDrest dataset. **(a)**Importance map. (b) p-Values differences connectivity. **(c)** Important connectivity thresholded.

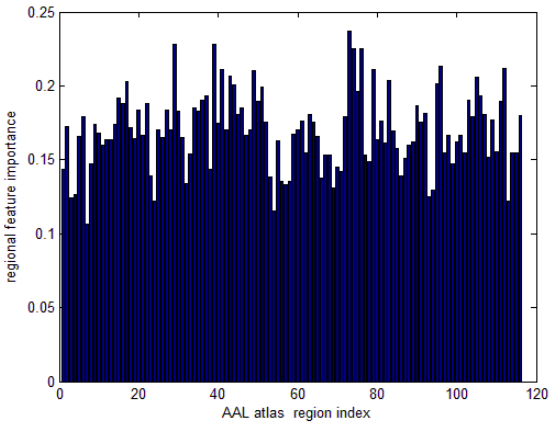


Figure 3. 15. (a) Regional feature importance connectivity on ADHDrest.

The differences among ADHDrest connectivity allow validating the testing set.It means difference connectivity regions are variably structured according to acquisition site.

3. 3. Conclusion

The changes the intensity searched through CMSTrest version A and ADHDrest dataset determine what spatial and temporal structures are existent. The average intensities between both scanners in each dataset (CMSTrest version A and ADHDrest) were different,but it was curious that significant regions was structured and located on region specific.

Using an intensity invariant technique could determine if this confounding factor based on changes on intensity is existent. The gradient on the mean intensity of the time-series could help us with this one and it was demonstrated there were structured regions different among ADHDrest dataset.

The temporal effectwas analyzed on time-variance of intensity. For this reason, the variance analysis gives spatial-temporal information.The temporal confounding factors are non-assured on CMSTrest dataset while on ADHDrest presented a significant structure located on Occipital and Parietal regions too.

Finally the ADHDrest connectivity matrix was evaluated and significant differences were appreciated on the brain left hemisphere. So, we could finish saying scanners effects are influencing on fMRI images from CMSTrest version A and ADHDrest datasets.

4. Conclusions and futures steps

4. 1. Conclusions

The neuroimaging is a young field that it advances quickly in the last years. The development of functional marker could diagnose difficult mental illness of recognizing. For making an optimal functional market is necessary to develop it with a huge quantity fMRI images. The fMRI tests for developing the functional market are taken from different scanners. This can induce a confounding factor which hampers to make good functional markers.

A confounding factor can be induced by acquisition changes realized by operator or changes of intensity on images acquired. The DICOM parameters were analyzed but these do not gave relevant information. There were not significant differences on DICOM parameters from the CMSTrest or CHUVrest dataset. On the other hand, machine learning allows analyzing intensity changes on ADHDrest and CMST version A datasets. The result established spatial and temporal effects structured existed on intensity information.

In an ideal case, the resting-state connectivity matrix should be similar or alike among subjects. The ADHDrest connectivity matrix was checked with machine learning algorithms and latter can validate the testing set correctly. So, functional characteristic dependent on their scanner have to exist and intensity changes could be the cause. A connectivity matrix is useful to develop functional markers but confounding factors do not allow realizing satisfactorily said task.

Inferring how confounding factors effect on fMRI images, standardizing the acquisition parameters, keeping the equipment the working in order, etc. could prevent this negative effects to develop good functional markers.

4. 2. Futures steps

The future objective is to remove the confounding factors. These can appear on images randomly and learning a generic method to remove them is a difficult task. For

developing a functional marker from huge quantity of images, the first step is to learn how the negative effect is and then this could be removed.

Negative effects can be corrected combining a multiplicative (a) and additive (b) factor.

$$x = ax + b$$

b can remove the average intensity differences while the multiplicative factor will consider others cases. The correction will try each voxels independently, thus the non-affected voxels by negative influences will be not altered.

Bibliography

[1] fivan. http://fivanblog.blogspot.com/2010/02/comunicarse-con-el-pensamiento.html. June, 2011.

[2] Daniel Y. Kimberg. “A functional connectivity primer”. June, 2011.

http://www.ncrrn.org/papers/methodology_papers/connectivityprimer.pdf. June 2011.

[3] mindpowercoach. http://mindpowercoach.net/the-brain-decoding-transmitting-beliefs. June, 2011.

[4] Francisco Pereira et Al. “Machine learning classifiers and fMRI: A tutorial overview.” Int. J. Neuroimagen, 2008.

[5] Jonas Richiardi et Al. “Decoding brain states from fMRI connectivity graphs”. Int. J. NeuroImagen, 2010.

[6] The ADHD – 200 Sample. http://fcon_1000.projects.nitrc.org/indi/adhd200/index.html.June, 2011.

[7] Statistical Parametric Mapping Manual. http://www.fil.ion.ucl.ac.uk/spm/doc/manual.pdf. March, 2011.

[8]. Wikipedia. The free encyclopaedia. http://en.wikipedia.org/wiki/Main_Page. March, 2011.

[9] Tom Michel. “Machine Learning”. McGraw Hill.

[10] Brown, Lawrence D., Cai, T. Tony, Dasgupta, Anirban, 1999. Interval estimation for a binomial proportion. Stat. Sci. 16, 101-133.Edwin B. Wilson, Probable Inference, the Law of Succession, and Statistical Inference, Journal of the American Statistical Association, Vol. 22, No. 158 (Jun., 1927), pp. 209-212.

[11] Naive Bayes classifier. http://en.wikipedia.org/wiki/Naive_Bayes_classifier. Wikipedia, April 2011.

[12] Support Vector Machines. http://www.dtreg.com/svm.htm. May 2011.

[13] Cross-validation. http://en.wikipedia.org/wiki/Cross-validation_(statistics). Wikipedia, May 2011.

[14] Point-Biserial correlation coefficient. http://en.wikipedia.org/wiki/Point-biserial_correlation_coefficient. Wikipedia, June, 2011.

[15] Reporting activity or lesion location. http://imaging.mrc-cbu.cam.ac.uk/imaging/BrodmannAreas. June, 2011.

[16] Statistic toolbox of Matlab. http://www.mathworks.com/help/toolbox/stats. March, 2011

[17] Non-Linear Dinamic. http://www.nonlinear.com/support/progenesis/samespots/faq/pq-values.aspx. March, 2011.

[18] “Monografias”. http://www.monografias.com/trabajos16/manual-estadistica/manual-estadistica.shtml. April, 2011

[19] Columbia University Medical Center. http://www.fmri.org/fmri.htm. June, 2011.

[20] Digital Imaging and Communication in Medicine. http://medical.nema.org/. June, 2011.