



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



UNIVERSITAT
DE VALÈNCIA



GENERALITAT VALENCIANA
CONSELLERIA DE SANITAT

**ANONIMIZACIÓN Y CONTROL DE CALIDAD CENTRALIZADO
(CENTRALIZED QUALITY CONTROL - CQC) EN IMAGEN DE
RESONANCIA MAGNÉTICA**

Trabajo Final de Máster

Máster en Ingeniería Biomédica

Septiembre 2015

Alumno: Ángel Fernández-Cañada Vilata

Directores: Montserrat Robles Viejo

María de la Iglesia Vayá

Contenidos

1	Agradecimientos	4
2	Introducción	5
3	Motivaciones	5
4	Objetivos	6
5	Ambito y entorno del proyecto	6
5.1	Centro de Excelencia e Innovación Tecnológica de Bioimagen de la Conselleria de Sanitat (CEIB-CS)	7
5.2	Banco digital de Imagen Médica de la Comunidad Valenciana (BIMCV)	8
5.2.1	Estudio poblacional y Big Data	8
5.2.2	Euro-Bioimaging y BIMCV	8
5.2.3	Gestor de Ensayos y programas de investigación con Bioimagen para la I+D (GEBID)	10
6	Aplicaciones precedentes	11
6.1	C-PAC Configurable Pipeline for the Analysis of Connectomes	11
6.2	Data Processing & Analysis of Brain Imaging (DPABI)	13
6.3	Proyecto EMBARC	15
6.4	Dicom Browser	16
7	Software utilizado y extendido	17
7.1	Matlab	17
7.2	Octave	18
7.2.1	FMRIB Software Library (FSL)	18
7.3	Statistical Parametric Mapping (SPM)	20
7.3.1	MRICron y DCM2NII	20
8	Actualización de CEIBANON - Clinical Trial Processor (CTP)	21
9	Desarrollo del sistema de envío de datos desde los centros al BIMCV	23
10	Implementación de la pipeline de conversión de DICOM a NIFTI automatizada	24
11	Desarrollo del Control de Calidad Centralizado (CQC)	25
11.1	Características básicas de la imagen	26
11.2	Estimación del movimiento	27
11.2.1	Framewise Displacement (FD)	29
11.2.2	DVARS	32
11.3	Rango dinámico	34

11.4	Contraste	35
11.4.1	Relación señal ruido temporal y relación contraste ruido temporal	37
11.5	Estimación de la heterogeneidad del campo de polarización	40
11.6	Espacio K	42
12	Implementación de algoritmos y herramientas de anonimización	43
12.1	Anonimización de cabeceras DICOM	43
12.2	Anonimización sobre la propia imagen	45
12.2.1	Anonimización de ecografías	46
12.2.2	Eliminación de facciones faciales	49
13	Creación de sistema de presentación visual de resultados e integración en GEBID	51
14	Presentación en congresos y premios	52
15	Conclusiones	56
16	Trabajo futuro	57
17	Bibliografía	58
18	Acrónimos	59

1 Agradecimientos

Agradezco la oportunidad, la ayuda, dirección, conocimiento y el apoyo que me han ofrecido la **Dra. Montserrat Robles Viejo** como mi directora en la **Universitat Politècnica de València (UPV)** y la **Dra. María de la Iglesia Vayá** como mi codirectora en el **Centro de Excelencia e Innovación Tecnológica de Bioimagen de la Conselleria de Sanitat (CEIB-CS)**.

No me gustaría olvidar a nadie sin quien este proyecto no habría sido posible haberse llevado a cabo. En especial doy gracias al **Dr. José Vicente Manjón (UPV)** por su orientación y el uso de parte de su script para la corrección de heterogeneidad de campo, a **José María Salinas Serrano (Responsable de la Unidad de Informática Departamento de Salud del Hospital San Juan de Alicante)** por sentar en parte las bases de este trabajo y la ayuda de su tesis, a **Cameron Craddock y John Pellman (Child Mind Institute, New York City)** por sus conocimientos sobre control de calidad y orientación en el uso del **Configurable Pipeline for the Analysis of Connectomes (C-PAC)**, gracias también a **Yan Chao-Gan (Institute of Psychology, Chinese Academy of Sciences)** por su implementación del **Framewise Displacement (FD)** y su herramienta de trabajo **Data Processing & Analysis of Brain Imaging (DPABI)** y a **Jonathan D. Power (Washington University)** por su ayuda en la corrección de mi implementación del marcador **DVARS** en lenguaje **Matlab**. Tampoco quiero olvidar al **Oceanogràfic** y en especial a **Daniel García Parraga** por cedernos las ecografías de delfines usadas en la aplicación de anonimización de ecografías.

Para finalizar, agradezco a mi amigo y compañero del **(CEIB-CS) Jorge Isnardo Altamariano** sus conocimientos prestados en la gestión y creación de pipelines de **eXtensible Neuroimaging Archive Toolkit (XNAT)**, en el aprendizaje del lenguaje de programación Python y en la administración de los sistemas operativos Linux. Del mismo modo, agradezco a mi amigo Físico e Ingeniero Biomédico **Fernando Fort Fabregat** la ayuda prestada para la comprensión de los fenómenos físicos que dan lugar a la imagen médica y sin olvidar a mi familia y a todos los demás que han estado a mi lado siempre y durante el desarrollo de este trabajo final de máster.

2 Introducción

La imagen biomédica apareció en el campo de la ciencia hace 115 años con la primera radiografía obtenida por Wilhelm Conrad Röntgen. Por entonces, el descubrimiento fue aclamado como uno de los mayores logros tecnológicos de la humanidad.

En el siglo XXI, la imagen biomédica ha pasado a ser una de las disciplinas centrales de la medicina actual a medida que aumenta el número de las patologías que pueden ser diagnosticadas y tratadas con su ayuda, a la vez que se desarrollan sistemas cada vez más sofisticados para la obtención de imágenes médicas.

Uno de los problemas fundamentales con los que se encuentran los investigadores en el campo de las técnicas de imagen biomédica es la heterogeneidad y la mala accesibilidad a las infraestructuras existentes o incluso el desconocimiento de su existencia.

El **Centro de Excelencia e Innovación Tecnológica de Bioimagen de la Conselleria de Sanitat (CEIB-CS)**, <http://ceib.san.gva.es/>) da solución a este problema mediante la creación del **Banco digital de Imagen Médica de la Comunidad Valenciana (BIMCV)**. BIMCV es un banco de imágenes concebido como un repositorio de conocimiento orientado a favorecer los avances tecnológicos en imagen médica y proporcionar servicios de cobertura tecnológica para el soporte a proyectos I+D. Dentro de BIMCV podemos destacar la plataforma **Gestor de Ensayos y programas de investigación con Bioimagen para la I+D (GEBID)**, <http://ceib.bioinfo.cipf.es/xnat/>) la cual a groso modo funciona como un gestor y repositorio de imágenes de ensayos clínicos.

3 Motivaciones

Las soluciones presentadas en el punto anterior son el esqueleto en el que se basa este trabajo. Es por ello, que a semejanza de un cuerpo humano completo aún queda mucho trabajo por hacer. Almacenar la información y ofrecerla al mundo sólo es uno de los primeros pasos a llevar a cabo. Por un lado, los contenidos almacenados deben ser anonimizados pues es un derecho fundamental de las personas que cedieron los datos. Por otro lado, se debe dotar a los investigadores de los indicadores necesarios de calidad de las imágenes depositadas en el Banco pues no es posible hacer buena ciencia a partir de datos erróneos o de una distorsión de la realidad. Los datos erróneos son sorprendentemente comunes. Sin embargo, incluso los más graves pueden ser difíciles de detectar si no los buscamos cuidadosamente.

Otro problema a resolver es que aunque existen muchas herramientas en el mercado, los presupuestos en estos momentos son ajustados, por lo que es necesaria una nueva generación de herramientas libres y gratuitas. Hay que comentar que existen herramientas libres y gratuitas que iremos presentando no obstante no hay ninguna que cumpla todas nuestras necesidades.

Finalmente, el problema más grave en cuanto al control de calidad se refiere (en nuestro caso) es la inexistencia de herramientas integradas en GEBID.

4 Objetivos

El objetivo principal del presente proyecto es ampliar el conjunto de herramientas de **BIMCV** dentro de la plataforma **GEBID** y asegurar con mayor robustez el anonimato de los sujetos a los que pertenecen las imágenes almacenadas de la forma más eficiente posible.

Por ello, que se proponen los siguientes objetivos secundarios para alcanzar dicho objetivo:

- Primer objetivo: buscar, adaptar e implementar procedimientos de control de calidad de imagen médica orientados a la resonancia magnética tanto cuantitativos como cualitativos.
 - En el control de calidad cuantitativo, se crean marcadores cuyo resultado nace de cálculos y el tratamiento de la imagen. Su representación final puede ser tanto numérica como otra (por ejemplo un código de colores, etc.) . Con este objetivo se pretende:
 - * Obtener las características generales de la imagen.
 - * Buscar y adaptar los mejores marcadores de movimiento asociados a las imágenes de resonancia magnética 4D.
 - * Buscar y adaptar al sistema algún estimador de la heterogeneidad de campo para imágenes de resonancia magnética (**RMI**).
 - * Implementar indicadores del nivel ruido, tanto temporales y por vóxel como absolutos.
 - * Obtener marcadores de contraste lo más próximos a la percepción del ser humano.
 - Por otro lado, el control de calidad cualitativo, es el control de calidad tradicional basado en la percepción y experiencia humana para la detección de artefactos. Para ello se quiere:
 - * Crear procedimientos para la generación de una muestra en base a un conjunto de capturas de la adquisición original.
 - * Regenerar a partir de la imagen su correspondiente en el espacio-k.
- Segundo objetivo: Implementar mecanismos de anonimización que permitan asegurar el anonimato de las personas que cedieron las imágenes.
 - Crear un sistema para la anonimización de cabeceras dentro del estándar **DICOM (Digital Imaging and Communication in Medicine)** utilizando herramientas externas y propias, que permita una gran flexibilidad respecto a los cambios en el estándar.
 - Desarrollar scripts para el enmascaramiento de rostro en **RMI (Imagen de Resonancia Magnética)** cerebral.
 - Implementar programas que permitan la eliminación automática de la información grabada sobre la imagen.
- Tercer Objetivo: Crear pipelines de conversión de formatos de imagen entre **DICOM** y **NIFTI (Neuroimaging Informatics Technology Initiative)**.
- Cuarto objetivo: Integrar el software desarrollado en la plataforma **GEBID (XNAT)** de **BIMCV**.
- Quinto: Organizar sistema de transmisión entre los distintos departamentos de la comunidad con **GEBID** mediante tecnologías libres y gratuitas.

5 Ambito y entorno del proyecto

En este punto se pretende dar a conocer el ámbito en el que se desarrolla este trabajo. Este proyecto nace en base a las necesidades de la **Conselleria de Sanitat de la Comunitat Valenciana** impulsado gracias a la propuesta de **María de la Iglesia Vayá**. El proyecto tiene su espacio físico en un laboratorio dentro del **Centro de Investigación Príncipe Felipe** y su longevidad mínima ha sido establecida en unos dos años. El alumno **Angel Fernández- Cañada Vilata** accedió a dicha beca tras el proceso de selección realizado por la propia **Conselleria de Sanitat de la Comunitat Valenciana**.

5.1 Centro de Excelencia e Innovación Tecnológica de Bioimagen de la Conselleria de Sanitat (CEIB-CS)

El **Centro de Excelencia e Innovación Tecnológica de Bioimagen de la Conselleria de Sanitat (CEIB-CS)** está vinculado a la investigación y el desarrollo de los aspectos cualitativos y cuantitativos de la imagen biomédica, contemplando tanto la gestión, desarrollo de nuevos programas y proyectos tecnológicos de innovación, así como la traslación de los resultados provenientes de la innovación tecnológica a los centros sanitarios para mejorar los procesos y la calidad asistencial.

La creación del **CEIB-CS** se enmarca, dentro de la estrategia de la Conselleria de Sanitat para fomentar el desarrollo de las **Tecnologías de la Información y la Comunicación (TIC)** con el objetivo de ponerlas al servicio del bienestar y la calidad de vida de los ciudadanos y para mejorar la eficacia, la eficiencia y la calidad de la atención médica. Concretamente el **CEIB-CS** nace como fruto de la colaboración y sinergias entre las siguientes entidades:

- La Conselleria de Sanitat proporciona respaldo institucional al **CEIB-CS** y la explotación de los repositorios de imagen médica del sistema sanitario público valenciano.
- El **Hospital Universitario y Politécnico la Fe (HUPLF)** a través del Área de Imagen Médica, dirigida por el Dr. Luis Martí-Bonmatí, cuyo **Grupo de Investigación Biomédica de Imagen (GIBI 2³⁰)** proporciona la experiencia clínica y científica sobre imagen médica y sus biomarcadores asociados.
- El **Centro de Investigación Príncipe Felipe (CIPF)** como entidad de apoyo de las actividades que surgen de las iniciativas en investigación e innovación del **CEIB-CS**.

Aunque el **CEIB-CS** está abierto a la colaboración con otras entidades y grupos de investigación. Actualmente el **CEIB-CS** colabora íntimamente con el grupo del laboratorio de '**Neurobiología**' dirigido por **Vicente Felipo** en el estudio de la encefalopatía hepática o el grupo de '**Genética y Medicina Molecular**' dirigido por **Francesc Palau** con el estudio de enfermedades raras neuronales.

Las áreas médicas y de imagen en las que interviene el **CEIB-CS** son: la anatomía patológica, la óptica, microscopía, endoscopia, radiología y medicina nuclear entre otras.

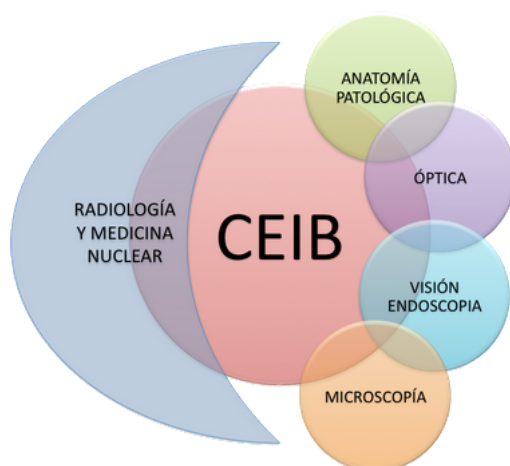


Figura 1: Areas médicas y de imagen en las que interviene el CEIB-CS

5.2 Banco digital de Imagen Médica de la Comunidad Valenciana (BIMCV)

Los biobancos de imágenes comienzan a percibirse como infraestructuras de investigación estratégicas. El **Banco digital de Imagen Médica de la Comunidad Valenciana (BIMCV)** es un banco de imágenes concebido como un repositorio de conocimiento orientado a favorecer los avances tecnológicos en imagen médica y proporcionar servicios de cobertura tecnológica para el soporte a proyectos I+D.

Actualmente, **BIMCV** tiene ámbito geográfico e incluye todos los departamentos de salud de la Comunidad Valenciana. No obstante, este sistema de gestión de la imagen médica digital integrada será posteriormente compatible con el nodo central del **Sistema Nacional de Salud (SNS)** y con otros nodos internacionales.

Sus principales objetivos son:

- Desarrollar e implementar estrategias para prevenir o tratar efectivamente las enfermedades mediante una infraestructura de investigación en imagen asociada a grandes estudios poblacionales.
- Proporcionar datos, herramientas y recursos de proceso para realizar estudios epidemiológicos y poblacionales avanzados en imagen.

5.2.1 Estudio poblacional y Big Data

El estudio poblacional y el análisis de los datos de imagen al que tiene acceso el **CEIB-CS** es una de las grandes tareas a desarrollar. Actualmente, el **CEIB-CS** trabaja con datos únicamente del Hospital San Juan de Alicante. Se han estudiado resonancias magnéticas cerebrales, obteniendo resultados sobre máximos y mínimos de (área, volumen, y espesor cortical) usando como covariables la edad y el sexo. Nuestro objetivo es llevar estos estudios a la totalidad de la Comunidad Valenciana. Para tal cantidad de datos es indispensable referirnos al popular término **Big Data**. Este término hace referencia a colecciones de conjuntos de datos muy grandes y complejos. Los cuales son imposibles de procesar con las habituales bases de datos y herramientas **TIC** más habituales.

Desde 1995 la gran cantidad de datos en neuroimagen recopilados se ha doblado aproximadamente cada 26 meses. A esta velocidad, en 2015 la cantidad de datos de neuroimagen adquiridos, sin incluir la información de las cabeceras, y los datos generados durante el procesamiento y análisis estadístico, podría exceder un por término medio los 20GB por cada estudio de investigación publicado.

La investigación en neuroimagen, por su propia naturaleza, maneja grandes volúmenes de datos, es multimodal y colaborativa. Sin embargo, la infraestructura necesaria para apoyar el avance de la investigación cerebral (donde los datos son cruciales) está todavía en desarrollo.

Los próximos pasos para el desarrollo de recursos que den soporte al **Big Data** de imagen cerebral a la escala de hexabytes requerirán la próxima creación de nuevas herramientas y servicios para la consulta de datos, integración, análisis y visualización.

5.2.2 Euro-Bioimaging y BIMCV



Figura 2: Logo Euro-Bioimaging

Euro-Bioimaging es un proyecto a gran escala cuyo objetivo es la creación de una infraestructura europea para la investigación dentro del plan de la **European Strategy Forum on Research Infrastructures (ESFRI)**.

La fase de preparación del proyecto tuvo lugar desde Diciembre del 2010 hasta Mayo del 2014. Actualmente, **Euro-BioImaging** está representado por doce países y la **European Molecular Biology Laboratory (EMBL)** como organización internacional, todos ellos se encuentran trabajando conjuntamente en la fase de construcción de la infraestructura.

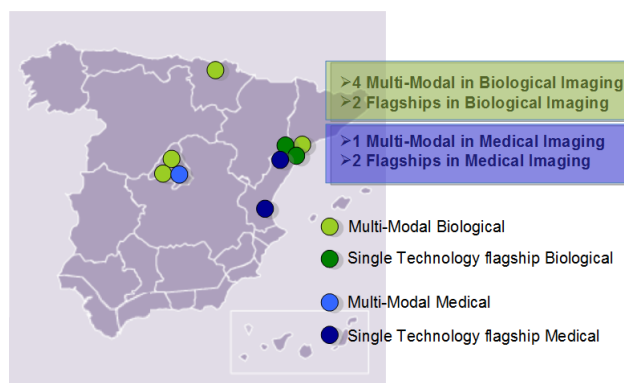


Figura 3: Nodos Euro-Bioimaging en España

En España existen nueve nodos. El nodo Valenciano **BIMCV** desarrolla y da acceso a una gran base de datos anonimizada de imágenes y registros clínicos asociados. Este repositorio **Big Data** tiene el objetivo de tener acceso a los datos de los hospitales de la región de Valencia (5 millones de habitantes viviendo en un área de 23.255 Km², con un promedio de 5.3 millones de casos por año procedentes de 210 técnicas de imágenes diferentes). Sin embargo, el nodo tiene la habilidad de incorporar datos de otras fuentes. Actualmente el nodo incorpora los datos del departamento 17 de la Comunidad Valenciana y varios proyectos de instituciones como el **Centro de Investigación Príncipe Felipe (CIPF)**, el **Hospital Arnau de Villanova (Valencia)**, **Hospital Quirón (Valencia)** o el **Oceanogràfic**.

El acceso a estos datos y herramientas supondrá un avance para la investigación y los estudios de imagen poblacional.

Servicios del nodo:

- **BIMCV** provee de un servicio de almacenamiento en un **VNA (Vendor Neutral Archive)**.
- El nodo integra el acceso a servicios computacionales de alto rendimiento de las infraestructuras locales y europeas (**neuGRID e-infraestructure**, **Infraestructura del CIPF** y la **UPV-I3M**).
- **BIMCV** define una metodología de acceso público (<http://ceib.bioinfo.cipf.es/xnat/>) para integrar diferentes tipos de datos para imagen poblacional, recursos cuantitativos y extracción de características.

5.2.3 Gestor de Ensayos y programas de investigación con Bioimagen para la I+D (GEBID)

El Gestor de Ensayos y programas de investigación con Bioimagen para la I+D (GEBID) es una instancia de **eXtensible Neuroimaging Archive Toolkit (XNAT)** perteneciente al **CEIB-CS**. Esta plataforma es un nodo **DICOM** orientado a proyectos de investigación que proporciona a la comunidad científica una herramienta sencilla y potente para poder organizar sus ensayos clínicos definidos con sus protocolos asociados. En nuestro caso aloja principalmente datos de neuroimagen.

La plataforma **XNAT** fue desarrollada y es mantenida por la Universidad de Wahington en St. Louis (Marcus et al., 2006-2007). Está diseñada para ayudar a los investigadores a almacenar, organizar y procesar datos de neuroimagen basándose en una estructura de proyectos, sujetos y sesiones.

Una de las grandes cualidades de **XNAT** es que es código abierto y puede ser instalado, configurado y modificado según las necesidades del grupo mediante el trabajo de programadores expertos, debido a su complejidad. Nuestro grupo se ve involucrado en su costumización y ampliación debido a que trabajamos con ella continuamente y los desarrolladores no pueden atender el rápido número de necesidades que vamos requiriendo. Gracias a esta flexibilidad, **GEBID** se puede utilizar para apoyar una amplia gama de ensayos clínicos y proyectos de investigación basados en bioimagen.

A su vez, **GEBID** permite compartir la información de sus proyectos, siempre bajo un entorno seguro y controlado, con otros gestores basados en **XNAT** o similares, lo que permite que la comunidad científica pueda colaborar entre sí y compartir datos de diferentes fuentes de una manera rápida y sencilla.

Otra de las bondades de **XNAT** es la posibilidad de ser ampliada mediante plugins que permiten la inclusión de nuevas funcionalidades. En nuestro caso las pipelines que estamos y hemos desarrollado son las siguientes:

- Conversión de **DICOM** a **NIFTI**.
- Control de calidad.
- Anonimización de datos impresos sobre la imagen.
- Eliminación de rasgos faciales de la imagen del paciente que puedan permitir su identificación (Defacing).

Por otro lado, **GEBID** incorpora un sistema avanzado de control de acceso a la información contenida en sus proyectos, permitiendo al administrador dar visibilidad individualizada a cada usuario, decidiendo quién accede a qué información. Otra de las posibilidades que permite **GEBID** es la capacidad de compartir datos entre diferentes proyectos, dando una visibilidad horizontal a la información. Además, para facilitar auditorías de seguridad **GEBID** incorpora un sistema de logs en dónde se puede rastrear toda acción realizada en el entorno.

GEBID también incorpora una serie de potentes servicios de consulta que permiten al usuario construir filtros avanzados que extraen la información deseada del proyecto, presentándola en pantalla o descargándola en diferentes formatos los datos para su análisis o uso externo. Estos filtros pueden configurarse mediante su interfaz web de **GEBID** o a través del uso de servicios **REST (Representational State Transfer)** para **XNAT** (Hadley, M. et al., 2009).

Como desventaja se debe plantear la dificultad de la instalación por las dependencias que acarrea y por la falta de información de la aparición de un error dentro de la instalación, es decir, visualmente puede parecer que la plataforma funciona y está bien instalada pero realmente no lo está. La dificultad de su gestión ha obligado al grupo a suscribirse a la cuenta de github de los desarrolladores para recibir al instante cada una de sus modificaciones.

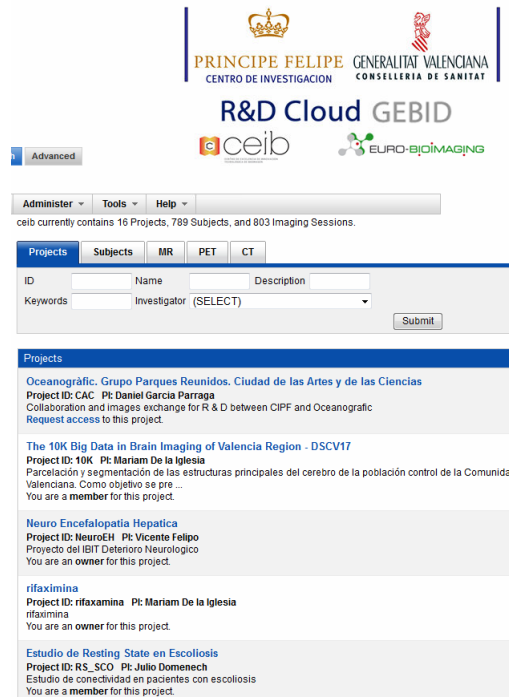


Figura 4: Página de inicio de la plataforma GEBID

6 Aplicaciones precedentes

En este apartado se da una muestra de las herramientas desarrolladas por otros grupos de investigación (con los que colaboramos estrechamente). Estas herramientas han sido utilizadas y ampliadas con el objetivo de conocer sus virtudes y defectos y de esta forma crear una herramienta adaptada a nuestras necesidades que contenga las características más deseables para nuestros propósitos.

6.1 C-PAC Configurable Pipeline for the Analysis of Connectomes



Figura 5: Configurable Pipeline for the Analysis of Connectomes

Es una plataforma configurable nacida en el **Child Mind Institute de New York** que permite organizar y ejecutar de forma flexible un conjunto de tareas sobre las imágenes médicas y que trabaja con muchas librerías (por ejemplo: **FSL**, **AFNI**, **ANTS**,...) por lo que son requisitos de la misma.

El equipo principal del **C-PAC** está formado por **Cameron Craddock Michael Milham, Steven Giavasis, Daniel Clark, Carol Froehlich, John Pellman**. Agradezco el trato recibido de todos ellos en especial de **Cameron Craddock, Michael Milham y John Pellman** con los que más he tratado y más ayuda me han prestado.

Una de las ventajas de este sistema es que es código abierto y gratuito, se encuentra escrito en código **Python** por lo que se puede extender con facilidad. Concretamente en **Nipype** el cual es un proyecto de **Python** que proporciona una interfaz entre el software de neuroimagen existente y facilita la colaboración de todos estas herramientas en un único flujo de trabajo.

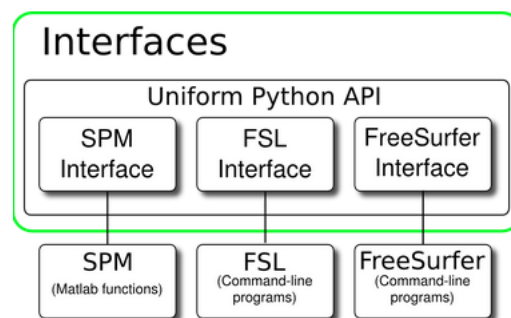


Figura 6: Esquema de interfaz Nipype

Tiene una interfaz de usuario muy bien organizada con explicaciones en cada uno de los pasos con el conocimiento subyacente de cada técnica con el objetivo de orientar fácilmente al usuario. Contiene tareas de todo tipo, como control de calidad, preprocesado, generación de mapas de conectividad (basados en la correlación obtenida a partir de semillas) extracción de elementos de series temporales en caso de no cumplir algunos requisitos de límite de movimiento (scrubbing), extracción del parénquima, parcelación en materia gris, blanca y líquido cefalorraquídeo basándose en probabilidades, etc. también permite lanzar estudios por paciente y grupo (siendo obligatorio lanzar primero un análisis de cada sujeto como prerrequisito para el de grupo). Además las pipelines pueden ser configuradas para excluir sujetos sin tener que modificar el espacio de trabajo cada vez que se quiera volver a pasar una pipeline a un sujeto o a un conjunto.

Otras ventajas que podemos citar son su facilidad de uso y su capacidad de ser configurado para ser lanzado en clusters de cómputo. Su instalación es sencilla pero un mal uso por desconocimiento de las técnicas y/o funcionamiento puede ocasionar que de siempre el mismo error o en el peor de los casos la sensación de que todo va bien generando resultados irreales. En esta aplicación importante repasar bien los logs.

También es importante saber lo que se desea ejecutar en cada momento, no es necesario pasar por todos los pasos de las pipelines una y otra vez en cada ejecución. Si se indica mantener el espacio de trabajo y no recalculan los datos de salida sólo se recalculará el paso que dio error o que no existía con anterioridad. Esto es importante pues no es un programa excesivamente rápido. Otro aspecto a mejorar es la representación gráfica de los resultados que tras varias ejecuciones puede dar problemas y la velocidad de cálculo la cual es superada por las aplicaciones basadas en Matlab.

Finalmente cabe destacar que **C-PAC** no funciona bajo **Windows** pero si que está probado en **Linux** y **OS X**.

Utilizamos esta herramienta durante un largo período de tiempo incluso en los principios de este proyecto la fuimos extendiendo con nuevo código, no obstante su lentitud y mala presentación de resultados nos hicieron abandonar esta opción. De ella extrajimos la idea de utilizar la librería **SPM (Statistical Parametric Mapping)** para determinar el movimiento entre instantes de tiempo y el uso de los marcadores de movimiento como el **Framewise Displacement** conjuntamente con el marcador **DVARS** que se explican más adelante.

6.2 Data Processing & Analysis of Brain Imaging (DPABI)

Está realizada por **Yan Chao-Gan** y su equipo. Fue programada en **Matlab** y se basa sobre todo en **FSL (FMRIB Software Library)** y **SPM** para realizar todas las operaciones. Para usarla basta con cumplir los prerequisites y agregar el directorio **DPABI** como una librería.

De todas las funcionalidades presentes las que más utilizamos son:

1. **Data Processing Assistant for Resting-State fMRI (DPARSF) 3.2 Advanced Edition.**
2. **DPARSF 3.2 Basic Edition.**
3. **Quality Control**
4. **Statistical Analysis**
5. **Viewer**



Figura 7: Interfaz principal de la herramienta DPABI

Las funciones uno y dos engloban las principales características. La diferencia entre la edición Advanced y la Basic es el número de parámetros por defecto que tiene. Su peor característica es la interfaz de usuario la cual si no conoces algo del tema puede marear. Por otro lado, uno de sus puntos fuertes es su rapidez.

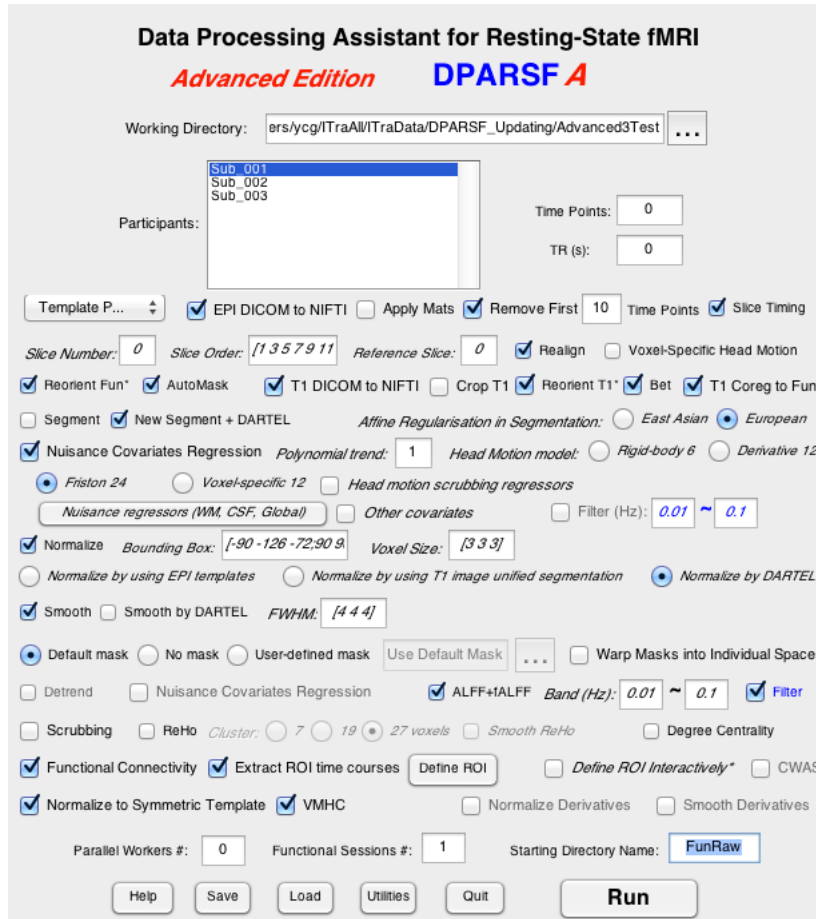


Figura 8: Interfaz DPARSF Avanced

En cuanto al control de calidad hay que destacar que en **DPABI** calculan el movimiento usando como primer volumen de referencia el primero de la serie sin embargo en el **C-PAC** lo hacen respecto a un volumen medio ficticio. Es por ello que ambas aplicaciones darán gráficas de movimiento diferentes con lo que esos cálculos pueden ser arrastrados al resto de cálculos. Otra diferencia con respecto al **C-PAC** es que **DPABI** calcula el **Framewise Displacement** siguiendo las ecuaciones de **Power, Van Dijk y Jenkinson** mientras que el **C-PAC** usa el de **Power** y el marcador **DVARs** del mismo. El **DVARs** es un marcador que no calcula **DPABI**.

Por otra parte esta aplicación permite realizar test estadísticos como la **ANOVA (ANalysis Of VAriance)** y el **Two Sample t-Student Test** y de esta forma podemos comparar grupos.

En cuanto al Viewer, con esta función podemos visualizar los resultados de los test estadísticos y ver las correlaciones (rojo) y anticorrelaciones (azules) entre los grupos designados. También nos permite seleccionar un umbral del estadístico p cuyos valores solemos usar entre 0.05 y 0.01 (en la imagen inferior no se ha establecido ningún umbral). Por último nos permite marcar el área anatómica en

la que se encuentran los máximos de cada región de interés siguiendo el atlas que deseemos (**AAL**, **Brodman**, **Harvard-Oxford**)

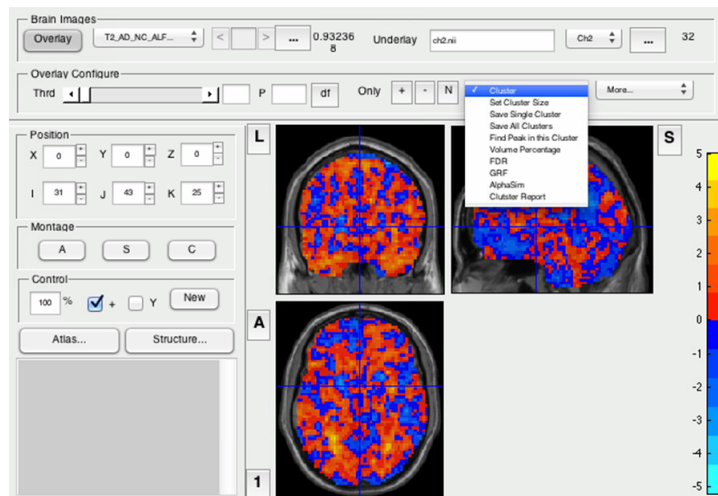


Figura 9: Interfaz que muestra correlaciones y anticorrelaciones entre áreas.

Esta herramienta también es libre y gratuita. De ella reutilizamos la implementación del **Frame-wise Displacement** siguiendo las ecuaciones de **Power**, **Van Dijk** y **Jenkison**, aunque después de analizar las tres y tras lo aprendido con el **C-PAC** nos quedamos con la de **Power**. Esta aplicación también nos ayudó a decantarnos por **Matlab** como plataforma de desarrollo y lenguaje gracias a la rapidez observada.

6.3 Proyecto EMBARC



Figura 10: Logo proyecto EMBARC

Fue un proyecto orientado a ampliar **XNAT** el cual tenía como objetivo realizar control de calidad dentro de la plataforma mediante la ejecución de una serie de pipelines. Se llegaron a publicar varios pósters como '**ExtendedBOLDQC Automated Functional MRI Quality Assessment**' y '**The Informatics Backbone of the Brain Genomics Superstruct Project Open Data Release**' presentados por el **Research Neuroinformatics Group (RNG)**, **Harvard University**. Incluso llegó a estar ofertado en el **XNAT MarketPlace** y un tutorial. No obstante, actualmente no aparece en el **XNAT MarketPlace** y después de investigar escribiendo a todos los autores de los pósters y tutoriales nos dimos cuenta que el proyecto había sido abandonado. Este hecho fue uno de los impulsores de la realización de este trabajo.

En la figura se muestra su integración con **XNAT** y los parámetros que pensaban obtener. El sistema apostaba principalmente por los marcadores de control de calidad cuantitativos algunos de ellos siendo mostrados de forma visual.

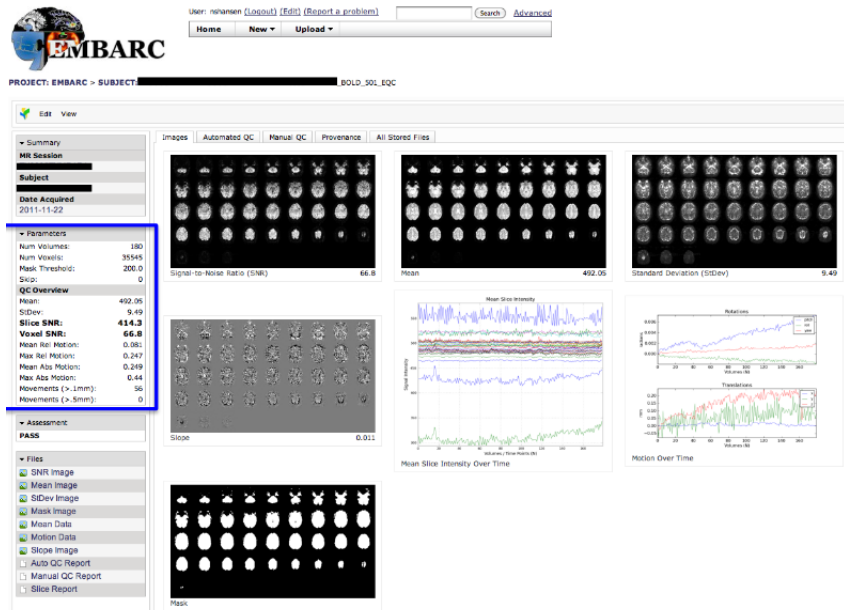


Figura 11: Prototipo Proyecto EMBARC

Este concepto de proyecto nos dio la idea de crear una serie de marcadores generales de las imágenes y una noción de cómo podía ser la presentación de resultados asociada e integración en **XNAT**.

6.4 Dicom Browser

Dicom Browser es una herramienta desarrollada por la **RNG (Research Neuroimaging Group)** cuya función principal es visualizar, inspeccionar las cabeceras **DICOM** y modificarlas ya sea mediante una interfaz o mediante scripts de extensión **.das**. Con estos mecanismos podemos modificar sesiones completas con una sola acción pues hay que tener en cuenta una única imagen puede dar lugar una gran cantidad de ficheros. Una vez realizado los cambios podemos crear una imagen nueva o modificar la original. También permite enviar los datos a través de la red a **XNAT** o a un sistema **PACS (Picture Archiving and Communications System)**.

Se encuentra disponible para muchos sistemas operativos **Linux, OS X y Windows**.

Los scripts **.das** pueden ser usados como scripts de anonimización pero es necesario indicar las acciones cabecera a cabecera. Por otro lado, tiene un escaso de instrucciones aplicables etiqueta a etiqueta que son muy básicas (no permiten bucles). En la siguiente web (<http://nrg.wustl.edu/software/dicomedit/dicomedit-reference/>) se puede comprobar las escasas posibilidades del lenguaje. En un principio se utilizó esta alternativa generando scripts **.das** a medida mediante scripts en **Python**. Dicom Browser permite aplicar dicho script y enviar los datos por línea de comandos. Bajo estas líneas se muestra un ejemplo de script de anonimización desarrollado por la **NRG**.

```

1 ///////////////////////////////////////////////////////////////////
2 // NRG sample anonymization script v. 001, 2010-06-21 by Kevin Archie <karchie@wustl.edu>
3 ///////////////////////////////////////////////////////////////////
4
5 // Instance Creator UID
6 (0008,0014) := ""
7
8 // Accession Number
9 (0008,0050) := ""
10
11 // Referenced Performed Procedure Step SQ
12 (0008,1111) := ""
13
14 // Referenced Patient SQ
15 (0008,1120) := ""
16
17 // Referenced Image SQ
18 (0008,1140) := ""

```

Figura 12: Ejemplo de script .das

Su limitada capacidad de generación de scripts de anonimización y falta de libertad de modificación nos hicieron buscar alternativas con estas cualidades.

7 Software utilizado y extendido

En este punto presentamos el software y herramientas utilizadas por el grupo para el desarrollo final de este trabajo. También se comentan sus ventajas y desventajas, la decisión del grupo de optar por el software libre y gratuito, así como algunos parámetros de las funciones usadas dentro de las herramientas y su razón de ser.

7.1 Matlab

El nombre **Matlab** nace de la unión de los términos "**matrix laboratory**". Fundamentalmente **Matlab** es un lenguaje de programación que ofrece un entorno de trabajo multiplataforma. **Matlab** fue ideado por **Cleve Moler** en 1984, en cuya primera versión utiliza una serie de subrutinas escritas en **Fortran** que proporcionan acceso sencillo al software de matrices **LINPACK** y **EISPACK** sin tener que usar **Fortran**.

Actualmente **Matlab** es un entorno de computación y desarrollo de aplicaciones que permite llevar a cabo proyectos con elevados requisitos de cálculos matemáticos, sistemas de visualización, procesamiento de señales y comunicaciones, automatización, etc, como es el caso del trabajo con imagen médica y **Big Data**.

Una de sus grandes ventajas es la capacidad de ampliación debida a una comunidad muy activa que genera nuevas herramientas de trabajo y a los propios desarrolladores de Matlab. Sin las cuales este trabajo no podría haberse llevado a cabo, entre ellas destacan (**DPARSF**, **SPM**, **MRI Inhomogeneity Correction** desarrollada por José Vicente Manjón en esta misma universidad).

Por otro lado, el lenguaje de **Matlab** proporciona soporte nativo para las operaciones de vectores y matrices que resultan fundamentales a fin de resolver problemas en imagen médica. Esto hecho permite un desarrollo y una ejecución rápida ya que no es necesario realizar tareas de bajo nivel tales como declarar variables, especificar tipos de datos y asignar memoria. En muchos casos, el soporte para las operaciones de vectores y matrices elimina la necesidad de bucles haciendo programas de menor longitud y de menor coste computacional. También es capaz de aumentar el rendimiento mediante técnicas multihilo las cuales permiten aprovechar todos los procesadores disponibles.

Otra herramienta incluida en **Matlab** es la ventana de comandos la cual nos permite introducir datos, ejecutar comandos o programas y mostrar los resultados de forma interactiva. Esta funcionalidad nos es útil tanto para la depuración de programas como el testeo de entrenamiento con las nuevas herramientas incorporadas. Asimismo, su editor tiene un sistema de depuración muy eficiente que ayuda al programador a mantener la eficiencia del código. También permite incorporar llamadas a aplicaciones desarrolladas en otros lenguajes lo que nos aporta un mayor abanico de posibilidades de desarrollo. **Matlab** permite llamar directamente código escrito en **C**, **C++**, **Java** y **.NET**. Del mismo modo se puede invocar código **Matlab** desde aplicaciones escritas en **ShellScript**, **C**, **C++** o **Fortran**.

Una de las desventajas de esta plataforma es no ser gratuita algo fundamental para los grupos de investigación con presupuesto ajustado y con mayor perjuicio cuando dicho grupo no tiene ningún objetivo lucrativo y cuyos resultados son abiertos al mundo entero.

7.2 Octave

Octave es una aplicación libre y gratuita capaz de interpretar código **Matlab**. Portar el sistema a **Octave** permite reducir costes aunque existen ciertas dificultades en este paso pues el lenguaje no permite ejecutar todas las instrucciones **Matlab**. Asimismo para ejecutar **Octave** con interfaz de usuario son necesarias aplicaciones extras.

Al igual que **Matlab**, **Octave** cuenta con un grupo activo de desarrolladores y un gran apoyo de una enorme comunidad de usuarios. Además funciona en prácticamente cualquier sistema operativo mayoritario como **Windows**, **Linux**, **OS X**,...

El uso de **Octave** en este proyecto es mayoritariamente económico, ya que nuestros programas funcionan en muchos centros y no es posible comprar tal cantidad de licencias. Por otra parte, en un entorno grid todos los recursos están abstraídos de manera que el usuario no sabe en realidad dónde está ejecutando cada tarea; es el middleware el que decide cuál es el entorno de ejecución más adecuado. Esto significa que debe haber una licencia de **Matlab** por cada tarea en grid que lo requiera, cuyo coste puede ser abrumador. Nuestras aplicaciones tienen como objetivo futuro trabajar en grandes clusters, de este modo el paso a **Octave** para grandes proyectos es casi obligatorio.

7.2.1 FMRIB Software Library (FSL)

FSL es un conjunto de librerías de análisis orientada a datos de imágenes de resonancia magnética convencional, funcional e imagen por tensor de difusión. Fue creada por el grupo '**Analysis Group, Oxford Centre for Functional MRI of the Brain**' (**FMRIB**), Reino Unido. **FSL** Está disponible para muchas plataformas (**Apple** and **Linux**, y **Windows** mediante el uso de máquinas virtuales). Las herramientas de **FSL** pueden ser ejecutadas mediante interfaz gráfica o mediante línea de comandos (nuestra forma más habitual). Las herramientas de **FSL** usadas en nuestro proyecto son **FSL VIEW**, **FSL BET** y **FSL MATHS**.

- **FSL VIEW** es usado al igual que otras herramientas como **MRICRON** y **IMAGEJ** para la visualización de las imágenes.

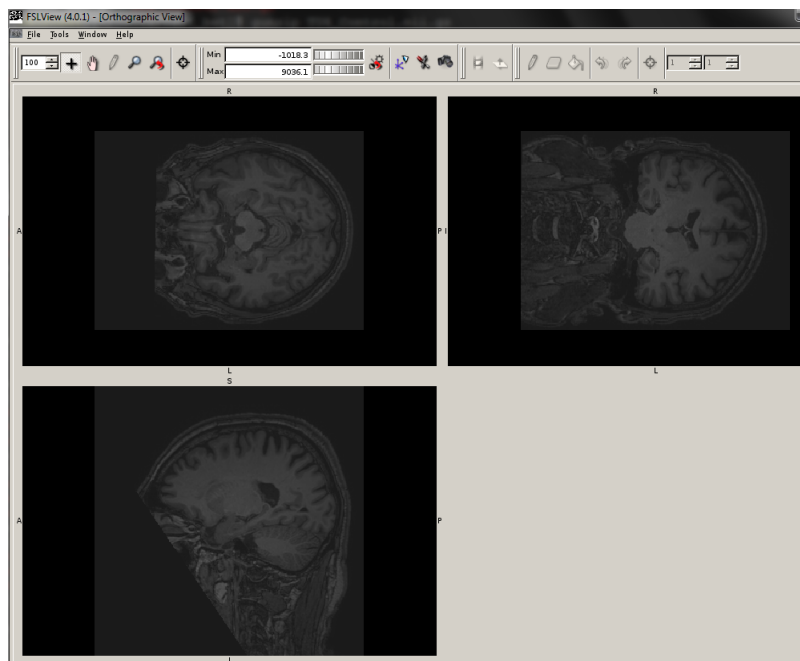


Figura 13: Captura de FSL VIEW de uno de los sujetos de estudio.

- **FSL BET (Brain Extraction Tool)** como su nombre indica trata de eliminar el tejido no cerebral de las imágenes quedándonos tan sólo con el parénquima para todo el cerebro completo. También tiene otras funcionalidades como estimar la superficie interna y externa del cráneo, etc. En nuestro trabajo lo utilizamos fundamentalmente para extraer de la imagen la zona del parénquima y para la creación de máscaras (parámetro -m).

Se han tomado otros parámetros para la ejecución de la extracción del parénquima que ahora paso a comentar:

- Robusto, Esta opción hace varias repeticiones del comando **BET** con las opciones seleccionadas y partiendo siempre de la misma imagen. En cada iteración estima un nuevo centro del cerebro en función del centro de gravedad estimado en la extracción anterior. De esta manera el centro del cerebro se acerca cada vez más hacia el centro real. Este proceso termina cuando el centro de gravedad ya no se mueve o un máximo de diez iteraciones.
- Umbral de intensidad fraccional, puede tomar valores entre 0 y 1 (con valor 0 no se observa ningún cambio entre la imagen de entrada y la de salida). Al aumentar el umbral más posibilidades hay de penetrar en el parénquima. Su valor por defecto es de 0.5. En este proyecto se han ido probando diferentes umbrales para varias imágenes, como resultado nos hemos seleccionado un umbral de 0.3 (en la imagen en color amarillo) podemos ver que se cubre casi todo el parénquima al contrario que en las otras opciones (0.2 en rosa, 0.4 en verde, 0.5 en azul y 0.8 en rojo)

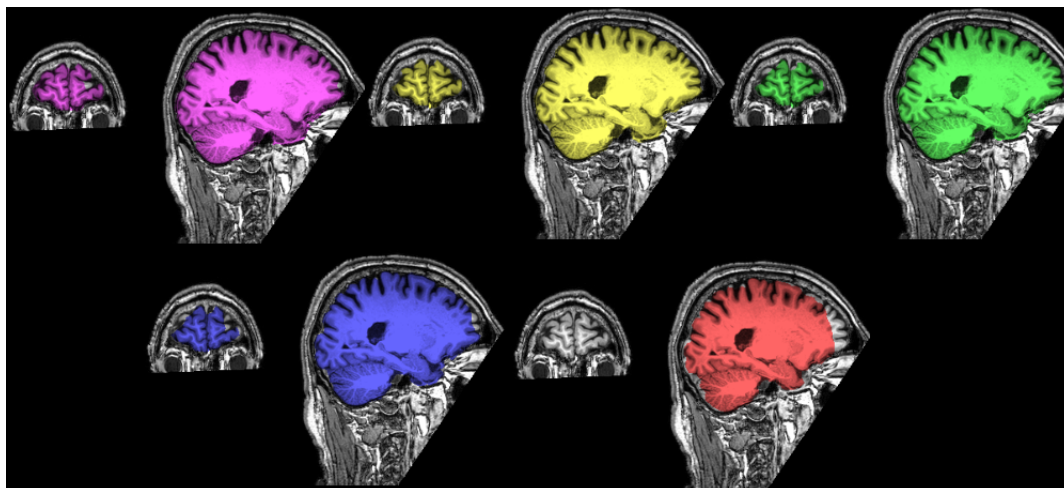


Figura 14: Resultado de aplicar diferentes umbrales de intensidad fraccional en una imagen de muestra.

- **FSL MATHS**, es una herramienta matemática para manipular las imágenes. En este proyecto ha sido usado para la normalización de las intensidades previo al cálculo del marcador **DVARS** del que hablaremos en puntos sucesivos.

7.3 Statistical Parametric Mapping (SPM)

SPM es una librería software diseñada para el análisis de las secuencias de imágenes cerebrales. Actualmente trabaja con imágenes de (**resonancia magnética funcional**) **fMRI**, (**tomografía por emisión de positrones**) **PET**, (**tomografía computarizada de emisión monofotónica**) **SPECT**, (**electroencefalograma**) **EEG** y (**magnetoencefalograma**) **MEG**. fue creada por el 'Wellcome Trust Centre for Neuroimaging. University College of London'.

En nuestro trabajo ha sido usado para las tareas de detección de movimiento, realineamiento y eliminación del rostro del paciente aunque tiene un gran repertorio de herramientas. Todas sus funciones pueden ser usadas mediante interfaz o sin ella.

El software está mayormente desarrollado en **Matlab** aunque también utiliza ficheros del tipo **MEX** escritos en lenguaje **C** y que deben ser compilados según la plataforma destino dónde vayamos a utilizarlos.

7.3.1 MRICron y DCM2NII

MRICron es un visor de imágenes en formato **NIFTI** el cual está disponible para los sistemas operativos **Windows**, **Linux** y **OS X**. Entre sus funciones destaca poder ver la imágenes tomográficas en varios planos, superponer capas y redenderizar volúmenes completos. Hay que comentar que en esta memoria se ha hecho de todas estas funciones ([véanse figuras 15 y 55](#)).

Por otra parte, incorpora la herramienta **dcm2nii** que permite convertir imágenes de formato **DICOM** a **NIFTI**. Esta función es muy importante pues casi todos los algoritmos que usamos suelen trabajar con la imagen en **NIFTI** y no en **DICOM**. También hay que comentar que esta conversión se puede hacer mediante código (nosotros lo hacemos en una pipeline dentro de **XNAT**) o mediante una interfaz de usuario.

Las opciones de la conversión que se presentan en la llamada por código se presentan a continuación. Como se puede observar el valor que pueden tomar son simplemente si o no.

- a de anonimizar, quita cierta información privada aunque básicamente sólo quita cuatro cabeceras por lo que es preferible la anonimización previa de la imagen **DICOM**. Aunque nosotros anonimizamos la **DICOM** al trabajar con la imagen **NIFTI** también borramos estos campos pues la anonimización en nuestro grupo es un principio muy importante.
- b si queremos cargar las opciones de un archivo preconfigurado. No solemos usar esta opción, sobre todo porque en la pipeline de **XNAT** ya se indica uno a uno cada uno de estos parámetros.
- c si queremos darle un directorio raíz y que el programa registre todas las subcarpetas y encuentre todas las imágenes y las trate como si estuvieran todas en la misma carpeta. Esta opción la usamos siempre pensamos que es muy útil.
- d añade la fecha al nombre del archivo resultante. Estos campos que hacen los nombres cada vez más largos son usados mínimamente. En este caso no se usa.
- e añade las series al nombre del archivo resultante. La serie si que la consideramos importante para diferenciar imágenes.
- f permite la transformación de formato **PAR/REC** de Philips a **NIFTI**. No se puede usar conjuntamente con la opción -n. Esta función no es usada normalmente pero como trabajamos con algunas máquinas Philips nos hemos visto obligados a utilizarla de vez en cuando.
- g si deseamos que el fichero de salida esté comprimido con gzip. En nuestro caso sí.
- i añade el ID del paciente del archivo original al resultante. Por tanto, no se usa.
- m activa el modo manual para que el usuario indique manualmente el formato de salida. Sólo se admite para imágenes de entrada de tipo **NIFTI**. Esta opción no es usada por nosotros en ningún caso.
- n si el valor es 'Y' como salida se obtiene un fichero .nii en caso contrario se crea un fichero del tipo .hdr/.img (**Analyze**).
- o esta opción indica el directorio de salida. Si no se indica nada es el mismo que el de entrada.
- p añade el protocolo al nombre del archivo resultante. Sí se usa.
- r si deseamos reorientar la imagen al plano ortogonal más cercano. No se usa.
- s indica si queremos que el formato de salida sea Analyze. En nuestro caso no.
- t crear informes. En nuestro caso no.
- v convertir todas las imágenes en el directorio. En nuestro caso sí.
- x reorientar y recortar las imágenes **NIFTI 3D**. En nuestro caso no.

8 Actualización de CEIBANON - Clinical Trial Processor (CTP)

El **Clinical Trials Processor (CTP)** fue desarrollado por la '**Radiological Society of North America**' (**RSNA**) bajo el proyecto '**Medical Imaging Resource Center**' (**MIRC**). El **CTP** es un software libre y gratuito que permite a los investigadores el intercambio seguro y la indexación de información (tanto de imágenes como otros documentos) para proyectos de imagen. La transmisión punto a punto se realiza de forma eficiente y está diseñada para soportar los protocolos de transporte estándar y trabaja con todos los sistemas **PACS (Picture Archiving and Communications System)**. Se encuentra desarrollado en Java por lo que es multiplataforma.

Dentro de los beneficios del **CTP** se encuentra el conjunto de pipelines que se pueden diseñar con ciertos módulos configurables incluidos. En nuestro caso hemos desarrollado/actualizado dos: una para la anonimización y envío directo a **GEVID** y otra para almacenamiento temporal en un repositorio previo. También permite programar tus propios plugins en lenguaje **Java** (aunque en nuestro caso no ha sido necesario).

El **CTP** usado en **CEIB-CS** es una instancia denominada **CEIBANON** pues su principal papel era el intercambio de archivos previamente anonimizados con **BIMCV**. Fue planteada por **José María Salinas Serrano** en su tesis doctoral '**Cloud CEIB I+D Sistema de gestión y extracción de conocimiento de la imagen médica**' dirigido por el **Dr. Miguel Angel Cazorla Quevedo**, **Universidad de Alicante, Hospital Universitario San Juan de Alicante**.

Pipelines desarrolladas o actualizadas por el alumno:

- **Actualización pipeline de anonimización de caberas según los perfiles del estándar DICOM de la National Electrical Manufacturers Association (NEMA) y envío a GEBID.** Esta pipeline transporta la información del centro a un espacio de trabajo común en el propio centro, la filtra, la registra, la anonimiza y la envía a **GEBID** en el exterior. La función de filtrado permite cribar el tipo de imagen. Actualmente trabajamos con **resonancia magnética** las cuales vienen marcadas como '**MR**' en el etiqueta (0008,0060) Modality, las imágenes rechazadas pasan a una cuarentena para su revisión. El paso del registro nos permite mantener un histórico de lo que se ha enviado. La sección de anonimización será explicada en puntos sucesivos. Y los pasos de envío y recepción identifican direcciones de entrada de salida y espacios de trabajo.
- **Pipeline recolectora.** Esta pipeline se encarga de almacenar toda la información que llega al nodo desde el punto origen (por ejemplo información del hospital al centro de sistemas de información del hospital). Este procedimiento es necesario cuando un centro desea enviarnos información debemos 'limpiarla' y seleccionarla antes de permitir el envío pues eso nos soluciona gran parte de las incidencias que pueden ocurrir.

Para esta función se desarrolló un script en Matlab. Para ello se utilizó una función oculta de Matlab conocida como `dicomparse` la cual permite leer etiquetas concretas del fichero **DICOM** en hexadecimal evitando la cargar completa en memoria de todas las etiquetas en una estructura. Esto supuso un gran aumento en la velocidad del script. En esta rutina también comprobaba la existencia de las etiquetas como (0018,1030) ProtocolName, necesarias para asignar el proyecto en **GEBID**, (0008,103E) Series Description, necesaria en **GEBID** para la organización de las imágenes y por último se buscaba la etiqueta (0002,0010) Transfer Syntax UID, la cual según una función interna de **Matlab** nos permitía saber si la imagen podía ser tratada por nuestro sistema o iba a causar fallos. Por otra parte, todas las imágenes excluidas quedan registradas en un registro.

Finalmente y con el tiempo nos vimos forzados a portar este código a **OCTAVE** pues no era viable el pago de una licencia **Matlab** en cada centro con el que trabajamos. Todo el código fue más o menos fácil de portar a excepción de la función `dicomparse` la cual es una función compilada en **C**. Como sustitución se utilizó la funcionalidad `gdcmanon` de **Grassroots DICOM (GDCM)** pues **GDCM** es un sistema más grande que implementa el sistema **DICOM** y tiene más funcionalidades.

Por otro lado, para aumentar la velocidad de la ejecución y dado de que cada fichero a la hora de ejecutar el limpiador es independiente de los demás tenemos como objetivo la paralelización del código.

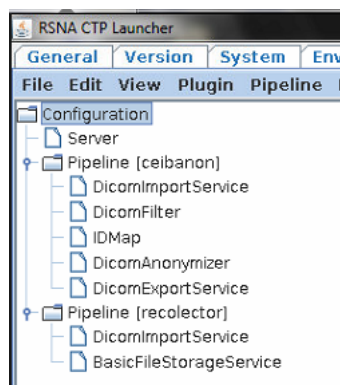


Figura 15: Pipelines desarrolladas en el CTP (CEIBANON)

El envío de imágenes al **CTP** ha sido testeado desde '**Dicom Browser**', '**OsiriX**' y '**File Sender**' (programa del mismo **MIRC**) sin problemas.

9 Desarrollo del sistema de envío de datos desde los centros al BIMCV

En este punto se explica el sistema de envío de datos desde un centro al **BIMCV**.

El primer procedimiento que se debe realizar es el envío de las imágenes desde cualquier aplicación con ese propósito a un nodo **DICOM**. Ejemplos de estas aplicaciones son, **OsiriX**, **Dicom Browser**, **File Sender**, etc. Estas aplicaciones envían dicha información al **CTP (CEIBANON)** dentro del propio centro, el cual tiene dos pipelines de trabajo. El puerto a seleccionar es el de la pipeline de recolección de datos. El cual almacena los datos a enviar dentro del propio centro dejándolos disponibles para la limpieza de las imágenes no usables.

Posteriormente y después de cribar las imágenes se envían los datos a la pipeline del **CTP** que filtra las resonancias magnéticas (pues por ahora sólo trabajamos con estas últimas), guarda los IDs anonimizados en una base de datos, anonimiza las cabeceras **DICOM** (como ya se ha explicado anteriormente), asocia el identificador del proyecto del XNAT a la etiqueta (0018,1030) 'ProtocolName' y envía los datos de forma segura a **BIMCV** por https aunque cuando este servicio se establezca definitivamente el acceso se establecerá mediante una **Virtual Private Network (VPN)** dentro de la Conselleria de Sanitat.

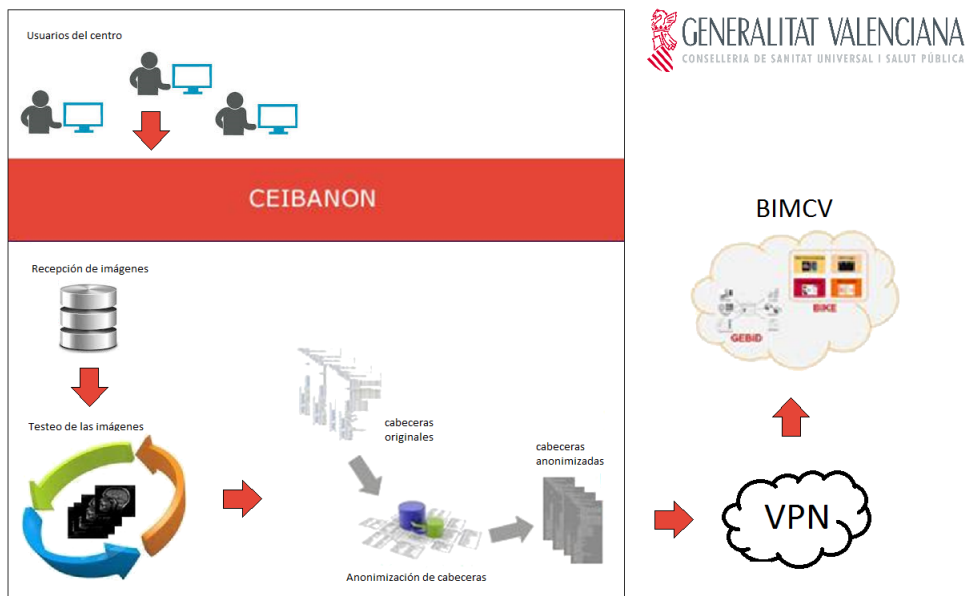


Figura 16: Sistema de envío de datos desde los centros al BIMCV

10 Implementación de la pipeline de conversión de DICOM a NIFTI automatizada

En este apartado se trata la creación de una pipeline. Una pipeline está compuesta por un conjunto de tareas que se deben de ejecutar en un cierto orden aunque cada tarea en si puede ser ejecutada en paralelo si las circunstancias lo permiten.

La pipeline que tratamos en este apartado se basa en el siguiente conjunto de tareas: seleccionar los archivos a ser tratados, crear un espacio de trabajo, copiar los archivos al espacio de trabajo, realizar la conversión de **DICOM** a **NIFTI** y posteriormente crear la carpeta de resultados para los sesiones tratadas y subir los resultados a dicha carpeta. Para definir las pipelines se debe crear por un lado un xml con los pasos a realizarse en el orden adecuado y por otro lado los xml que definen las tareas/comandos y sus parámetros. Bajo estas líneas podemos ver una de las tareas de la conversión con los parámetros que se están utilizando.

```

<resource name="dcm2nii" location="pipeline-dcm2nii.gz">
  <argument id="4">
    <value>Y</value>
  </argument>
  <argument id="a">
    <value>Y</value>
  </argument>
  <argument id="c">
    <value>Y</value>
  </argument>
  <argument id="d">
    <value>N</value>
  </argument>
  <argument id="e">
    <value>Y</value>
  </argument>
  <argument id="f">
    <value>Y</value>
  </argument>
  <argument id="g">
    <value>Y</value>
  </argument>
  <argument id="i">

```

Figura 17: Ejemplo de tarea de la pipeline de conversión de **DICOM** a **NIFTI**

Una vez creada la pipeline se debe indicar a **XNAT** las pipelines a las que se tiene acceso y asignar para cada proyecto las pipelines que estarán disponibles y si éstas serán de ejecución automática al recibir una nueva sesión o ejecutadas a voluntad del propietario del proyecto. En la figura inferior podemos ver como a cierto proyecto se le han asignado las pipelines de conversión de **DICOM** a **NIFTI** y de control de calidad.

				Applies To	Generates	Name	Description
	Edit	Details		All Datatypes		pipeline-dcm2nii.gz	Genera una carpeta NIFTI a nivel de SCAN con los archivos *.nii.gz, *.bval y *.bvec de la serie.
	Edit	Details		All Datatypes		WebBasedQCImageCreator	Pipeline for generation of QC Images

Add More Pipelines

Figura 18: Ejemplo de sistema de gestión de las pipelines de un proyecto de **XNAT**.

11 Desarrollo del Control de Calidad Centralizado (CQC)

Después de convertir las imágenes anonimizadas a formato **NIFTI** ya somos capaces de extraer información interesante de los datos facilitados. Todo el trabajo basado en control de calidad ha sido programado en Matlab e integrado en **GEBID** al igual que se muestra en el apartado anterior. El objetivo principal **Control de calidad Centralizado (CQC)** es suministrar a los investigadores que deseen utilizar las imágenes almacenadas, los marcadores de calidad que les aseguren unos resultados veraces pues no se puede hacer buena ciencia a partir de datos que den una imagen deformada de la realidad.

Existen principalmente dos aproximaciones para el análisis de la calidad de una imagen médica:

- **Análisis cualitativo:** guiado por un ojo experto. En esta aproximación se recorren los cortes de la imagen (caso de imagen tomográfica) en busca de artefactos.
- **Análisis cuantitativo:** guiado por los datos. En esta aproximación se evalúan índices numéricos calculados por un sistema automatizado.

En este proyecto se intenta dar solución a ambas aproximaciones aunque centrándose prioritariamente en la aproximación cuantitativa. Actualmente para la aproximación cualitativa se dan recursos para que el interesado busque los artefactos guiado por nuestros resultados cuantitativos u observe la imagen dentro de una serie de capturas a base de ejemplo. En un futuro se prevé integrar un visualizador completo que facilite dicha tarea manual. Por otro lado, las imágenes tratadas mediante el **CQC** por el momento son de tipo tomográfico y sobre todo de resonancia magnética.

11.1 Características básicas de la imagen

En este primer apartado del informe de control de calidad se trata la información mínima necesaria para el conocimiento de la imagen almacenada. Como muestra y ayuda para el control cualitativo de la imagen lo primero que se muestra es un montaje con ciertos cortes de la imagen tomográfica. Si la imagen sólo incluye un volumen se muestra una selección como máximo de 16 capturas a lo largo de dicho volumen, en el caso de existir una serie temporal la colección de cortes corresponden al volumen central de la serie. En la imagen inferior podemos ver un ejemplo de una imagen estructural con un único volumen.

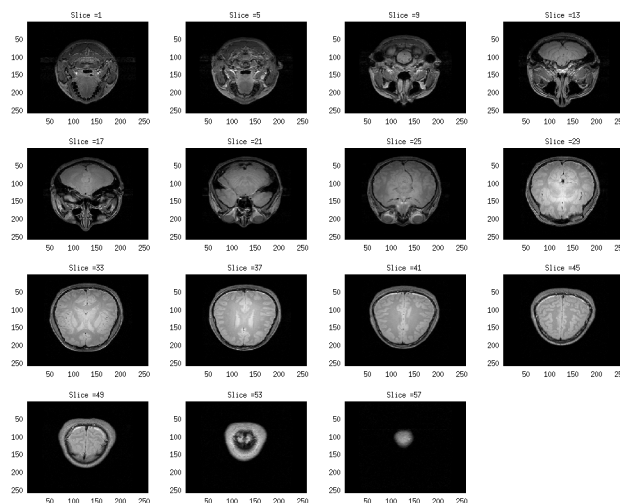


Figura 19: Capturas de muestra del volumen completo de una imagen estructural cedida por José Vicente Manjón

También se muestran valores comunes a las imágenes (estos valores pueden variar dependiendo del tipo de la imagen). Los parámetros mostrados son:

- **Número de vóxels:** este puede corresponder a un volumen (en una imagen estructural 3D), una secuencia de volúmenes (por ejemplo: en una resonancia magnética funcional).

- **Las dimensiones de los vóxeles:** corresponden al ancho, alto y espesor. Como se puede ver en el ejemplo bajo estas líneas los vóxeles son cuadrados aunque esto no siempre es así. El tamaño del vóxel es importante pues de su tamaño depende la cantidad de información que contiene. Vóxeles pequeños nos ayudan a obtener resultados más exactos (menos estructuras y tejidos en un mismo vóxel), también mejora la resolución espacial y el contraste no obstante, se ven más afectados por el movimiento del sujeto. Si el vóxel es grande, contendrá muchas estructuras y tejidos diferentes. Sin embargo, cuanto mayor sea el tamaño de vóxel la intensidad de la señal será más alta por lo que mejora la relación señal-ruido.
- **Número de cortes:** puede venir determinado en parte por el tamaño del vóxel utilizado o el tamaño volumétrico a estudiar.
- **Resolución de la imagen:** es el número de filas y columnas de intensidades de la matriz que contienen los valores que forman la imagen en un corte. En este caso la imagen es cuadrada por lo que la representación de la resolución se ha simplificado de 256 x 256 vóxeles a 256 vóxeles.
- **Relación señal - ruido (SNR):** es la proporción de ruido respecto a la señal de interés que se encuentra en la imagen. Mientras mayor sea esta proporción la imagen será de mejor calidad. En este caso utilizamos como medida de la señal la media de las intensidades del volumen/es dividido por la desviación típica de las intensidades del volumen/es. Los valores son normalizados mediante la aplicación de $10 * \log_{10} * \text{SNR}$.
- **Media, mediana (segundo cuartil) y desviación típica:** son marcadores estadísticos bien conocidos que nos ayudan a conocer la distribución de los datos.

Name	Value
Voxels number	3932160
Voxel width	1 mm
Voxel height	1 mm
Slice thickness	1 mm
Number of slices	60
Image slice resolution	256 voxels
SNR	5.3721 dB
Mean	99.9506
Median	105
STD	30.5481

Figura 20: Datos de ejemplo de la imagen de la figura anterior

11.2 Estimación del movimiento

Este paso de preprocesado únicamente se utiliza cuando tenemos varias imágenes de un sujeto en diferentes instantes de tiempo (por ejemplo el caso de una imagen de resonancia magnética funcional). Esta técnica consiste en estimar la diferencia de posición de la cabeza (tanto rotación como traslación) entre las distintas imágenes e instantes de tiempo. La estimación de estos movimientos nos permite corregirlos en etapas posteriores. No obstante otra de las aplicaciones para las que se puede usar esta estimación es para la obtención de marcadores de calidad pues mientras mayor sea el movimiento del sujeto mayor será la posibilidad de la aparición de artefactos dentro de la imagen y peor será la calidad de la imagen. Para corregir el movimiento se aplican las traslaciones y rotaciones adecuadas compensatorias, de este modo se consigue que las imágenes del mismo corte pero diferente tiempo coincidan en un espacio común.

Para la estimación del movimiento y realineamiento hemos utilizado la funcionalidad que nos presta **SPM**. Esta aplicación toma como imagen de referencia la primera adquirida. En nuestro programa en **Matlab** usamos las funciones `spm_realign.m` y `spm_reslice.m`. En este apartado nos interesa especialmente la estimación del movimiento respecto al primer instante de tiempo cuyos resultados se obtienen usando `spm_realign.m`. Los resultados se almacenan en un fichero llamado `rp_*.txt` (dónde el valor de * es el nombre de la imagen de entrada). Este fichero contiene seis columnas, las tres primeras hacen referencia a los parámetros de movimiento de traslación en los ejes x, y, z. Las últimas tres columnas hacen referencia al giro en cada uno de estos tres ejes, es decir, cabeceo, balanceo y 'derrape'.

Aunque los movimientos del paciente sean de baja amplitud (una traslación de un 1 mm o una rotación de 1 grado) son capaces de dar lugar artefactos de movimiento, el aumento de la cantidad de ruido e incluso causar áreas de falsa activación que pueden alterar los resultados estadísticos. En las imágenes inferiores aparece marcado el movimiento de rotación y traslación. Los asteriscos marcan movimientos de más de 1 mm respecto a la posición de referencia. Los movimientos de rotación han sido transformados a distancias mediante el cálculo del desplazamiento en la superficie de una esfera de radio 50 mm (aproximadamente la distancia media desde el córtex cerebral al centro de la cabeza). Para calcular la longitud del arco correspondiente a un ángulo de un grado se usa el siguiente procedimiento:

$$L = \frac{2 * \pi * r * \alpha}{360}$$

Figura 21: Ecuación para la obtención del arco de un grado x

dónde α son grados sexagesimales y r es el radio de la esfera (50mm).

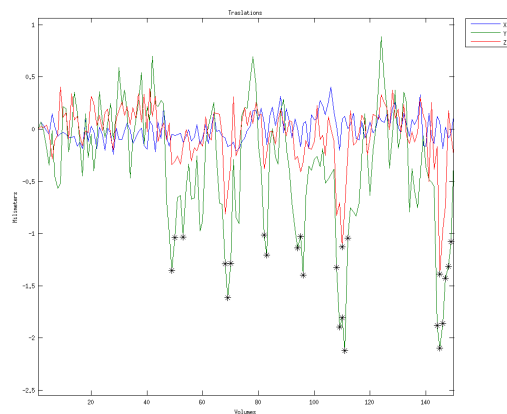


Figura 22: fMRI Traslaciones respecto al primer instante temporal obtenida a través de una imagen cedida por el grupo de Neurobiología (CIPF)

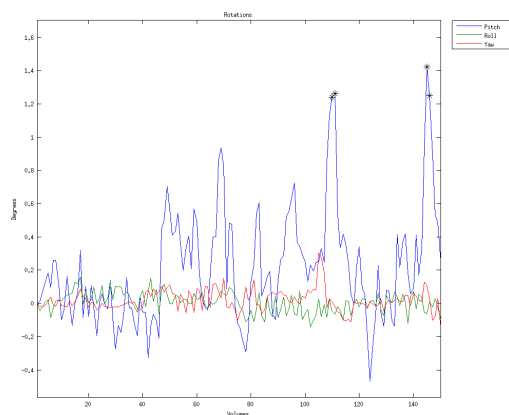


Figura 23: fMRI Rotaciones respecto al primer instante temporal obtenida a través de una imagen cedida por el grupo de Neurobiología (CIPF)

Con estos resultados además de los valores marcados se pueden medir más indicadores, entre ellos el movimiento de los arcos y traslaciones de más de 1 y 5 mm para ambos casos, así como el movimiento máximo y medio (absoluto y relativo, es decir respecto al instante de referencia o el instante anterior). En la imagen inferior se puede ver los resultados correspondientes a la imagen utilizada en este apartado. En los subsiguientes puntos 11.2.1 y 11.2.2 se va a seguir utilizando la misma imagen funcional que en este punto por lo que todos los resultados mostrados están relacionados.

Name	Value
Movements more than 1 mm	24
Movements more than 5 mm	0
Movements arcs more than 1 mm	4
Movements arcs more than 5 mm	0
Mean relative motion	0.00053667 mm
Max relative motion	1.6518 mm
Mean absolute motion	0.010799 mm
Max absolute motion	1.2405 mm

Figura 24: Valores obtenidos del análisis de la imagen cedida por el grupo de Neurobiología (CIPF)

11.2.1 Framewise Displacement (FD)

Existen varias medidas llamadas todas ellas **Framewise Displacement** y todas ellas intentan representar el movimiento de la cabeza de un sujeto durante la adquisición. El grupo ha utilizado la implementación de **Yan Chao-Gan (Institute of Psychology, Chinese Academy of Sciences)** como base. A continuación se presentan las tres alternativas más conocidas:

- La versión de **Jenkinson** nace como una medida para estimar la diferencia entre dos transformaciones con el objetivo de asegurar la precisión y robustez de los métodos de registro. Estas

medidas no pueden ser obtenidas seleccionando una serie de puntos de la imagen (por ejemplo las esquinas), es por ello que como marcador se utiliza la desviación media cuadrática o **Root Mean Square (RMS)** ya que es fácil de calcular y el error que se comete es pequeño (se supone que el volumen cerebral es una esfera).

Por otro lado, la transformación entre dos volúmenes puede actuar como coordenadas de vóxel, coordenadas globales (del mundo) o una combinación de ambas. Estos sistemas de coordenadas se relacionan mediante una transformación afín del tipo $\mathbf{T}\mathbf{w}: \mathbf{T} \rightarrow \mathbf{W}$, donde \mathbf{V} es el espacio de coordenadas de vóxel y \mathbf{W} son las coordenadas del mundo. De este modo Jenkinson expresa las transformaciones entre ambos sistemas de coordenadas lo que es más eficiente y robusto. A modo de ejemplo, $\mathbf{T}: \mathbf{V}_1 \rightarrow \mathbf{V}_2$ (siendo transformaciones entre dos espacios de coordenadas de vóxeles) se puede expresar en el sistema de coordenadas del mundo como $\mathbf{T}^{-1}_{\mathbf{w}_1} * \mathbf{T} * \mathbf{T}_{\mathbf{w}_2}$.

Una vez comentado todo lo anterior la media cuadrática de la distancia en mm (nuestro FD) se obtiene ,

$$d_{\text{RMS}} = \sqrt{\frac{1}{5}R^2\text{Tr}(\mathbf{M}^T\mathbf{M}) + \mathbf{t}^T\mathbf{t}}$$

Figura 25: Ecuación FD Jenkinson

dónde \mathbf{R} es el radio del volumen de interés y $\mathbf{T}^{-1}_{\mathbf{w}_1} * \mathbf{T} * \mathbf{T}_{\mathbf{w}_2}$ es usado para calcular la matriz \mathbf{M} y el vector \mathbf{t} . Tr o la operación **Trace** de una matriz cuadrada se define como la suma de los elementos de la diagonal principal (la que va desde arriba a la izquierda a abajo a la derecha)

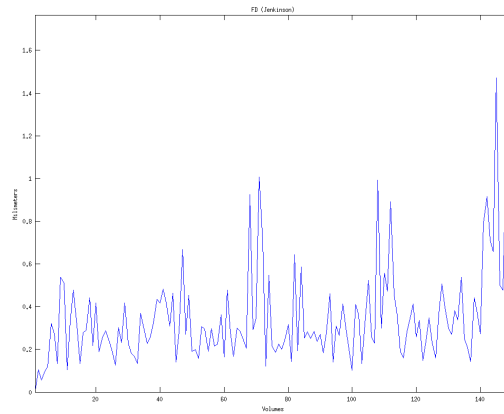


Figura 26: Resultados FD Jenkinson

- En la versión de **Van Dijk**, conocida como **Mean Framewise Displacement** se representa el desplazamiento absoluto medio entre el volumen actual y el anterior (en mm) estimado a partir de la traslación en cada uno de los ejes (x,y,z). El movimiento medio y absoluto para cada volumen se calcula como la media cuadrática (**RMS**) de los parámetros traslacionales. No tiene en cuenta los giros sobre cada uno de los tres ejes principales. Por este motivo, se ha descartado su uso, seguidamente se muestra un ejemplo y su fórmula de cálculo.

$$FD_i = \sqrt{x^2 + y^2 + z^2}$$

Figura 27: Ecuación FD Van Dijk

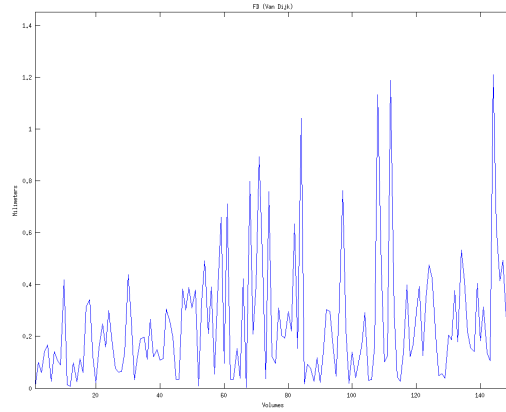


Figura 28: Resultados FD Van Dijk

- El marcador propuesto por **Power** representa el movimiento de la cabeza en cada instante de tiempo y es el resultado de agrupar seis parámetros de movimiento (tres desplazamientos y tres ángulos) en un único valor según la siguiente fórmula: $FD_i = |\Delta d_{ix}| + |\Delta d_{iy}| + |\Delta d_{iz}| + |\Delta \alpha_i| + |\Delta \beta_i| + |\Delta \gamma_i|$ donde $\Delta d_{ix} = d(i-1)_x - d_{ix}$ y de la misma forma para los otros parámetros. Los movimientos rotacionales son transformados a distancias mediante el cálculo del desplazamiento en la superficie de una esfera de radio 50 mm (distancia media desde el córtex cerebral al centro de la cabeza). Para calcular la longitud del arco de un determinado grado se ha usado la ecuación mostrada en la [figura 21](#).

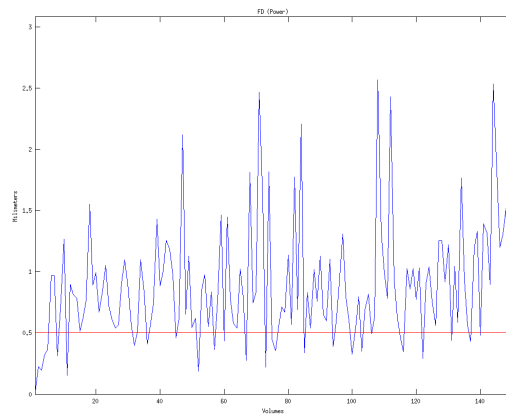


Figura 29: Resultados FD Power

Los tres marcadores vistos guardan una alta correlación. El **FD** de **Power** toma valores al menos el doble que los valores de **Jenkinson** y los de **Jenkinson** (por lo general) alcanzan valores mayores que los de **Van Dijk**. Esta información se ha extraído del artículo '**Addressing head motion dependencies for small-world topologies in functional connectomics**'[14].

Por otra parte, una de las cuestiones a tener en cuenta, es que cuanto mayor sea el **tiempo de repetición (TR)** mayor será la posibilidad de que ocurran movimientos de mayor amplitud. También cabe señalar que estas medidas pueden verse influenciadas por cómo se ha preprocesado la imagen pues la información del movimiento se extrae por completo de la propia imagen y no usa la ayuda de ninguna medida externa. Del mismo modo, la precisión de la medida se ve influenciada por la precisión de la adquisición en sí misma.

Seguidamente se muestran los resultados para la adquisición de **fMRI** (también usada en el punto anterior) con cada una de las tres metodologías. En todas ellas se ha utilizado como volumen de referencia el primero por lo que el movimiento inicial siempre es cero. Nótese también que aparece un umbral (0.5) para la versión del **FD** de **Power** proveniente de un artículo suyo del año 2012 '**Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion**'[4].

11.2.2 DVARS

Este marcador promedia el cambio de la señal **BOLD (Blood Oxygen Level Dependent)** a lo largo de cada volumen de la imagen, (la **D** hace referencia a la derivada temporal de los instantes de tiempo y **VARs** hace referencia a la varianza de la media cuadrática **RMS** sobre los vóxeles). Se calcula como sigue:

$$DVARS(t) = \sqrt{\frac{1}{I} \sum_i (Y_{it} - Y_{it-1})^2}$$

Figura 30: Cálculo de DVARS

dónde es el valor del vóxel i en el instante de tiempo t e I el número total de vóxeles de la imagen. Para la imagen tratada en el caso del Framewise Displacement y cedida por el grupo de Neurobiología (CIPF), se han obtenido los siguientes valores (la línea azul marca el límite recomendado de aceptación):

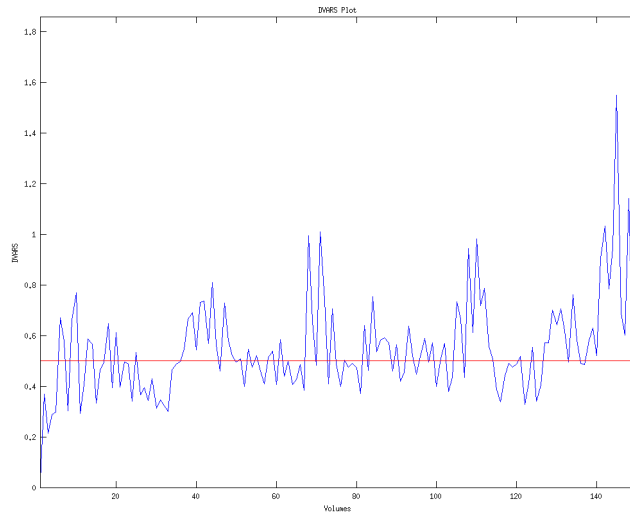


Figura 31: Resultados DVARS para la imagen de ejemplo.

El indicador **DVARS** no se puede usar por sí sólo, pues un cambio brusco en la intensidad de la señal no tiene por qué estar asociado al movimiento sino a causas fisiológicas o propias de la tarea realizada en el caso de (**fMRI** basada en tarea). Es por ello, que se debe de usar en conjunto con el marcador **FD**. Si ambos marcadores superan un cierto umbral se debe examinar la imagen en busca de artefactos. Según “**Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion**”[4] Power et al. (2012) un valor umbral adecuado para **FD** es 0,5 y para **DVARS** un 0.5% de la intensidad **BOLD**. Bajo estas líneas se muestra el resultado de límite de umbral para la imagen de muestra. Un valor de 1 indica umbral no superado.



Figura 32: Superación de los umbrales de aceptación para FD y DVARS

Volumes with excessive movement	6, 7, 9, 10, 13, 14, 18, 20, 25, 37, 38, 39, 40, 41, 42, 43, 44, 47, 48, 49, 51, 53, 55, 58, 59, 61, 63, 68, 69, 71, 72, 74, 77, 82, 84, 86, 87, 88, 90, 93, 96, 97, 99, 101, 102, 105, 108, 109, 110, 111, 112, 113, 114, 121, 124, 127, 128, 129, 130, 132, 134, 135, 138, 139, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150
---------------------------------	--

Figura 33: Volúmenes con excesivo movimiento.

Como se puede observar hay muchos volúmenes que son objeto de sospecha de excesivo movimiento debido a la baja resolución de las imágenes funcionales éstas podrían pasar desapercibidas ante el análisis visual cualitativo.

Existen tendencias que sugieren la eliminación de estos volúmenes (**Scrubbing**) no obstante, no es posible utilizar esta técnica cuando la cantidad de volúmenes a eliminar es muy elevada. Por otro lado, se ha demostrado que esta técnica mejora la correlación de las regiones vecinas de la imagen y empeora la de las más lejanas, por otra parte, la eliminación de varios volúmenes consecutivos puede alterar significativamente los resultados de análisis posteriores por lo que por ahora no nos planteamos su uso. Como reseña, el **C-PAC** implementa el **Scrubbing** usando como umbrales los valores configurados por el usuario de **FD** pero no dejando indicar un umbral de **DVARs** aunque internamente si lo usa.

11.3 Rango dinámico

Es el rango de diferencia tonal entre la parte más clara y la más oscura de una imagen. Cuanto más alto sea el rango dinámico, se pueden potencialmente representar más matices. No obstante, el rango dinámico no se correlaciona de forma directa con la cantidad de tonos reproducidos.

En nuestro desarrollo se ha calculado el rango dinámico para imágenes 2D (**RGB** y escala de grises), 3D y 4D. Para las imágenes 2D se obtiene un valor único si la imagen es en escala de grises y tres si la imagen es **RGB** (una por cada color, rojo, verde y azul), para las 3D se obtiene una gráfica en la que se representa rango dinámico contra corte, por último en el caso de las 4D se ha decidido usar la imagen media de todos los instantes de tiempo y actuar como en el caso de la 3D.

En la siguiente imagen podemos ver cómo cambia el rango dinámico a lo largo de los cortes. Hay que tener en cuenta que la cantidad y tonalidad de las estructuras cerebrales que aparecen en cada corte es diferente por lo que no es extraño que el rango dinámico sea constante. Por otra parte, podemos ver una caída del rango dinámico en los últimos cortes, esto es debido a que la imagen visualizada se ha captado en el plano transversal y en los últimos cortes apenas aparecen estructuras e incluso en el último corte desaparece por completo.

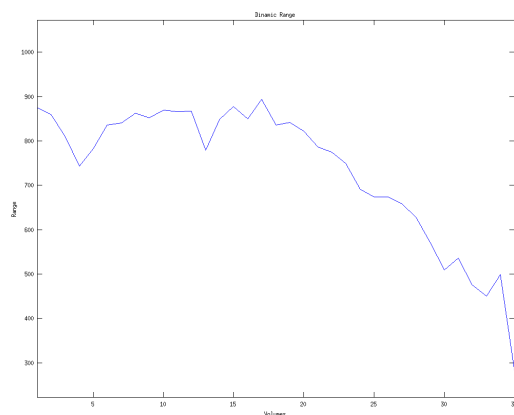


Figura 34: Rango dinámico de la imagen (fMRI) cedida por el grupo de Neurobiología (CIPF)

11.4 Contraste

El contraste se define como la diferencia entre la luz y las partes oscuras de la imagen. Un menor contraste da lugar a una imagen más plana, otras características visuales hacen que los objetos sean distinguibles o la diferencia entre dos puntos adyacentes de la imagen. Tan sólo medir la diferencia entre los puntos más o menos claros en la imagen no es una buena predicción de la percepción de contraste por las personas ya que el efecto del contraste depende de todos los tonos adyacentes. El contraste es importante clínicamente pues nos permite distinguir entre diferentes áreas o estructuras de interés.

En el artículo '**Evaluation of contrast measures in relation to observers perceived contrast**'[5] se elabora una función de contraste para imágenes en el espacio de color **CIELAB** (**Comission Internationale de l'Éclairage, Lightness** donde **A** y **B** son dimensiones de color opuestas) el cual se obtiene realizando una serie de operaciones con raíces cúbicas. Esta operación se debe realizar para todos los canales de color. Las ventajas de este espacio de colores es que incluye todos los colores perceptibles, haciendo que su gama sea superior a las de los modelos de color **RGB** (**Red Green Blue**) y **CMYK** (**Cyan Magenta Yellow black**). Uno de las características más importantes de **CIELAB** es la independencia del dispositivo, es decir, los colores se definen independientemente de la naturaleza de la adquisición o del dispositivo en que se muestran las imágenes.

El marcador llamado **Lab variance** explicado en el artículo citado se calcula como la media geométrica de la varianza de cada canal en el espacio de color **CIELAB**. Para el caso de las imágenes en escala de grises se ha utilizado como medida la desviación típica de los valores de intensidad. En la imagen inferior podemos ver como evoluciona el valor del marcador con los cambios de contraste provocados artificialmente. El valor 0 representa la imagen original, los números positivos un aumento (50 y 100) un aumento artificial en el contraste y los negativos (-50 y -100) una disminución del mismo. Como podemos ver a menor contraste menor es el valor del marcador Lab variance. No obstante, no hemos podido definir cuáles son los valores ideales de este marcador para una imagen.

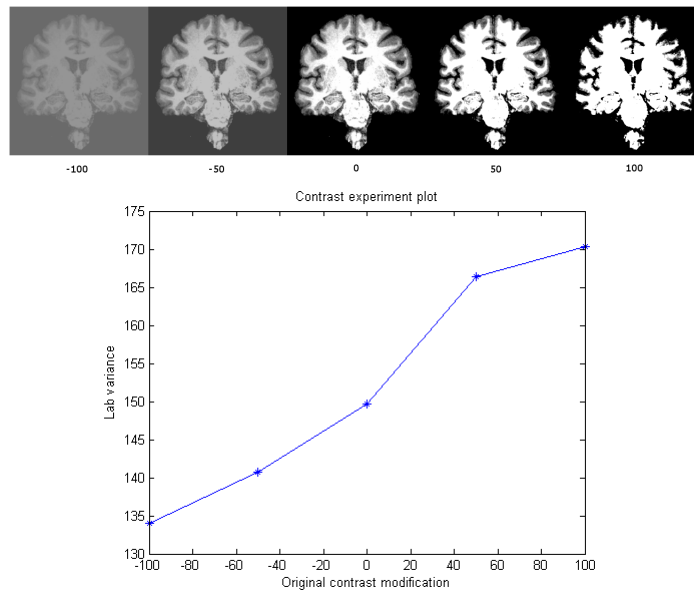


Figura 35: Experimento diferencia del marcador ante diferentes contrastes

En el caso de una imagen 3D se realizan los cálculos para cada corte y en una 4D para cada corte y para cada instante. En el caso 1D y 2D obtenemos valores mientras que en el caso 3D se obtiene una gráfica de líneas (al principio y al final de las gráficas no hay valores pues en los cortes suministrados al principio y al final no hay parénquima, en el corte primero se ha aplicado un algoritmo de detección de bordes (Sobel) para ilustrar mejor la cuestión). Finalmente para la representación de la 4D se ha realizado un mosaico con una línea por cada corte y un color que representa el valor alcanzado (el cual no ha resultado muy intuitivo).

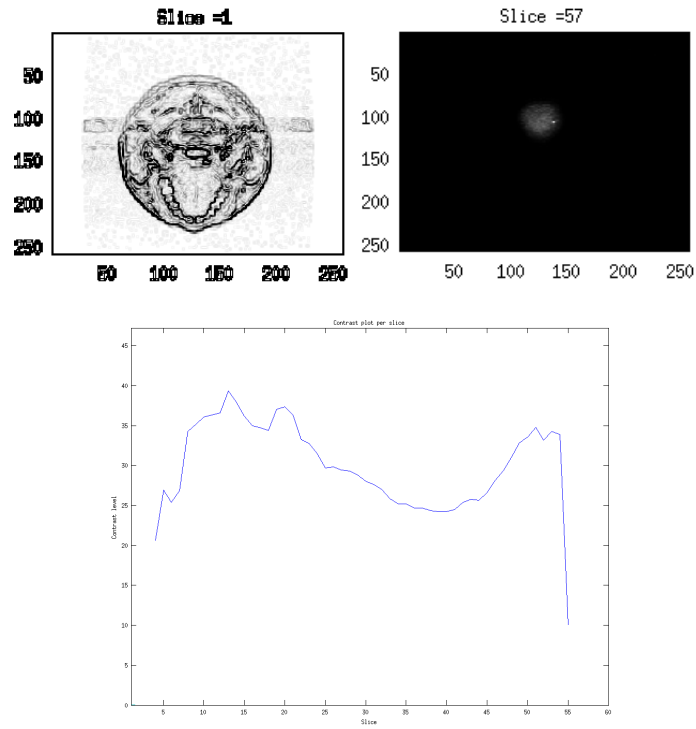


Figura 36: Representación del nivel de contraste por cortes

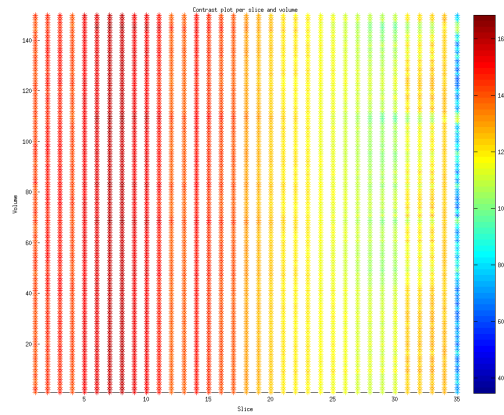


Figura 37: Representación del contraste por cortes e instantes de tiempo

11.4.1 Relación señal ruido temporal y relación contraste ruido temporal

- Relación Señal Ruido Temporal (tSNR)

Es la relación entre la amplitud de la señal y el ruido de la misma a lo largo del tiempo. Dado que es deseable que los valores de señal sean altos en un sistema de imagen y que el ruido sea lo

más bajo posible, los valores de **tSNR** deberán ser elevados. Es importante controlar el **tSNR** pues el nivel de ruido incluido en las imágenes reduce la fiabilidad de las medidas. El umbral de **tSNR** que se observa en la bibliografía suele encontrarse entre 1,5 y 2. Seguidamente mostramos como se calcula el **tSNR** para un vóxel determinado:

$$tSNR = \frac{\mu}{\sigma} = \frac{\mu}{\sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2}}$$

Figura 38: Fórmula tSNR

dónde N es el número de instantes de tiempo, μ es la media de toda la serie temporal, σ es la desviación estándar, X_i es la serie temporal medida en un vóxel, es por ello que la representación se realiza por vóxel. Previamente se ha obtenido y se ha aplicado una máscara mediante el comando **FSL BET**. Los valores obtenidos se han representado en decibelios (dB) mediante una selección de unos 16 cortes. Mostramos un ejemplo a continuación:

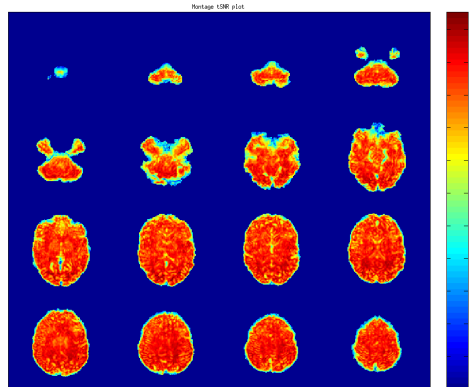


Figura 39: Ejemplo tSNR funcional

Name	Value
Mean tSNR	3.6037
tSNR First quartil	0
tSNR Second quartil (Median)	17.4352
tSNR Third quartil	19.34
tSNR data Interquartile Range	19.34
Data over threshold 1.5	72.1 %

Figura 40: Distribución del tSNR

- **Relación Contraste Ruido Temporal (tCNR)**

Este marcador intenta plasmar las diferencias de intensidad entre las regiones de la imagen con respecto al ruido. Al contrario que la proporción señal-ruido mientras más bajo sea este valor mejor calidad tendrá la imagen. En este caso el cálculo se realiza para cada vóxel para toda la serie temporal de forma independiente, de esta manera vemos el **tCNR** espacial. En este caso esta relación se ha calculado siguiendo la ecuación siguiente que como se puede ver muestra los resultados en decibelios:

$$tCNR = 10 * \log_{10} \frac{A^2}{\sigma^2}$$

Figura 41: Ecuación para el cálculo del tCNR

Ahora, en el siguiente gráfico podemos ver la representación en 16 cortes del **tCNR** por vóxel. Teóricamente las zonas con un valor más elevado podrían tener una mayor activación debido a que la imagen es una (**fMRI**) y la ecuación se basa en la fuerza de la señal.

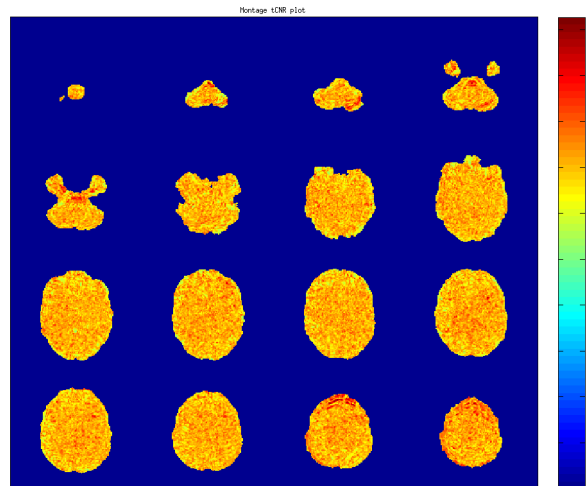


Figura 42: Ejemplo tCNR imagen funcional

Por otro lado, no hemos encontrado un umbral claramente diferenciador entre un buen y un mal contraste. Al igual que en el caso de **tSNR** también se puede visualizar la distribución de los valores en los tres modos (gráfico de cajas, texto correspondiente al gráfico de cajas e histograma).

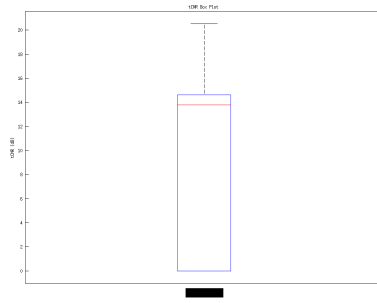


Figura 43: Distribución del tCNR de la imagen funcional.

Finalmente, hemos decidido comparar la evolución del **tSNR** y el **tCNR**. Recordemos que el **tSNR** cuanto más alto sea mejor será la imagen y lo contrario para el **tCNR**. De este modo, una imagen de buena calidad suele tener la línea correspondiente al **tSNR** por encima de la del **tCNR**. Otra opción cuantitativa podría ser obtener el promedio para ambos valores. a continuación mostramos la comparación entre una imagen de mejor calidad con otro de peor calidad.

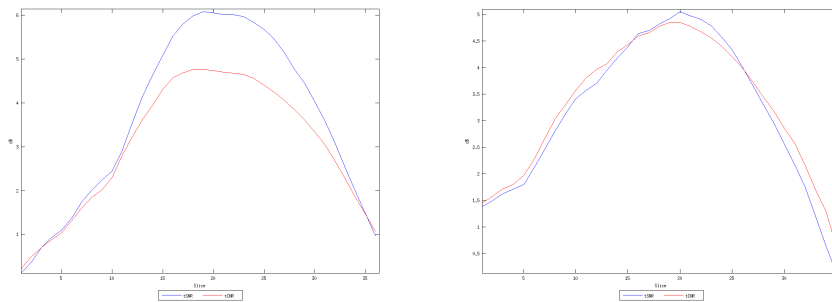


Figura 44: Imagen de mejor calidad contra una de mala calidad

11.5 Estimación de la heterogeneidad del campo de polarización

Un artefacto de susceptibilidad se observa como un área de color oscuro que de manera anormal aparece en la imagen. La zona oscura está limitada normalmente mediante una banda brillante, un área de ondas luminosas y oscuras u otras distorsiones en la anatomía circundante. Una causa común de este artefacto es el metal. Otras sustancias pueden ser magnetizadas en diferente grado (por lo que su susceptibilidad magnética es diferente) y por tanto, cuando se someten a un campo uniforme la imagen captada aparece distorsionada mediante una (pérdida de señal) o es demasiado brillante (hiperintensidad).

El problema que nos ocupa en este caso es muy parecido al anterior la diferencia con este último radica en que la diferencia de magnetización no está causada por ningún elemento en el paciente sino un problema dentro de la máquina de adquisición. Las imágenes de resonancia magnética utilizan una bobina receptora para recoger la señal de radiofrecuencia utilizada para crear la imagen. Si la bobina receptora no es sensible de manera uniforme, se lee incorrectamente obteniendo una señal con

más intensidad en algunas zonas y más débilmente en otras. La homogeneidad de la señal puede variar notablemente entre las diferentes bobinas receptoras ya que nunca son 100% homogéneas. En la imagen inferior se puede observar las diferencias y semejanzas de ambos artefactos.

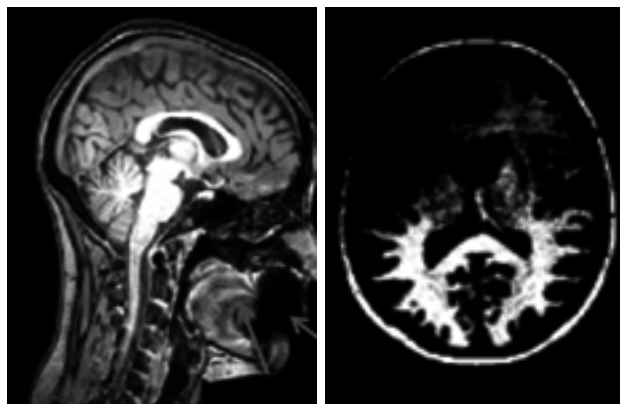


Figura 45: Diferente susceptibilidad (izquierda) contra heterogeneidad de campo (derecha)

Por todo lo comentado, es crucial identificar la heterogeneidad del campo. Es por ello, que es necesario estimarla para poderla corregirla en la imagen y avisar a la entidad productora de las imágenes del problema observado en el caso de que la heterogeneidad sea grave. Para la estimación de la heterogeneidad se ha utilizado parte de un algoritmo descrito en '**A nonparametric MRI inhomogeneity correction method. Medical Image Analysis**'[1] perteneciente a **Jose Vicente Manjón**. En este artículo se propone una función de coste relacionada con la información contenida (entropía) en combinación con características de intensidad y gradiente de la imagen. Dicho código está desarrollado en **Matlab**, su última versión es del 2013 y está disponible públicamente en <http://personales.upv.es/jmanjon/bias/index.htm>. Bajo estas líneas se puede observar la estimación de heterogeneidad obtenida para la imagen estructural suministrada junto con la herramienta.

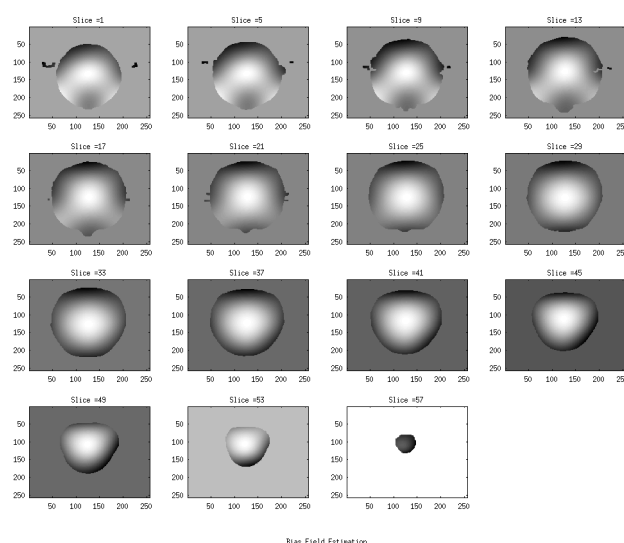


Figura 46: Capturas de la estimación de la heterogeneidad de campo a partir de la imagen estructural cedida por José Vicente Manjón (UPV)

11.6 Espacio K

El espacio k representa señales de voltaje en función de coordenadas de frecuencia espacial (K_x codifica las frecuencias y K_y se encarga de la codificación de fases) mientras que el espacio de la imagen recoge intensidad de señal en función de coordenadas espaciales x e y . La herramienta matemática que relaciona ambos espacios es la **transformada de Fourier**. La matriz de datos de una imagen tiene el mismo tamaño que la matriz correspondiente de datos en el espacio k. No obstante, la información de cada píxel de la imagen está formada por todos los puntos del espacio k.

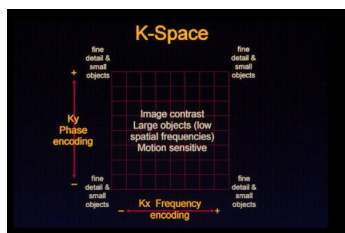


Figura 47: Esquema espacio k

Los datos del espacio-k se pueden representar como una imagen 2D en el centro de la imagen se sitúan las bajas frecuencias y en los extremos y en la periferia se sitúan las altas frecuencias. Las bajas frecuencias contienen la información correspondiente a la forma general de la imagen es decir, muestran el contenido o el contraste de la imagen. Por otro lado, las altas frecuencias representan los detalles más finos de la imagen así como el contorno de los objetos. La mayor parte de la información de una imagen de resonancia magnética se encuentra en las bajas frecuencias por lo que unos valores elevados en zonas no centrales son señal de algún artefacto. En la figura inferior podemos ver la correlación existente entre algunos artefactos en la imagen y el espacio k.

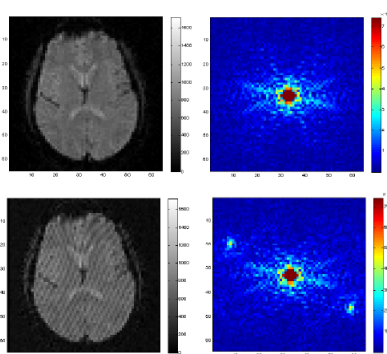


Figura 48: Imagen contra espacio K

En el caso de nuestra aplicación, pasamos del dominio del tiempo al dominio de frecuencia mediante la **FFA (Fast Fourier Transform)**. Esta operación se realiza para cada corte y/o volumen/es dependiendo si la imagen es 2D, 3D o 4D). En la imagen inferior se muestra un subconjunto de todas los cortes del volumen de la imagen estructural cedida por José Vicente Manjón, si la imagen fuera

4D el subconjunto de cortes correspondería al instante de tiempo central. No obstante el sistema realiza esta operación para todos los cortes de cada volumen para toda la serie temporal, lo guarda en formato **NIFTI** y lo deja disponible en la web para su descarga. De esta forma, los usuarios pueden bajar el espacio **K** completo y revisarlo manualmente en busca de valores atípicos.

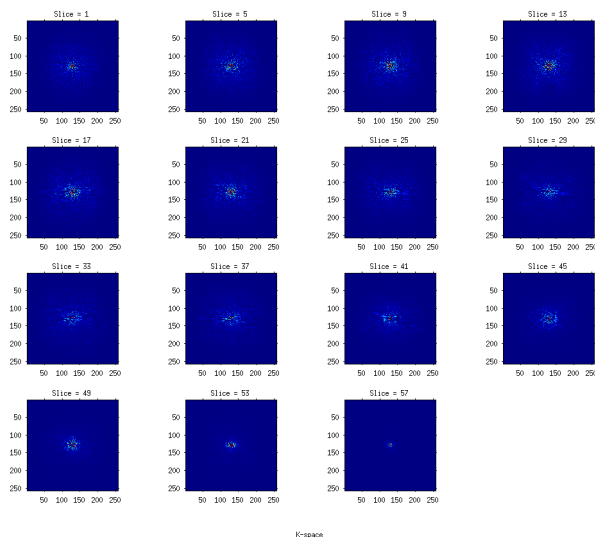


Figura 49: Capturas del espacio **K** del volumen de la imagen cedida por José Vicente Manjón (UPV)

12 Implementación de algoritmos y herramientas de anonimización

12.1 Anonimización de cabeceras DICOM

La anonimización de las cabeceras **DICOM** se ha realizado en el nodo **CEIBANON** mediante la pipeline de anonimización de cabeceras según los perfiles del estándar **DICOM** de la **National Electrical Manufacturers Association (NEMA)**.

La antigua pipeline funcionaba con unos perfiles preconfigurados y no actualizados. Nuestro sistema se basa en una serie de scripts en Python que son capaces de extraer de la página del estándar <http://dicom.nema.org/> las cabeceras presentes en el mismo y los diferentes perfiles de anonimización aplicables y las acciones a aplicar dependiendo del estándar seleccionado. Toda esta información es guardada en ficheros **.json** debido a su fácil manejo y eficiente uso.

Para cada proyecto se genera un script de anonimización en el cual se debe indicar los perfiles del estándar a aplicar. Seguidamente se muestran los diferentes perfiles disponibles en el **apartado 15 del estándar titulado Security and System Management Profiles, en su anexo E**:

- **Basic Profile**: es el perfil básico, obligatorio en nuestra aplicación y más restrictivo
- **Retain Safe Private**: contiene información propietaria del fabricante conocida sólo por el mismo. Ciertas técnicas podrían hacer uso de estos campos.
- **Retain UIDs**: se guardan algunos identificadores sobre el estudios, series, instancias, etc. Esos identificadores no pueden identificar directamente a un individuo fuera de contexto no obstante

si se dispone del apoyo de una base de datos que relacione ambos se podría identificar al sujeto. Sin embargo, existen casos de uso que requieren la capacidad de mantener la relación del sujeto con las imágenes originales. En definitiva, esta opción se usa cuando se sopesa que el riesgo de ganar acceso a la información original a través de los identificadores es bajo con respecto al beneficio de retenerlos. Siempre con el requisito previo de la aplicación y aceptación del contenido del **consentimiento informado** limitando su uso a la investigación o docencia.

- **Retain Device Identification:** La información identificativa de la máquina usada se considera de riesgo debido a que limita las personas a las cuales pertenece una imagen. No obstante, hay aplicaciones que requieren de la identificación del dispositivo como por ejemplo cuando se produce heterogeneidad de campo debido al mal estado de las bobinas de la máquina es necesario identificarla para repararla.
- **Retain Patient Characteristics:** se mantiene información privada pero no identificativa por sí misma de la persona. Teniendo como requisito previo la aplicación y aceptación del contenido del **consentimiento informado** limitando su uso a la investigación o docencia.
- **Retain Longitudinal Temporal Information With Full Dates:** se mantienen todas las fechas y horas, se debe añadir la etiqueta (0028,0303) con el valor "UNMODIFIED".
- **Retain Longitudinal Temporal Information With Modified Dates:** en este caso las fechas y horas del estándar son modificadas. La modificación debe ser llevada de forma que:
 - modifica la fecha o la transforma.
 - mantener la relación en distancia temporal en los estudios longitudinales.
 - mantener la relación entre las imágenes y los acontecimientos del mundo real que permite extender el análisis de las imágenes.

Se debe añadir la etiqueta (0028,0303) con el valor "MODIFIED".

- **Clean Descriptors:** se deben revisar las cabeceras con contenido de texto libre para limpiar los datos identificativos.
- **Clean Structured Content:** en este caso cualquier información en las etiquetas de tipo secuencia deben de ser limpiadas como por ejemplo Acquisition Context o Specimen Preparation Sequence Items.
- **Clean Graphics:** se limpia la información que aparece en la imagen, anotaciones, texto superpuesto, etc.

Las acciones disponibles para cada uno de las etiquetas sensibles que contiene el estándar son:

- **D:** sustituir contenido por un valor no nulo de longitud igual o diferente al valor sustituido.
- **Z:** reemplazar contenido por un valor nulo de longitud cero, o por un valor ficticio y coherente con el contenido del valor sustituido.
- **X:** borrar contenido.
- **K:** mantener el contenido para etiquetas que no son secuencias y limpiar para etiquetas que sí lo son. Limpiar corresponde a reemplazar los datos por valores ficticios en el caso que sea necesario.
- **C:** limpiar campos, es decir reemplazar contenido con valores de significado similar y coherente con el **VR (Value Representation)** del dato.
- **U:** sustituir contenido por un UID diferente al contenido pero internamente sea consistente dentro de un conjunto de instancias. En este caso se han utilizado las acciones con las que trabaja el CTP para estos casos.

- Z o D: Se aplica Z antes que D.
- X o Z: Se aplica X antes que Z.
- X o D:- Se aplica X antes que D.
- X o Z o D: Se aplica X antes que Z y Z antes que D.
- X o Z o U: Se aplica X antes que Z y Z antes que U.

En el caso en el que hay diferentes opciones a aplicar se está utilizando la más restrictiva. Por otro lado, la operación de limpiado se ha comprobado el VR del dato y se ha sustituido el valor por uno consecuente en el caso de que fuera posible, en cualquier otro caso se ha borrado el contenido.

También fueron eliminados los campos fuera del estándar en el caso de que no fueran necesarios para el uso de la imagen. Este es el caso de las 'imágenes por Tensor de Difusión' (DTI) en las que cada fabricante guarda parte de los datos necesarios para su uso en campos no estándar, los cuales, sí que son guardados.

Finalmente, el script guarda el nombre del proyecto en la etiqueta (0008,1030) 'Study Description', pues posteriormente es necesaria para asignar la imagen a un proyecto en GEBID. En la figura inferior podemos ver un ejemplo del código generado por el script.

```
<e en="T" t="0040DB0D" n="Template Extension Creator UID">@keep()</e>
<e en="T" t="00101010" n="Patient's Age">@keep()</e>
<e en="T" t="00102180" n="Occupation">@remove()</e>
<e en="T" t="4008010A" n="Interpretation Transcriber">@remove()</e>
<e en="T" t="00189517" n="End Acquisition DateTime">@keep()</e>
<e en="T" t="00400244" n="Performed Procedure Step Start Date">@keep()</e>
<e en="T" t="00100030" n="Patient's Birth Date">@keep()</e>
<e en="T" t="00082112" n="Source Image Sequence">@keep()</e>
<e en="T" t="00101081" n="Branch of Service">@remove()</e>
<e en="T" t="00204000" n="Image Comments">@remove()</e>
<e en="T" t="00101040" n="Patient Address">@remove()</e>
<e en="T" t="00102297" n="Responsible Person">@remove()</e>
<e en="T" t="00400280" n="Comments on the Performed Procedure Step">@remove()</e>
<e en="T" t="00380300" n="Current Patient Location">@remove()</e>
<e en="T" t="00400003" n="Scheduled Procedure Step Start Time">@keep()</e>
<e en="T" t="04000402" n="Referenced Digital Signature Sequence">@remove()</e>
<e en="T" t="00321032" n="Requesting Physician">@remove()</e>
<e en="T" t="00404052" n="Procedure Step Cancellation DateTime">@keep()</e>
<e en="T" t="00080020" n="Study Date">@keep()</e>
<e en="T" t="00402010" n="Order Callback Phone Number">@remove()</e>
<e en="T" t="00081048" n="Physician(s) of Record">@remove()</e>
<e en="T" t="00404034" n="Scheduled Human Performers Sequence">@remove()</e>
<e en="T" t="00403001" n="Confidentiality Constraint on Patient Data Description">@keep()</e>
<e en="T" t="00080014" n="Instance Creator UID">@keep()</e>
<e en="T" t="0070031A" n="Fiducial UID">@keep()</e>
<e en="T" t="0008010D" n="Context Group Extension Creator UID">@keep()</e>
<e en="T" t="00081120" n="Referenced Patient Sequence">@remove()</e>
<e en="T" t="4008010C" n="Interpretation Author">@remove()</e>
<e en="T" t="40004000" n="Text Comments">@remove()</e>
<e en="T" t="00181400" n="Acquisition Device Processing Description">@remove()</e>
<e en="T" t="00700086" n="Content Creator's Identification Code Sequence">@remov
<e en="T" t="00400242" n="Performed Station Name">@keep()</e>
<e en="T" t="00101030" n="Patient's Weight">@keep()</e>
```

Figura 50: Resultado de instrucciones de anonimización para el CTP

12.2 Anonimización sobre la propia imagen

No sólo las cabeceras de la imagen contienen información identificativa del sujeto. Por ejemplo, a partir de imágenes tomográficas como la resonancia magnética estructural se puede recomponer el aspecto externo de una persona lo que se vuelve crítico si la resonancia afecta al rostro.

Otro caso de posible identificación a partir de la imagen es la información que puede aparecer grabada sobre la propia imagen (un caso típico son las ecografías).

Por estos dos motivos se ha desarrollado un script propio para la anonimización automática de texto sobre la imagen y un script de anonimización de rostros tanto en **NIFTI** como en **DICOM**.

12.2.1 Anonimización de ecografías

La información personal puede aparecer en capas distintas de la imagen clínica. Esta situación está contemplada en el estándar **DICOM**. Por tanto, se puede eliminar dicha información mediante el borrado del contenido de cierta información **DICOM** gracias al perfil **Clean Graphics**.

No obstante, si toda la información aparece mezclada dentro de una única imagen se dice que la imagen médica contiene información quemada.

Una solución que puede hacer frente a estos casos es la que presenta el **CTP**. Con el módulo '**Pixel Anonymizer**' se pueden definir regiones fijas de la imagen a ser borradas lo que conlleva un conocimiento minucioso de la localización de la información a anonimizar (asume que la información está siempre en la misma posición). En nuestro caso el origen de las imágenes es de muchas máquinas y fabricantes lo que hace difícil aplicar esta solución.

Otras herramientas como '**Dicom Cleaner**' permiten seleccionar a mano zonas rectangulares de la imagen. El programa se limita rellenar dichas zonas de negro cuando el usuario le indica que la imagen ha sido limpiada. Esta forma parece mucho más manual que la del **CTP** y parece ser que no asume que la siguiente imagen vaya a ser igual a la anterior lo que hace que se tenga que repetir el trabajo imagen a imagen.

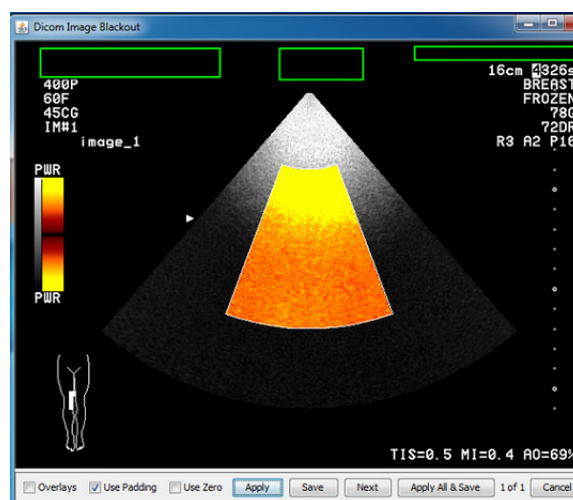


Figura 51: Dicom Cleaner Image Blackout

En cuanto a la solución que nosotros hemos aplicado está basada en la segmentación del texto del resto de la imagen. Para ello se han utilizado regiones de interés seleccionadas una función de **Matlab**. Con la información obtenida se han ido aplicando criterios para descartar **regiones de interés (ROI)** hasta quedarnos sólo con el texto de la imagen. A partir de este texto se ha generado una máscara que lo ha eliminado de la imagen. Por otro lado, para la selección de los umbrales de cada criterio se ha utilizado la distribución de los datos basándose en cuartiles. Finalmente hay que señalar que las imágenes utilizadas han sido cedidas por el **Oceanogràfic** gracias a **Daniel García Parraga** y corresponden a los delfines de sus instalaciones.

Para cada imagen de un archivo **NIFTI** se han realizado las siguientes tareas:

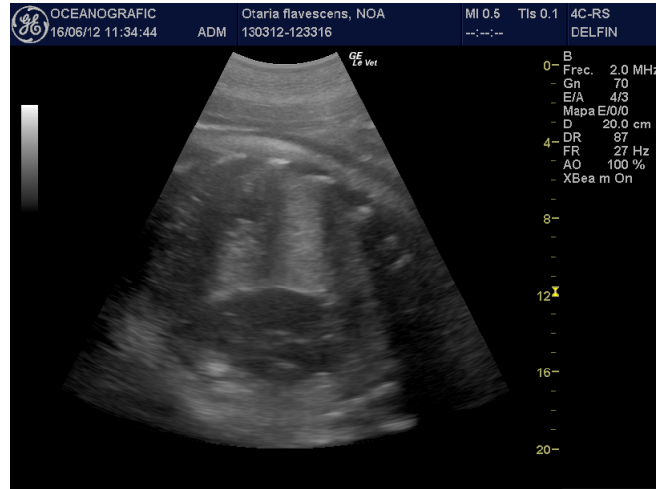


Figura 52: Imagen original

- Primeramente se han eliminado las **ROIs** con un tamaño mayor que un 68% de los datos, pues en una ecografía pertenecen a los datos gráficos ecográficos. También podemos ver como la imagen ha pasado a ser del tipo blanco y negro.

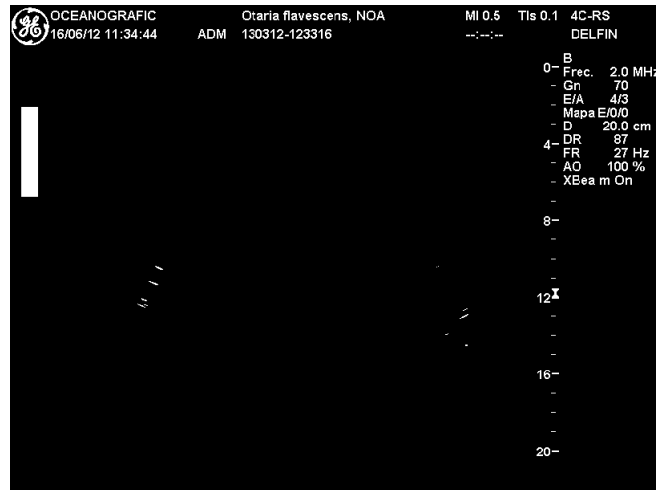


Figura 53: Imagen sin las áreas más grandes de la imagen.

- Seguidamente, como sabemos que el texto debe seguir más o menos una misma proporción de altura se han eliminado aquellos datos cuya altura era menor que el primer cuartil de las **ROIs**, es decir los elementos de menor altura.

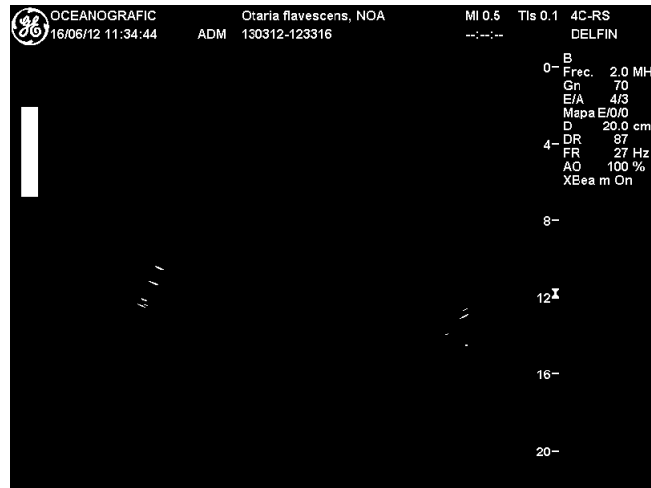


Figura 54: Imagen después de aplicar el criterio de proporción

- De las **ROI**s restantes se han calculado los centroides de las mismas y se han tomado distancias (euclídeas) entre todas ellas. Estos datos han sido usados para eliminar las **ROI**s que tienen menos vecinos espaciales. El proceso se ha repetido varias veces pues eliminar una **ROI** en una iteración reduce el número de vecinos cercanos que pueden aparecer en la siguiente (parando este proceso en caso de convergencia o al final de diez iteraciones).

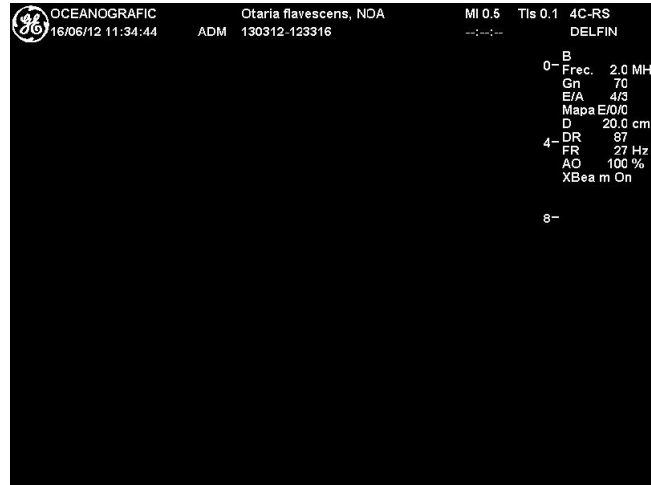


Figura 55: Imagen después del proceso de segmentación

- Finalmente se han enmarcado en cajas rojas los datos eliminados para que el encargado de revisar las imágenes pueda detectar problemas rápidamente y confirmar la anonimización.

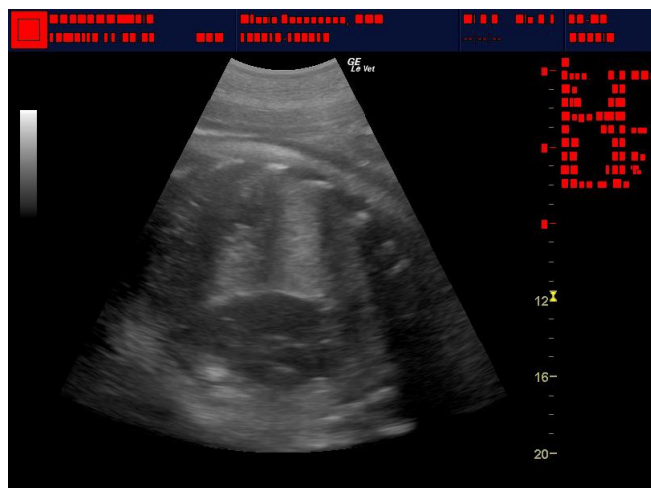


Figura 56: Imagen original con zonas anonimizadas

Este trabajo pensamos que podría mejorarse mediante **OCR (Optical Character Recognition)** sobre el texto segmentado con el objetivo de borrar únicamente datos identificativos. Aunque esta es una idea que necesita ser pensada más detenidamente debido a la gran variedad de datos identificativos que pueden presentarse. Otro aspecto que cabe señalar es que aunque esta técnica ha sido probada con ecografías creemos que puede ser trasladada a radiografía convencional.

12.2.2 Eliminación de facciones faciales

La eliminación de rasgos faciales es un problema sobre todo en las técnicas tomográficas que permiten la renderización del volumen completo sobre todo cuando afectan a rasgos identificadores de los sujetos. Algunos grupos de investigación solucionan este problema mediante la aplicación de algoritmos de extracción del parénquima como es el caso del **FSL BET**.

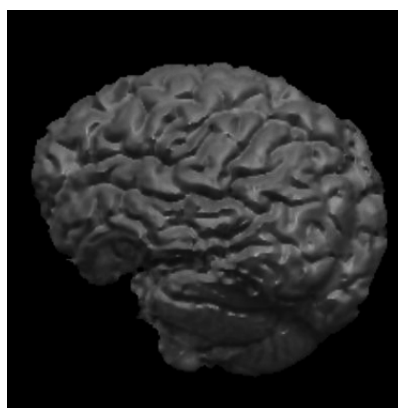


Figura 57: Renderizado después de la extracción del parénquima

No obstante, para el estudio y procesado de **electroencefalogramas (EEG)** y de **magnetoencefalogramas (MEG)** combinados con **resonancias magnéticas funcionales (fMRI)** es necesario

mantener las orejas durante el protocolo de desfiguración del rostro pues las orejas suelen servir como puntos de referencia.

Por otro lado, la solución de la **NRG** es el uso de la aplicación Face Masking, la cual oculta las facciones faciales de los scans de resonancia magnética con el objetivo de reducir el riesgo de identificación del sujeto. Unas de sus mejores características es que es mínimamente invasivo y diseñado para funcionar con los otros sistemas desarrollados por la **NRG**. Se fundamenta en la herramienta **FSL** y es posible utilizarla separadamente del **XNAT**. No obstante, se suministra junto con un ejemplo con el objetivo de integrarla dentro de la plataforma. Como elemento en contra tiene la escasa documentación de instalación que intenta simular una facilidad de instalación que realmente no existe. En resumen esta herramienta parecía muy prometedora pero no hemos conseguido que funcione en nuestro sistema.

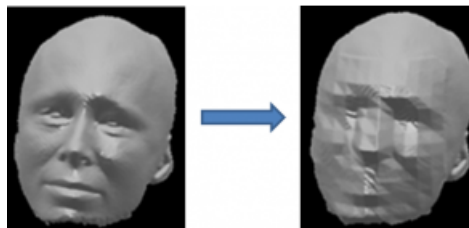


Figura 58: Resultado del Defacing según Face Masking de NRG

Nuestra solución se ha basado en la aplicación sobre la imagen en **NIFTI** el **Defacing** implementado por **SPM**. Una vez obtenida la imagen en **NIFTI** sin rostro es necesario trasladar estos resultados a la imagen **DICOM** pues no sirve para nada anonimizar un formato sin trasladar el resultado a los demás. De este modo se han sustituido los contenidos de la imagen **DICOM** original por los desfigurados manteniendo las cabeceras **DICOM** referentes a otras secciones. Seguidamente se muestra la imagen en sus tres planos principales (coronal, sagital y axial) y su renderización. Podemos ver que de este modo conservamos las orejas pero ocultamos la identidad de la persona.

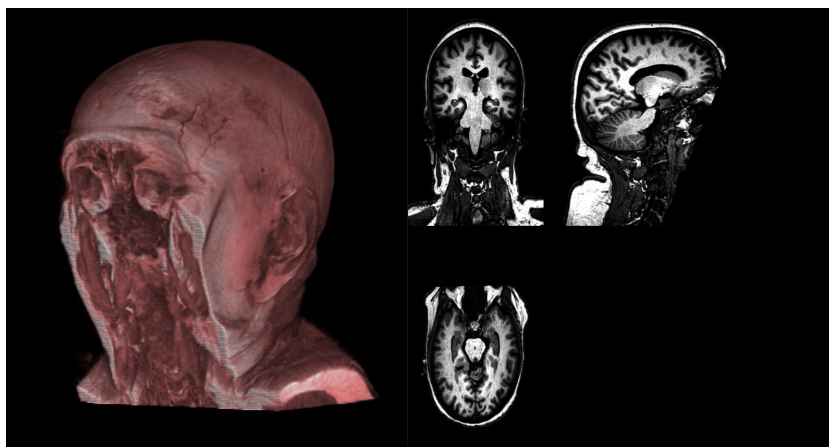


Figura 59: Renderizado de la imagen sin rostro y visión en los tres planos (coronal, sagital y axial)

13 Creación de sistema de presentación visual de resultados e integración en GEBID

El último paso de la pipeline consiste en presentar de forma amigable al usuario los resultados obtenidos. Con este objetivo, se hace uso de **Bootstrap** el cual es un framework que facilita el desarrollo web multidispositivo y establece un sistema de plantillas para la página web, las hojas de estilo **CSS** las cuales ayudan a formatear el aspecto de los contenidos y la librería de **JavaScript jQuery** que facilita la rápida implementación del código de ensamblado de la web. De esta forma al hacer click en el enlace a los resultados un código se ejecuta en el navegador web del usuario autogenerando de forma automática una página web mediante la carga de un fichero .json que actúa de índice con los contenidos gráficos y textuales a visualizar. Este fichero json ha sido previamente creado durante la ejecución de los scripts de control de calidad, incluye:

- Títulos
- Direcciones a las imágenes en el servidor
- Los datos numéricos del control de calidad cuantitativo
- Descripciones de cada una de las técnicas utilizadas

The screenshot displays the XNAT web interface for a specific MR session. The header includes the logo of the Generalitat Valenciana and navigation menus. The main content area shows the MR Session ID: 104638564041574873013745558076671369035. Below this, there are tabs for 'Details' and 'Projects'. The 'Details' tab is active, showing a table of scan parameters:

Accession #:	ceibcs_E01663	Subject:	62048358155769569994966808337176457304
Date Added:	07/20/2015 09:40:18 (jlsnardo)	Gender:	
Date:	03/26/2014	Handedness:	
Time:	09:25:12	Age:	--
Scanner:	SJRM GE MEDICAL SYSTEMS Optima MR450w		

To the right of the details is an 'Actions' menu with options: Edit, View, Download, Email, Pipelines, Manage Files, and Delete. Below the details is a 'Scans' table:

Scan	Type	Series Desc	Usability	Files	Note
1	LOCALIZER	LOCALIZER	usable	4.5 MB in 22 files	Quality Control (QC)
3	SagT1FLAIR	SagT1FLAIR	usable	33.9 MB in 30 files	

Under the 'SagT1FLAIR' row, there is a grid of MRI scan images. Below the images, the text 'Quality Series Desc' is visible, with 'usable' and 'SagT1FLAIR' listed. A 'Quality Control (QC)' link is also present at the bottom right of the image grid.

Figura 60: Integración del control de calidad dentro de XNAT

14 Presentación en congresos y premios

El comité científico de la **Asociación Valenciana de Informáticos Sanitarios (AVIS@)** en el IV concurso de pósters celebrado en la **VII Jornada Técnica AVIS@: "Movilidad, Seguridad, Privacidad y BI en Sanidad"** otorgó el primer premio al póster '**Implementación de los perfiles de confidencialidad del estándar DICOM y Control de Calidad (QC) en Imagen Médica Digital para proyectos de I+D+I**' presentado por el CEIB junto con sus colaboradores. Dicho póster está íntimamente relacionado con este trabajo final de máster y ha sido realizado por María de la Iglesia Vayá¹, José María Salinas², José Miguel Puig Saqués³, Diego Guijarro⁴, Jorge Isnardo Altamariano² y Angel Fernández-Cañada Vilata².

Las instituciones que han colaborado en este proyecto han sido: ¹Unidad mixta FISABIO y Centro de Investigación Príncipe Felipe (CIPF), Valencia, España, ²Centro de Excelencia en Imagen Biomédica (CEIB), Conselleria de Sanitat de la Comunitat Valenciana, Valencia, España, ³Sistemes d'Informació per a la Salut de la Conselleria de Sanitat de la Comunitat Valenciana, ⁴ Hospital San Juan de Alicante.

Implementación de los perfiles de confidencialidad del estándar DICOM y Control de Calidad (QC) en Imagen Médica Digital para proyectos de I+D+i

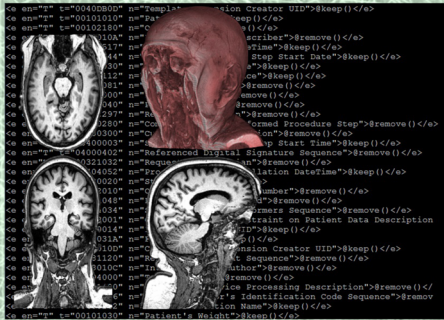
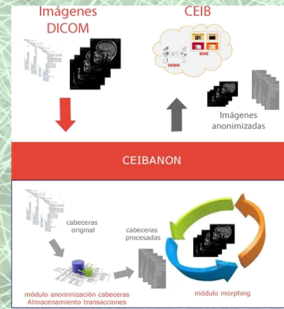
María de la Iglesia, José María Salinas, Diego Guijarro, José Miguel Puig

Introducción

En la actualidad, la imagen biomédica es una de las disciplinas y eje central de la medicina debido al incremento del número de las patologías que pueden ser diagnosticadas y tratadas con su ayuda.

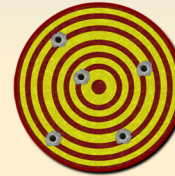
En este marco, uno de los problemas fundamentales con los que se encuentran los investigadores cuando se aplican técnicas de análisis de imagen biomédica es la heterogeneidad y la mala accesibilidad a las infraestructuras existentes o incluso el desconocimiento de las mismas. No obstante, almacenar la información y ofrecerla al mundo sólo es uno de los primeros pasos. También es importante que los contenidos almacenados posean indicadores de calidad para informar a los investigadores sobre la naturaleza de los datos. Del mismo modo es fundamental que los datos estén debidamente anonimizados, ya que se trata de un derecho fundamental de las personas que cedieron los mismos.

La Subdirección General de Sistemas de Información para la Salud, está desarrollando soluciones para dar cobertura a esta problemática, a través de BIMCV (Banco de imágenes de la CV).



Objetivos

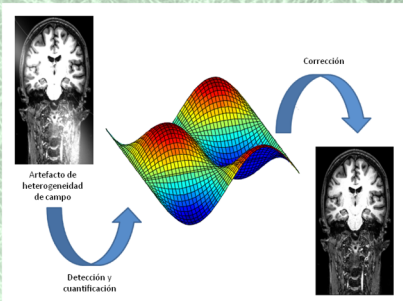
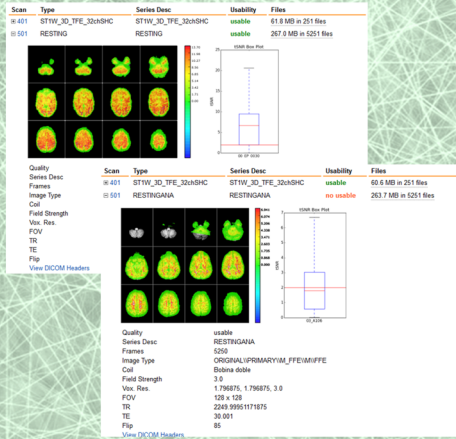
- ✓ Dotar a los investigadores de un mínimo conjunto de indicadores que garanticen la calidad de las imágenes depositadas cada proyecto en marcha.
- ✓ Proteger el derecho de confidencialidad de las personas que cedieron las imágenes.
- ✓ Integración en GEBID-CS de las herramientas mencionadas.



Metodología

Con el objetivo de obtener marcadores de calidad de las imágenes almacenadas se han implementado aplicaciones en Matlab basadas en la plataforma "Configurable Pipeline for the Analysis of Connectomes" y "Data Processing & Analysis of Brain Imaging" aunando lo mejor de ambos mundos.

Para asegurar el anonimato de los sujetos se han implementado los perfiles de confidencialidad que establece estándar DICOM para la anonimización de las cabeceras DICOM que suministran información sensible, además se están desarrollando técnicas de anonimización de imagen como el "defacing" para eliminación digital del rostro en caso que sea necesario y dependiendo de la valoración de los comités éticos en relación al proyecto en curso.



Trabajo futuro

El desarrollo de las herramientas mencionadas supone una nueva base para la mejora de las herramientas disponibles en la plataforma GEBID-CS. No obstante, el progreso de actualización de la plataforma GEBID-CS es un proceso constante que en un cercano futuro dotará a los investigadores de nuevos mecanismos de evaluación y adecuación de la imagen médica. En este orden de cosas se pretende obtener nuevos marcadores de calidad y detección automática de anomalías en la imagen así como herramientas de procesado y mejora de imagen.

Trabajo desarrollado por Ángel Fernández-Cañada Vilata y Jorge Isnardo Altamirano

VII Jornada Técnica AVIS@ - "Movilidad, Seguridad, Privacidad y BI en Sanidad" - Valencia, 13 de junio de 2015

Figura 61: Implementación de los perfiles de confidencialidad del estándar DICOM y Control de Calidad (QC) en Imagen Médica Digital para proyectos de I+D+i

Junto a las ideas expuestas en el anterior póster se añadió otra de las líneas de estudio del grupo en sinergia con lo anterior , estas son el **estudio poblacional** y el **Big Data**.

En el congreso de la '**Organization for Human Brain Mapping**' celebrado en el año **2015 en Honolulu** se expuso el trabajo del grupo (infraestructura diseñada, Big Data, anonimización, control de calidad,etc.) mediante el póster '**BIMCV: Synergy between data in population medical imaging, computer aided diagnosis and AVR**'. María de la Iglesia¹, Angel Fernández-Cañada Vilata², Jorge Isnardo Altamirano², F. Xavier Castellanos³, Gonzalo Rojas Costa⁴, Erika Proal⁵, José María Salinas² , Jacobo Martínez⁶.

Las instituciones que han colaborado en este proyecto han sido: ¹Unidad mixta FISABIO y Centro de Investigación Príncipe Felipe (CIPF), Valencia, España, ²Centro de Excelencia en Imagen Biomédica (CEIB), Conselleria de Sanitat de la Comunitat Valenciana, Valencia, España, ³Phyllis Green and Randolph Cōwen Institute for Pediatric Neuroscience, NYU Langone Medical Center, New York, NY, ⁴Advanced Medical Image Processing Lab, Department of Radiology, Clínica las Condes, Santiago, Chile, ⁵NEUROingenia, mexico city, Mexico, ⁶Fundación para el Fomento de la Investigación Sanitaria de la Comunitat Valenciana, FISABIO, Valencia, Spain.

15 Conclusiones

Tomando en consideración los objetivos propuestos y los resultados obtenidos, se puede afirmar que se ha cumplido con el objetivo principal de ampliar el conjunto de herramientas de **BIMCV**, del tal forma que es posible formular las siguientes conclusiones:

1. El conjunto de marcadores seleccionados, adaptados e implementados, propiciará el buen uso de las imágenes y la toma de conciencia de la importancia de la adquisición de las mismas así como su selección para un estudio. De esta forma, creemos que esta nueva línea de presentar la información en los bancos de imágenes contribuirá a la mejora de la calidad de las investigaciones y con ello al porcentaje de éxito en la traslación de los resultados obtenidos en el laboratorio a la práctica clínica. Sabemos que la búsqueda de marcadores más útiles acaba de empezar y somos conscientes de la poca facilidad de interpretación de algunos de los marcadores propuestos como en algunos casos de sus valores umbrales entre la buena y mala calidad de la imagen.
2. Los algoritmos y scripts de anonimización generados serán un exponente ante la falta de conciencia sobre el derecho de preservación de identidad de los pacientes. De esta forma el **CEIB-CS** será uno de los abanderados de la defensa de este derecho y servirá de ejemplo para otras instituciones. Somos también conscientes de que es un arduo camino a recorrer y que la mejora en nuestros algoritmos de anonimización automática deben ser robustecidos, pues perder datos clínicos es muy crítico. No obstante, sabemos por este motivo que son pocos algoritmos automáticos como el nuestro los que se pueden encontrar . Esto revaloriza la pequeña meta alcanzada y nos anima a seguir mejorando. Del mismo modo, la anonimización de cabeceras realizada ha resultado robusta en **RMI** y del mismo modo sabemos que es flexible. No obstante tenemos el reto de comprobar si dichos algoritmos podrán ser utilizados en otras técnicas de adquisición de imagen.
3. El sistema del **CTP** para el envío anonimizado y con detección de errores mejorará la usabilidad del banco de imágenes pues en la mayoría de los casos se evita de esta forma el almacenamiento en la plataforma de imágenes inservibles o corruptas.
4. El uso de software de código abierto, ha demostrado ser capaz de afrontar los objetivos de este trabajo, siendo una opción de flexibilidad manifiesta capaz de adaptarse a los fines particulares perseguidos. No obstante, el uso de este tipo de software plantea nuevos retos tales como la particularización y mantenimiento de los proyectos adaptados y la posible deriva paralela del desarrollo del software libre original.
5. La elaboración de este trabajo ha ofrecido un gran valor añadido en concepto de formación, tan requerida a nivel profesional a un ingeniero biomédico destinado en un departamento de imagen médica. No solo por haber implementado de forma satisfactoria herramientas útiles en este ámbito, sino también por haber sido capaz de emplear, adaptar y contribuir a la comunidad mundial desarrolladora con nuevo código abierto.
6. Las aplicaciones de imagen médica han demostrado conllevar un gran gasto en recursos, no sólo por la adquisición de la propia imagen, sino también por el almacenamiento para guardar dichas imágenes , ancho de banda para transmitir las y cómputo para procesarlas. La eficiencia alcanzada puede considerarse buena en cuanto a que mayormente utilizamos software generado por nosotros mismos o gratuito. No obstante, recortar en almacenamiento o cómputo es mucho más complicado. La colaboración entre organizaciones nacionales e internacionales parece ser la única salida viable en estos momentos.

16 Trabajo futuro

- Ampliar el control de calidad a más secuencias y técnicas de adquisición de imagen.
- Selección de nuevos y mejores marcadores de calidad.
- Cribado de indicadores redundantes o de poca utilidad.
- Simplificar los resultados acercando el conocimiento al personal consumidor (no expertos en imagen) de **GEBID**.
- Integración de visores completos para las imágenes almacenadas.
- Minimizar la agresividad de la anonimización por deformación facial.
- Mejorar el tiempo de cómputo mediante la paralelización y uso de clusters.
- Mejorar la presentación de resultados.
- Mejorar la integración de nuestros desarrollos dentro de **GEBID**.
- Comenzar a explotar y extraer conocimiento de los datos almacenados.
- Ampliar la capacidad de almacenamiento y cómputo.
- Ampliar la seguridad con nuevas medidas para la transmisión de la información.
- Migrar aplicaciones que funcionan con software de pago a software libre y gratuito.
- Evaluar la usabilidad de la aplicación con los usuarios objetivo y captación de opiniones mediante encuestas.

17 Bibliografía

1. Manjon JV, Lull JJ, Carbonell-Caballero J, Garcia-Marti G, Mart-Bonmati L, Robles M. A nonparametric MRI inhomogeneity correction method. *Med Image Anal.* 2007 Mar; 11:336–345
2. Power JD, Schlaggar BL, Petersen SE. Recent progress and outstanding issues in motion correction in resting state fMRI. *Neuroimage.* 2015 Jan 15;105:536-51
3. Power J.D; Barnes K.A; Snyder A. Z; Schlaggar B. L; Petersen S. E. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *Neuroimage.* 2012; 59(3):2142-54
4. Pedersen M; Rizzi A; Hardeberg J.Y; Simone G. Evaluation of contrast measures in relation to observers perceived contrast. *Society for imaging science and technology. Conference on Colour in Graphics, Imaging, and Vision.* 2008; p. 253-8
5. Marcus D.S., Olsen T; Ramaratnam M; Buckner R. L. XNAT: a software framework for managing neuroimaging laboratory data. *Organization for Human Brain Mapping. Proceedings of the 12th Annual Meeting. Florence (Italy) , ed Corbetta M.* 2006
6. Marcus D.S; Olsen T.R; Ramaratnam M; Buckner R.L. The extensible neuroimaging archive toolkit: an informatics platform for managing, exploring, and sharing neuroimaging data. *Neuroinformatics* 5, 2007; 11–34
7. Woods R.P; Cherry S.R; Mazziotta J.C. Rapid automated algorithm for aligning and reslicing PET images. *Journal of computer assisted tomography.* 1992; 16(4): 620-33
8. Andronache A; Rosazza C; Sattin D; Leonardi M; D’Incerti L; Minati L. Impact of functional MRI data preprocessing pipeline on default-mode network detectability in patients with disorders of consciousness. *Frontiers in Neuroinformatics.* 2013, vol. 7.
9. Salinas Serrano, J. M. Cloud CEIB I+D: Sistema de gestión y extracción de conocimiento de la imagen médica. (Tesis). Universidad de Alicante. Departamento de Ciencia de la Computación e Inteligencia Artificial. 2013
10. Gantes Cabrera, R. Técnicas de reconstrucción de imagen en resonancia magnética. (Tesis) Universidad Complutense de Madrid. Máster en Física Biomédica. 2009.
11. Yan C-Y; Craddock R. C; He Y; Milham M. P. Addressing head motion dependencies for small-world topologies in functional connectomics. *Frontiers in Human Neuroscience.* 2013 vol. 7.
12. Jenkinson M; Bannister P; Brady M; Smith S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage.* 2002; 17(2):825-41.
13. Welvaert M, Rosseel Y. On the definition of signal-to-noise ratio and contrast-to-noise ratio for fMRI data. *PLoS Journals One.* 2013
14. Van Dijk K.R; Sabuncu M.R; Buckner R. L. The influence of head motion on intrinsic functional connectivity MRI. *Neuroimage.* 2012; 59(1):431-8.
15. Jenkinson M; Bannister P; Brady M; Smith S. Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images. *NeuroImage.* octubre de 2012;17(2):825-41.
16. DICOM Standards Committee, DICOM PS3.15 2015c - Security and System Management Profiles. NEMA. 2015. <<http://dicom.nema.org/medical/dicom/current/output/html/part15.html>> [Consulta: 18 de noviembre de 2014].
17. Yan C-Y. Data Processing Assistant for Resting-State fMRI (DPARSF) V2.3. 2. 2013. <<http://www.rfmri.org/DPA>> [Consulta: 20 de diciembre 2014].
18. Hadley M; Sandoz P. JAX-RS: Java™ API for RESTful Web Services. *Java Specification Request (JSR).* 2009. <<http://java.net/nonav/projects/jsr311/sources/svn/content/trunk/www/drafts/spec20080827>> [Consulta: 4 de febrero 2015]

18 Acrónimos

- **ANOVA:** ANalysis Of VAriance
- **AVIS@:** Asociación Valenciana de Informáticos Sanitarios
- **BET:** Brain Extraction Tool
- **BIMCV:** Banco digital de Imagen Médica de la Comunidad Valenciana
- **BOLD:** Blood Oxygen Level Dependent
- **C-PAC:** Configurable Pipeline for the Analysis of Connectomes
- **CEIB-CS:** Centro de Excelencia e Innovación Tecnológica de Bioimagen de la Conselleria de Sanitat
- **CIE LAB:** Comission Internationale de l'Éclairage, Lightness A and B are the color-opponent dimensions
- **CIPF:** Centro de Investigación Príncipe Felipe
- **CMYK:** Cyan Magenta Yellow black
- **CQC:** Centralized Quality Control
- **CTP:** Clinical Trial Processor
- **DICOM:** Digital Imaging and Communication in Medicine
- **DPABI:** Data Processing & Analysis of Brain Imaging
- **DPARSF:** Data Processing Assistant for Resting-State fMRI
- **DTI:** Difusion Tensor Image
- **DVARS:** D hace referencia a la derivada temporal de los instantes de tiempo y VARS hace referencia a la varianza de la media cuadrática RMS
- **EEG:** Electroencefalograma
- **ESFRI:** European Strategy Forum on Research Infrastructures
- **FD:** Framewise Displacement
- **FFA:** Fast Foruier Transform
- **fMRI:** Functional Magnetic Resonance Imaging
- **FMRIB:** Analysis Group Oxford Centre for Functional MRI of the Brain
- **FSL:** FMRIB Software Library
- **GDCM:** Grassroots DICOM
- **GEVID:** Gestor de Ensayos y programas de investigación con Bioimagen para la I+D
- **GIBI:** Grupo de Investigación Biomédica de Imagen
- **HUPLF:** Hospital Universitario y Politécnico la Fe
- **MEG:** Magnetoencefalograma
- **MIRC:** Medical Imaging Resource Center
- **NEMA:** National Electrical Manufacturers Association
- **OCR:** Optical Character Recognition

- **PACS:** Picture Archiving and Communications System
- **PET:** Positron Emission Tomography
- **QC:** Quality Control
- **REST:** Representational State Transfer
- **RGB:** Red Green Blue
- **RMI:** Resonance Magnetic Imaging
- **RMS:** Root Mean Square
- **RNG:** Research Neuroinformatics Group
- **ROI:** Region Of Interest
- **RSNA:** Radiological Society of North America
- **SNR:** Signal to Noise Ratio
- **SNS:** Sistema Nacional de Salud
- **SPECT:** Single Photon Emission Computed Tomography
- **SPM:** Statistical Parametric Mapping
- **tCNR:** Temporal Contrast to Noise Ratio
- **TIC:** Tecnologías de la Información y la Comunicación
- **tSNR:** Temporal Signal to Noise Ratio
- **UPV:** Universitat Politècnica de València
- **VNA:** Vendor Neutral Archive
- **VPN:** Virtual Private Network
- **XNAT:** eXtensible Neuroimaging Archive Toolkit
- **NIFTI:** Neuroimaging InFormatics Technology Initiative