

Imagen médica y molecular, compartición de información, entomo Grid y búsqueda de conocimiento

Montserrat Robles Viejo^{1,2,6}, José Millet Roig^{1,6}, David Moratal Pérez¹, Luis Martí-Bonmatí^{2,3,6}, Vicente Hernández García^{2,4,6}, Bernardo Celda Muñoz^{2,5,6}

¹Grupo BET de la Universidad Politécnica de Valencia

²ADIRM (Asociación para el Desarrollo y la Investigación en RM)

³Servicio de Radiología. Clínica Quirón. Valencia

⁴GRYCAP (Grupo de Redes y Altas Prestaciones). Universidad Politécnica de Valencia

⁵Departamento Química Física. Universidad de Valencia

⁶IM3 Red temática de investigación cooperativa

En este artículo se muestra una forma de trabajo multidisciplinar y coordinado dentro de la red IM3¹, creando una red de conocimiento mediante la colaboración de diversas especialidades y utilizando las técnicas más modernas a nuestro alcance. El objetivo final es optimizar tanto los procesos de atención al paciente relacionados con la resonancia magnética de imagen (RMI) y espectroscópica (ERM) como la investigación conducente a realizar diagnósticos con mayor conocimiento y, por tanto, más precisos.

La necesidad de colaboración nace de la complejidad actual de la imagen médica y molecular y de las enormes posibilidades que las Tecnologías de Información y Comunicaciones (TIC) actuales nos ofrecen. La coordinación de esta colaboración multidisciplinar se realiza desde ADIRM (Asociación para el Desarrollo y la Investigación en Resonancia Magnética). Debido a los avances en los sistemas de adquisición y procesamiento de imágenes convencionales de RM y moleculares mediante técnicas espectroscópicas se está produciendo un gran crecimiento del volumen de datos de cada paciente, datos que necesitan ser analizados y compartidos por diversos especialistas de la medicina y de la investigación para lo que se necesita una infraestructura de tecnologías de información y comunicaciones cada vez más sofisticada. Simultáneamente, la velocidad y conectividad de las redes va en aumento lo que permite una mayor cooperación entre equipos y recursos distribuidos. La mayor potencia de procesamiento permite realizar análisis más complejos y presentar visualizaciones y resultados de imagen en menor tiempo y de mayor calidad. La tecnología Grid es clave para facilitar esta forma de trabajo, utilizando y aprovechando los recursos distribuidos desde una perspectiva global, en la que un conjunto de usuarios y centros (Comunidad Virtual) comparten de forma segura y eficiente computadores, sistemas de almacenamiento, bases de datos u otros recursos sin perder el control individual de cada centro por sus recursos. El Grid se puede entender como una evolución de la World Wide Web, en la que no sólo se comparten

datos, sino recursos, y además esta compartición se realiza desde una perspectiva más eficiente, segura y transparente.

Las especialidades que se incluyen en esta colaboración provienen del campo de la RM (radiología, imagen de RM, espectroscopia de RM, imágenes moleculares espectroscópicas, imágenes de RM cardiaca, RM funcional, mapas cerebrales paramétricos de perfusión y de difusión) y de las tecnologías de la información y comunicaciones (procesamiento y análisis de imagen, visualización interactiva, gestión de la información, ontologías, integración de la información clínica y de investigación y tecnología Grid).

La forma de trabajo, mostrada esquemáticamente en la figura 1, conlleva los siguientes pasos: tras una justificación clínica adecuada se realiza una exploración de los pacientes mediante RM (imágenes anatómicas, de difusión, de perfusión, funcionales, moleculares espectroscópicas, espectroscopia de vóxel único) para analizar su proceso patológico. Toda la información generada se archiva en soporte digital y se procesa aplicando diversas técnicas de posproceso y análisis. La información generada por estas herramientas también se archiva digitalmente. Esta

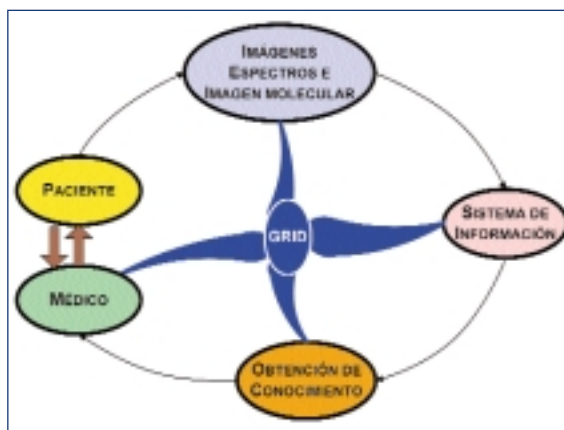


Figura 1. Forma de trabajo coordinado en una red de conocimiento.

Especial: Imagen Médica Molecular y Multimodalidad

información se puede integrar mediante un sistema de información, a continuación un sistema de obtención de conocimiento basado en técnicas de minería de datos accede a través del sistema de información a la información seleccionada de acuerdo con los actores involucrados en cada escenario y aporta y/o confirma resultados, pudiendo funcionar también como Sistema de Orientación Clínica (SOC) para ayuda a la decisión. Estos resultados se utilizan por los profesionales de la salud para un mejor diagnóstico y seguimiento de la salud de los pacientes. Cualquier información relevante para la salud de los pacientes (como es la clínica, genética, ambiental, epidemiológica, además de la propia de imagen) debe formar parte de esta estructura de trabajo, pero no se hace referencia a ella en este artículo por quedar fuera del contexto de este monográfico. Existe otra red temática INBIOMED^{2,3} más orientadas hacia la integración de la información de todos los niveles. Las tecnologías Grid permiten combinar la información de diferentes bases de datos y realizar de forma simultánea su proceso utilizando todos los recursos disponibles.

A lo largo de este trabajo se van a exponer distintas técnicas o modalidades de imagen médica y espectroscopia involucradas, así como las distintas tecnologías de integración, comunicación y análisis de la información puestas en juego en el proceso.

2. ESPECTROSCOPIA DE RESONANCIA MAGNÉTICA E IMAGEN MOLECULAR

La RMN es una técnica espectroscópica que obtiene información de la estructura y composición de la materia a través de las propiedades magnéticas de los núcleos atómicos. Sólo aquellos núcleos con un número protones y neutrones impar/par son activos en presencia de un campo magnético.

Entre los diversos núcleos con propiedades magnéticas destaca por su abundancia y máxima sensibilidad el isótopo ^1H (protón) del átomo de hidrógeno. Así las imágenes de RMN (RMI) se basan fundamentalmente en las diferentes propiedades de relajación (T_1 y T_2) de los ^1H del agua, componente mayoritario de los organismos, aunque también se utilizan en menor medida la densidad protónica, procesos de transferencia de magnetización, propiedades físicas del agua (difusión y perfusión) así como modificaciones de las propiedades magnéticas del entorno (diamagnetismo y paramagnetismo). Sin embargo, la explotación de una de las propiedades diferenciales de la RMN, el desplazamiento químico, apantallamiento del campo magnético real percibido por un núcleo concreto por el entorno electrónico que le rodea en la estructura particular de la molécula a la que pertenece, posibilita el desarrollo de una aplicación en la RMI: la denominada espectroscopia de RMI (ERM). El resultado directo de la ERM es un espectro o conjunto de espectros que representan el conjunto de las diferentes energías absorbidas por el mismo tipo de núcleo, habitualmente ^1H , en diferentes posiciones de una molécula o en distintas moléculas presentes en una determinada localización de un organismo (Figuras 2A, y 2B, esta última realizada con el programa SIVIEW⁴).

La información contenida en un espectro permite identificar los diferentes componentes de una muestra de tejido concreta, y por lo tanto la composición del conjunto de metabolitos observables que lo constituyen. Además, a través de la deconvolución y la integración de las señales que componen un espectro (Figura 2A) se puede determinar la cantidad de cada uno de los metabolitos observados. Asimismo, se pueden generar simultáneamente diversas imágenes moleculares de los distintos metabolitos observados (Figura 2B). En este caso, además de una

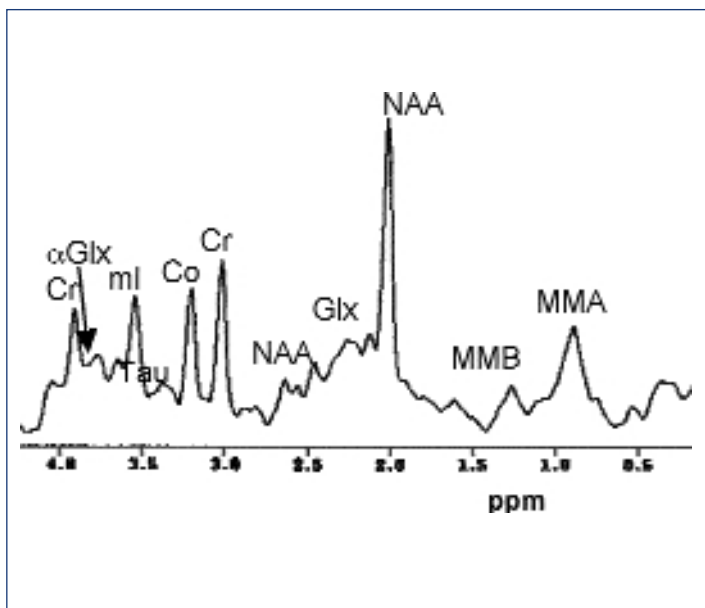


Figura 2A. ^1H espectro de volumen único.

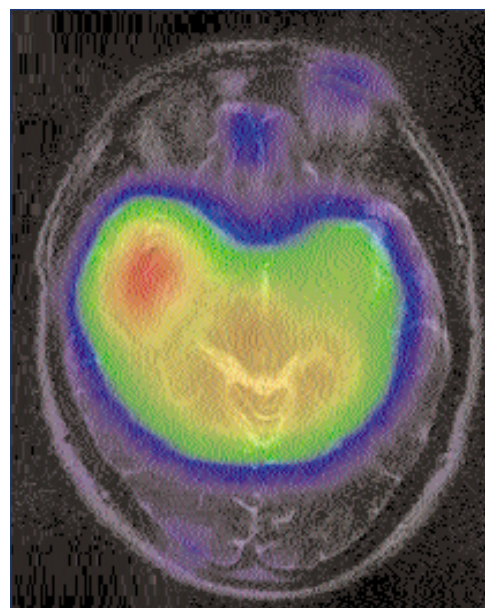


Figura 2B. Imagen Molecular Colina (SIVIEW).

Especial: Imagen Médica Molecular y Multimodalidad

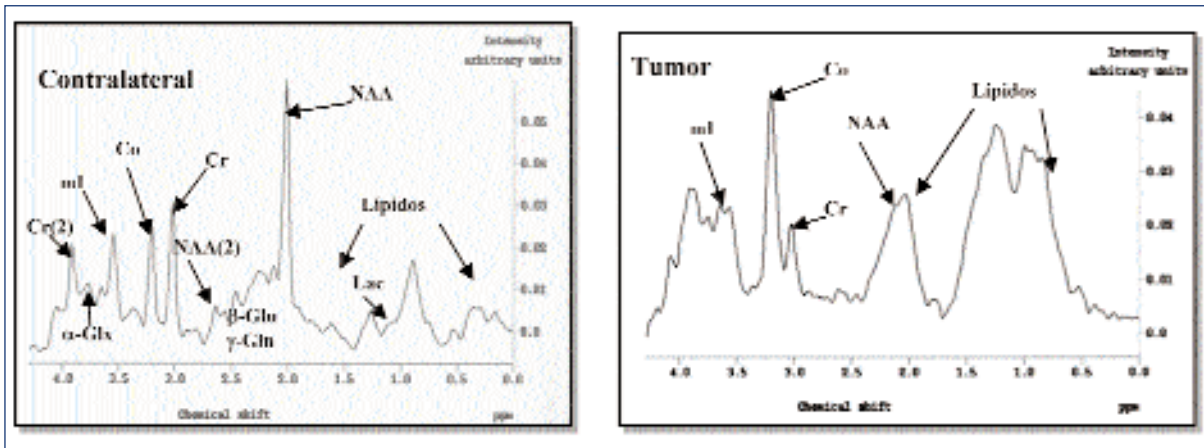


Figura 3. 1H Espectros de Volumen único del cerebro: contralateral y tumor.

adquisición simultánea de imágenes moleculares éstas corresponden a compuestos endógenos, metabolitos propios del órgano/s bajo estudio, por consiguiente, sin la necesidad de inyectar sustancias exógenas con mayor o menor riesgo médico. Aunque el número de metabolitos

observable y cuantificable en espectros de ERM es limitado, por los bajos campos magnéticos usados en clínica (1,5 y 3T) resulta suficiente para: i) diferenciar metabólicamente diferentes órganos y tejidos; ii) distinguir entre tejido no afecto y patológico (Figura 3), siendo específica-

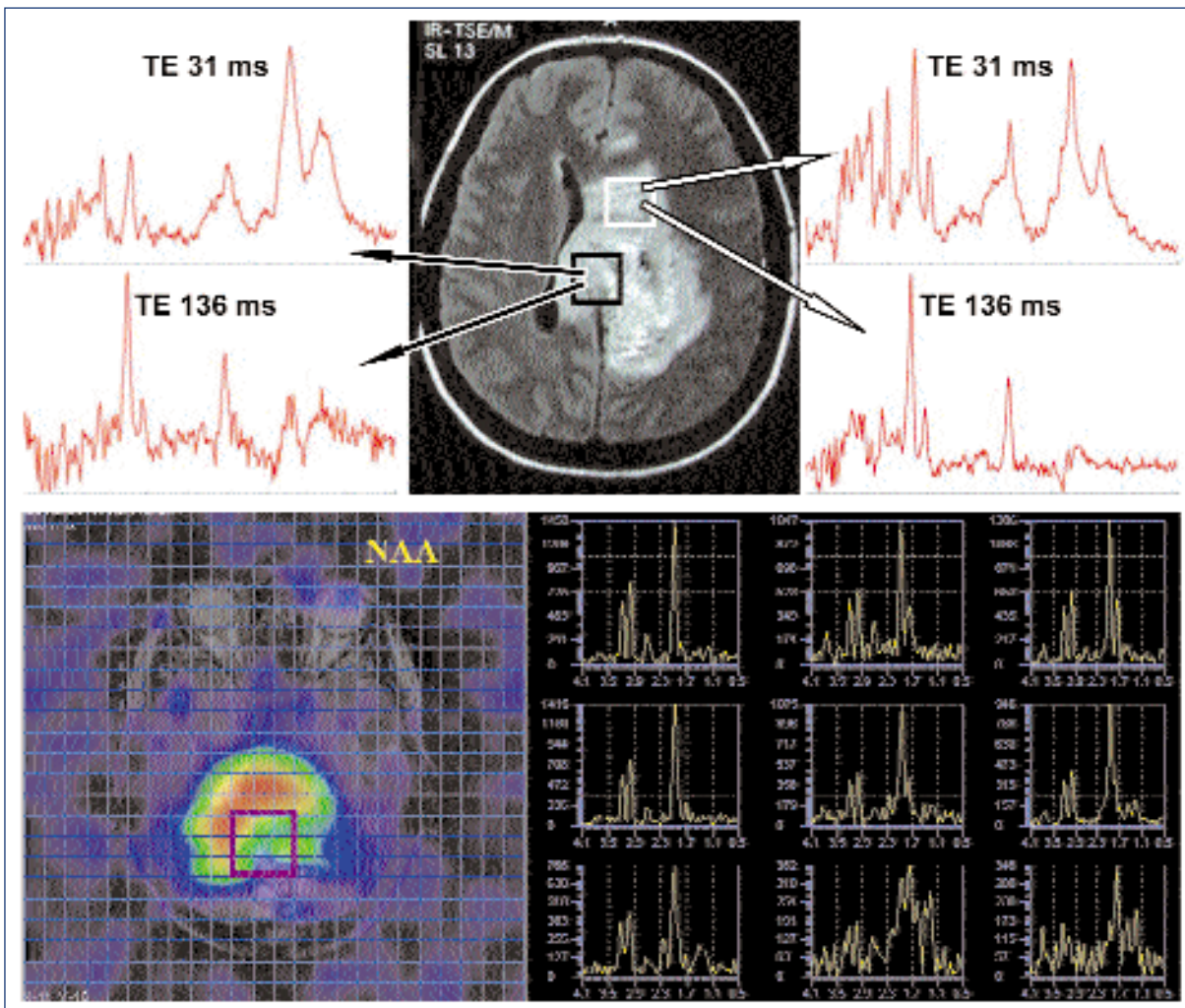


Figura 4 (Arriba). 1H Espectros. Dos Volúmenes únicos en dos localizaciones y dos TE
Figura 5. Imagen Molecular de NAA de una lesión del SNC: Recurrencia. (SIVIEW)

Especial: Imagen Médica Molecular y Multimodalidad

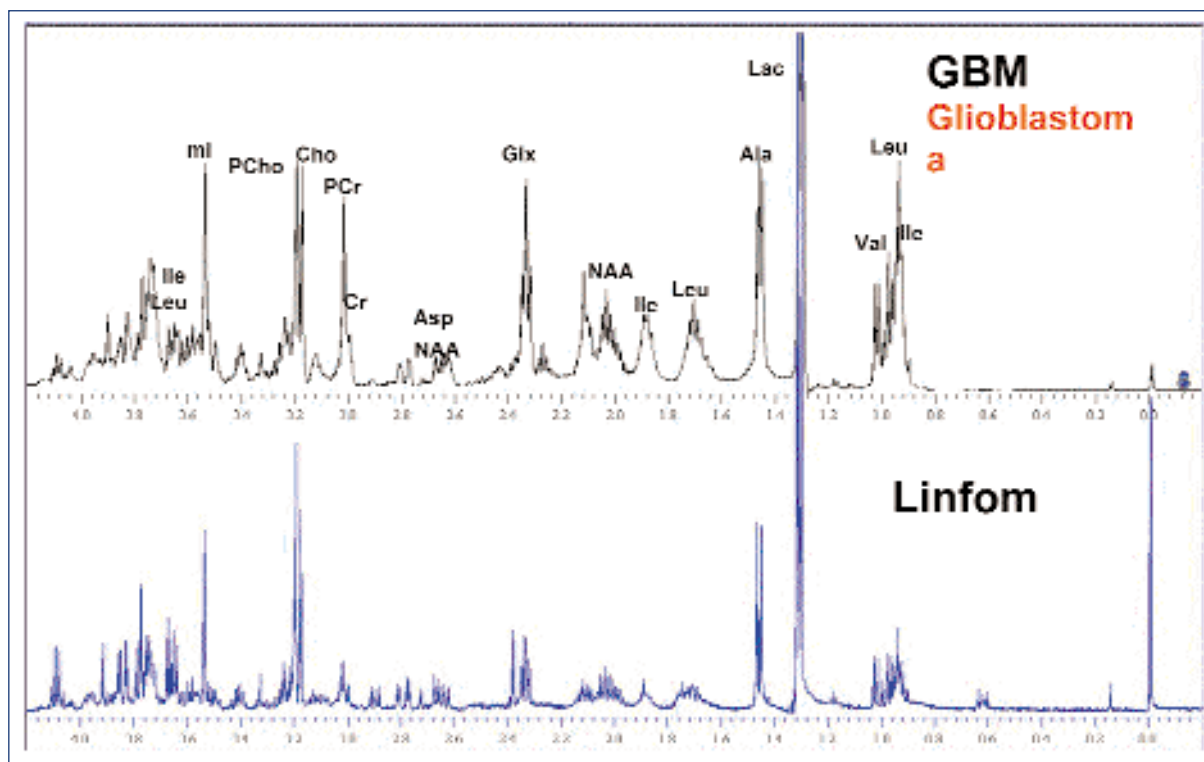


Figura 6. 1H 1D espectros HR-MAS a 500 MHz de dos biopsias de dos tumores del SNC

mente útil en aquellos casos en donde las diferencias anatómico-fisiológicas no son evidentes.

La metodología de adquisición es doble: a) composición metabólica promedio de un volumen concreto seleccionado (VOI) (Figura 4) donde además se muestran espectros adquiridos con diferentes tiempos de eco (TE) con el fin de establecer un filtrado selectivo por la diferente movilidad de los distintos metabolitos, y b) composición metabólica individualizada de un volumen elevado a partir del conjunto de pequeños paralelepípedos que lo componen (espectroscopía de imagen) (Figura 5), generándose distribuciones espaciales, dentro del volumen elegido, de la concentración de diferentes metabolitos, es decir, se obtiene un conjunto de distintas imágenes moleculares de manera simultánea, (p.e. de NAA, de Cr, de Co).

Otra de las posibilidades de la RMN es el estudio de la composición detallada, como mínimo 30 metabolitos diferentes, de tejidos (HR-MAS) (Figura 6) y biofluidos (RMN de alta resolución) de manera directa a altos y ultra altos campos magnéticos (11 a 21 T). De esta manera se puede identificar la composición metabólica detallada (Metabolómica) de diferentes sistemas vivos (desde células hasta tejidos pasando por biofluidos). En este caso además de los espectros en una dimensión, y habitualmente de un sólo núcleo, ^1H (homonucleares), se adquieren y procesan espectros de más de una dimensión y de varios núcleos (heteronucleares) (Figura 7).

Por consiguiente, tanto el almacenamiento y organización de los datos originales de ERM y HR-MAS y RMN de alta resolución como el conjunto de espectros transformados requieren de un sistema informático de organización y localización específicos. Además hay que unir a los requerimientos anteriores la necesidad de compartir el enorme conjunto de datos analizados, cruzados y los diferentes patrones obtenidos. Finalmente, la comparación de los resultados obtenidos a partir de la ERM y HR-MAS con los datos clínicos y de imagen también necesitan de un sistema informático complejo y robusto.

3. MODALIDADES DE IMÁGENES DE RESONANCIA MAGNÉTICA

Aunque la radiología ha avanzado notablemente en los últimos años, son las técnicas topográficas y la imagen digital los pilares de la imagen médica actual. En los servicios de radiología existe un entorno digital en el que la imagen se genera, manipula y analiza con todas sus propiedades. Esta digitalización del trabajo, asociada al desarrollo de secuencias de imagen de muy alta resolución espacial, ha cambiado el trabajo hospitalario. Además, el desarrollo tecnológico y de la RM permite en la actualidad analizar fenómenos hasta ahora impensables, como la estructura macromolecular (transferencia de magnetización), el consumo de oxígeno (estudios funcionales), la permeabilidad de la membrana de los capilares (perfusión farmacocinética) y la distribución metabólica (imagen molecular mediante espectroscopía).

Todas estas técnicas generan imágenes digitales que se

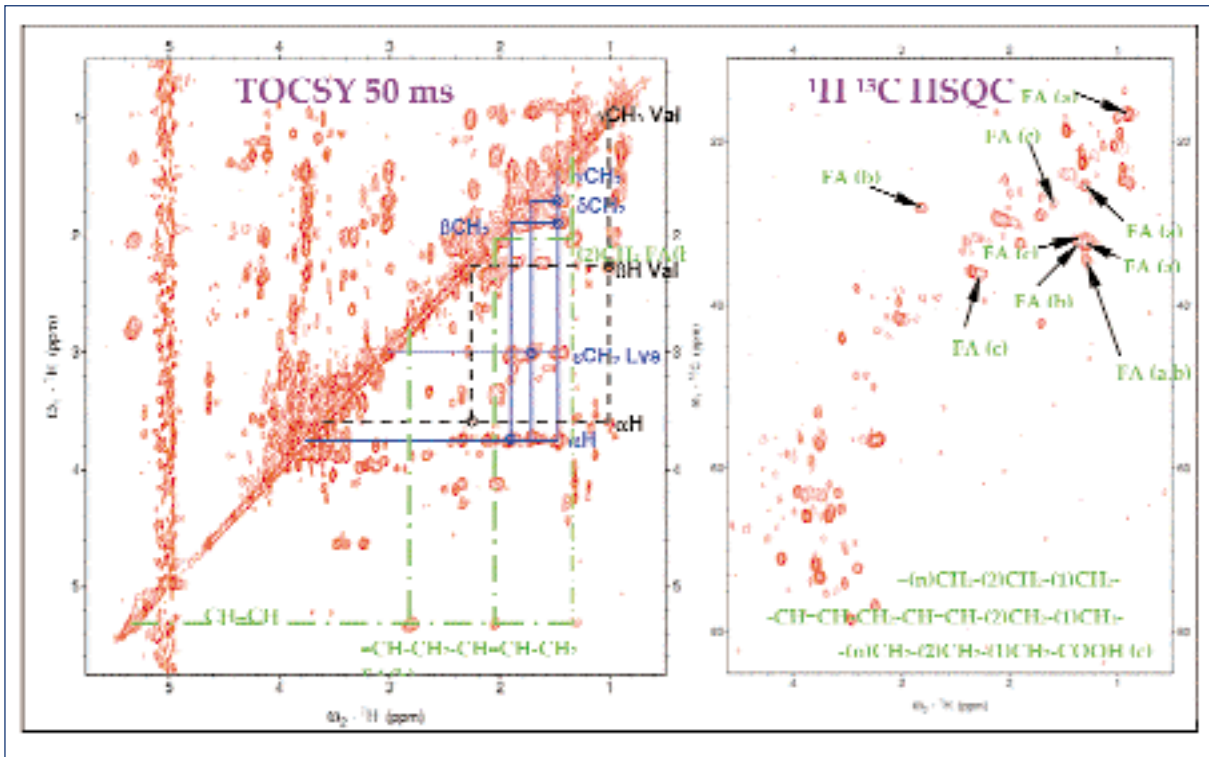


Figura 7. 2D espectros homo(1H) (TOCSY) y heteronuclear (1H-13C) HR-MAS a 500 MHz de una biopsia de tumor del SNC

benefician notablemente del posproceso mediante el que la información se visualiza de forma distinta a como se adquiere (visualización del fenómeno).

3.1 RM funcional

La RM funcional (RMf) genera una enorme cantidad de información. Así, un experimento típico utiliza 8 bloques de dinámicos de adquisición (4 para reposo, 4 para activación), cada bloque constituido por 10 volúmenes cerebrales completos, y cada volumen con 24 imágenes (total de $24 \times 10 \times 8 = 1920$ imágenes) de 128×128 píxeles. Cada imagen ocupa 37 KB, así que para una única sesión de activación se ocupan 71 MB. Si en un experimento estándar se adquieren al menos 2 sesiones (con estímulos distintos, mano derecha - mano izquierda; auditiva con estímulo emotivo - auditiva neutra) más los cortes anatómicos para situar la activación (~ 13 MB), hará un total de 155 MB para este estudio (sin añadir el espacio extra que va ocupando al realizar el posproceso: realineamiento, estadísticos).

Las imágenes de activación funcional son imágenes donde se pueden apreciar las zonas cerebrales que se activan al realizar una acción o percibir un estímulo. El estudio de activación mediante RM no necesita de la administración de un marcador radioactivo, como el PET y SPECT, presenta una resolución espacial y temporal muy adecuada, y genera estudios fácilmente repetibles.

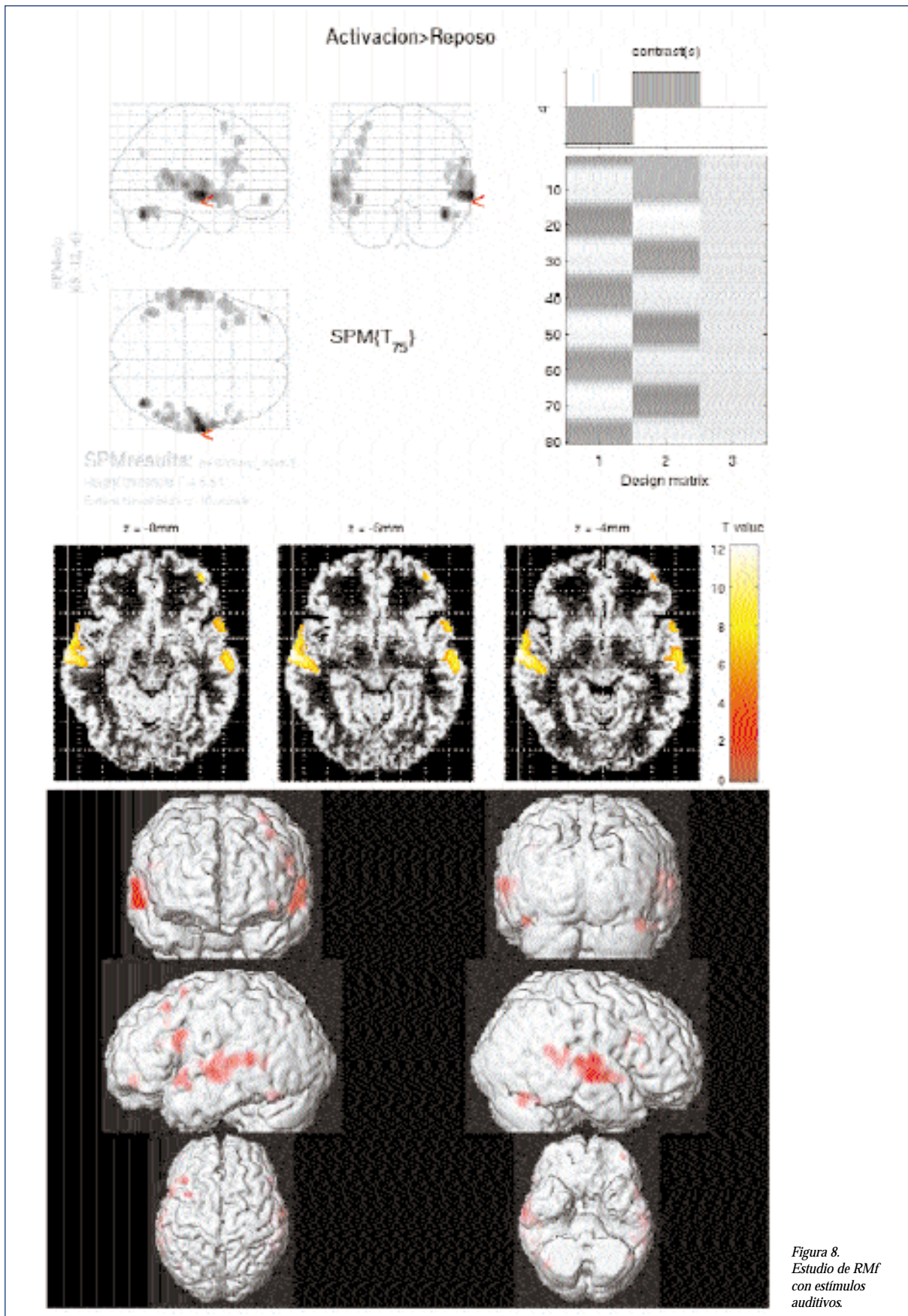
Para adquirir las imágenes de RMf se emplea la técnica

BOLD (blood oxygenation level dependant contrast), que utiliza la hemoglobina como contraste endógeno y una secuencia volumétrica EG-EPI potenciada en T2*. Los mapas de activación se basan en las diferencias temporales en la señal de un píxel al comparar las intensidades registradas durante dos estados distintos del cerebro, denominados estados de "reposo" y de "activación". Durante un estudio completo los dos estados se alternan periódicamente (para dotar de mayor potencia al clasificador) y el volumen de interés se explora varias veces durante cada uno de los periodos de control y de activación⁵. En la figura 8 se muestra, a modo de ejemplo, el resultado de un estudio de RMf con estímulos auditivos de un paciente de esquizofrenia, analizada con software SPM (Statistical Parametric Mapping, FIL, Londres⁶. Del análisis de estas imágenes se obtiene una matriz de resultados de la que se puede extraer la información más importante, consistente en:

- extensión de activación o número de voxels activos en el cerebro
- índice de lateralidad o cantidad de asimetría cerebral para la tarea realizada
- curva hemodinámica, curso que sigue la señal, tanto para el punto de mayor activación cerebral como para regiones de interés

3.2 RM perfusión

La perfusión describe el suministro sanguíneo nutritivo a un órgano o tejido por vóxel. Un estudio de perfusión cere-



Especial: Imagen Médica Molecular y Multimodalidad

bral con RM consiste en la adquisición de una serie temporal de imágenes del cerebro durante el tránsito del medio de contraste. En aquellas áreas del cerebro por donde circula el gadolinio se produce una caída de señal por el efecto paramagnético del contraste. Aplicando una serie de transformaciones matemáticas la información se parametriza, representando cada píxel un parámetro hemodinámico (flujo, volumen, tiempo). La secuencia de perfusión se repite también con una estructura 4D (una adquisición volumétrica 3D repetida dinámicamente con una alta resolución espacial hasta analizar todo el fenómeno) que genera un volumen de información similar (~130 MB). Las imágenes paramétricas analizan el primer paso del contraste a través de los capilares. Los parámetros más importantes son el Volumen Sanguíneo Cerebral Regional (rCBV, *regional cerebral blood volume*), el Tiempo de Tránsito Medio (MTT, *mean transit time*), el Flujo Sanguíneo Cerebral Regional (rCBF, *regional cerebral blood flow*), la Máxima Caída de Intensidad (IMAX), la Concentración Máxima (CMAX), la Llegada del Bolo (BAT, *bolus arrival time*), y el Tiempo al Pico (TTP, *Time To Peak*)^{5, 9, 10}.

3.3 Imágenes potenciadas en difusión y tensoriales

La difusión representa el análisis del movimiento aleatorio o browniano de partículas o moléculas debido a los choques con las moléculas del medio. La difusión se conoce como isotrópica cuando las restricciones son iguales en cualquier dirección. Una difusión anisotrópica con direcciones preferenciales para el agua expresan barreras a la difusión. La Difusión Preferencial tensorial representa al vector neto resultante de las múltiples direcciones espaciales en función de su dirección y magnitud^{5, 7, 8}. Los mapas de tractos (tractografía) producen una información radiológica muy útil mediante el post-proceso de múltiples imágenes (si se estudia todo el cerebro en 24 cortes y cada corte se estudia con 2 valores b y digamos 12 direcciones el número total de imágenes es de $24 \times 12 = 576$).

3.4 Imágenes de resonancia magnética cardiaca

La reducción del tiempo de adquisición de las imágenes sin afectar a su calidad constituye un reto que redundaría en una mayor efectividad y comodidad en la realización del examen clínico. Las técnicas actuales para la adquisición de imágenes dinámicas mediante RM, como por ejemplo la Resonancia Magnética Cardíaca (RMC), requieren la adquisición repetida de imágenes en las que sólo una parte del campo de visión (CDV) cambia a lo largo del tiempo. Las técnicas convencionales de adquisición de imágenes dinámicas adquieren todo el conjunto de datos del espacio-k de cada una de las imágenes que forman este tipo de secuencias animadas. Ello conlleva una redundancia en los datos, por lo que se han propuesto diversos métodos^{11, 12} para gestionar de distintas maneras dicha redundancia.

Nuestro grupo en colaboración con la Emory University ha introducido una novedosa técnica de aceleración llamada Noquist^{13, 14}, que no se sustenta ni en la sustitución ni en la interpolación de los datos. Permite reconstruir sin artefactos de aliasing una secuencia de imágenes dinámicas muestreadas a una menor tasa, mediante la reducción del tamaño del modelo de Fourier discreto del espacio-k de la imagen dinámica. La imagen se reconstruye por inversión directa de este modelo reducido. En la figura 9 se muestra el fundamento de la técnica de aceleración en la adquisición de imágenes dinámicas cardíacas denominada Noquist. En una escena dinámica –en este caso una sección transversal del pecho y corazón– el campo de visión de codificación de fase contiene N líneas independientes, de las cuales ND cambian con el tiempo y NS no lo hacen. La zona no rayada corresponde a la región dinámica de la animación.

En resumen, Noquist se basa en la formulación de la reconstrucción de la imagen como un problema inverso. Un sistema de ecuaciones lineales relacionan los datos del espacio-k con los píxeles de la imagen. El proceso de reconstrucción de la imagen se define mediante la inversión de este sistema, el cual se realiza mediante la Transformada de Fourier Discreta inversa.

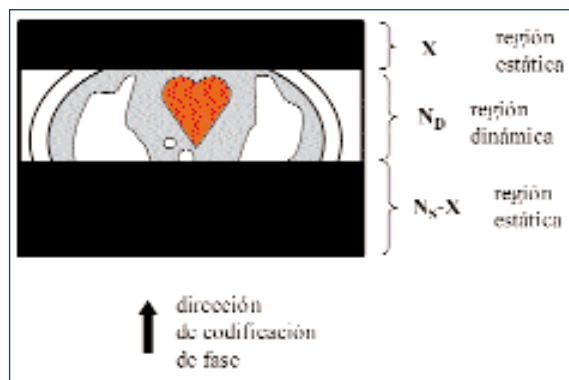


Figura 9. Fundamento de la técnica de aceleración en la adquisición de imágenes dinámicas cardíacas denominada Noquist.

4. INTEGRACIÓN Y COMPARTICIÓN DE LA INFORMACIÓN DISTRIBUIDA

La integración de datos consiste en combinar datos que residen en distintos sistemas, posiblemente heterogéneos entre sí y presentar a los usuarios finales una vista unificada de esta información. Generalmente existen diversas fuentes de datos, con distintos requerimientos y propósitos y que contienen información complementaria y a veces redundante. El objetivo de la integración de datos es ofrecer a todos los usuarios de los mismos un acceso a dichos datos que sea uniforme y transparente al hecho de estar distribuidos y ser heterogéneos. Cuando no se dispone de este acceso uniforme y transparente a los datos cada persona del grupo de trabajo que necesite acceder a los datos almacenados en varios sistemas debe saber qué bases de

Especial: Imagen Médica Molecular y Multimodalidad

datos están disponibles, qué información hay en cada base de datos para así poder determinar su relevancia, saber cómo acceder a cada una de ellas y cómo consultarla^{15, 16}.

Un buen sistema de integración debe dar la impresión a los usuarios de que están trabajando con un único sistema de información local, homogénea y consistente.

La integración de datos ha sido y sigue siendo un campo de investigación muy activo. El uso de la Web como red global de comunicación ha abierto un campo de nuevas posibilidades y un interés creciente en esta disciplina. La tecnología actual aporta gran cantidad de soluciones al problema de integración de los datos por lo que vamos a comentar sólo algunas que puedan ser de ayuda en el problema de integración y compartición segura de la información en un grupo de investigación como el nuestro.

4.1 VPN

Una posible solución sería la creación de una red privada virtual (VPN *Virtual Private Network*) que es una red privada que se extiende, mediante un proceso de encapsulación, y en su caso, de encriptación, de los paquetes de datos a distintos puntos remotos mediante el uso de infraestructuras públicas de transporte¹⁷.

Los paquetes de datos de la red privada viajan por medio de un "túnel" definido en la red pública. En el caso de acceso remoto, la VPN permite al usuario acceder a su red corporativa, asignándole a su ordenador remoto las direcciones y privilegios de la misma, aunque la conexión la haya realizado por medio de un acceso a Internet público.

4.2 Web services

Otra de las tecnologías que nos puede servir en este propósito son los Web services como facilitadores de la interconexión de aplicaciones distribuidas bajamente acopladas. Un *Web Service* (WS) es una aplicación que puede ser descrita, publicada, localizada e invocada a través de una red, generalmente Internet y utilizando una mensajería estándar basada en XML¹⁸.

El protocolo encargado de gestionar esta mensajería se denomina SOAP (*Simple Object Acces Protocol*). Este protocolo define una estructura estándar, reglas de codificado y asociación para transportar documentos XML a través de otra serie de protocolos como HTTP, SMTP, FTP^{18,19}. SOAP permite la comunicación entre aplicaciones heterogéneas de modo que clientes de diferentes plataformas o lenguajes de programación puedan interoperar entre sí.

Los WSs (*Web Services*), además son un mecanismo para construir aplicaciones distribuidas que puedan compartir e intercambiar información entre ellas utilizando como protocolo de comunicación SOAP.

No puede decirse que los WS constituyen la herramienta ideal para el diseño e implementación de cualquier sistema de información distribuido, de hecho, son herramientas similares a otras tecnologías asentadas hace años, tipo CORBA, RMI, DCOM etc., pero en la actualidad parecen ser el punto de referencia en estándares de comunicación de aplicaciones informáticas.

4.3 Mediadores

Otra metodología relacionada con la integración son los Mediadores. El término mediador (mediator) fue introducido por Wiederhold (20) y desde entonces se utiliza en la literatura sobre técnicas y proyectos de integración de datos. Los mediadores son programas informáticos especializados que obtienen la información a partir de una o más fuentes de datos o de otros mediadores, es decir, de los componentes que están por debajo de él, y proporcionan información a los componentes que están por encima (otros mediadores) y a los usuarios externos del sistema. Las fuentes de datos están "envueltas" por una capa de software, denominada adaptador o *wrapper*, el cual traduce entre el lenguaje, modelos y conceptos de la fuente de datos y el lenguaje, modelo y conceptos utilizados en el mediador. Un mediador ofrece una vista unificada e integrada de la información que se encuentra almacenada en las diversas fuentes de datos. El mediador no almacena datos, pero pueden ser consultados como si lo hiciese, su tarea es acceder a sus fuentes de datos y encontrar la respuesta a la consulta.

Tras comentar algunos de los mecanismos que pueden ayudar a la integración de la información distribuida, como son las VPN, los web services y los mediadores hay que decir que existen otras muchas soluciones: bases de datos federadas, utilización de CORBA para integrar con servicios e interfaces bien definidas para el ámbito sanitario como las del Healthcare task force de OMG etc. Pero que cada situación particular requiere de un estudio particular que intente aprovechar al máximo las capacidades y sistemas de base con los que se parte para intentar minimizar al máximo el proceso de integración.

Finalmente comentar que el GRID computing es también una solución interesante para la integración y que es una de las propuestas con más proyección de futuro, sobre todo para necesidades de computación paralela.

Es en este aspecto de integración y compartición de la información y de las herramientas hay que decir que aún existiendo las tecnologías informáticas para hacer posible el acceso a fuentes de información distribuidas desde cualquier lugar y en cualquier momento todavía no se están utilizando al menos con todo el potencial que éstas ofrecen. Observando que el número de estudios y el de proyectos de investigación va en aumento y que para cada

Especial: Imagen Médica Molecular y Multimodalidad

proyecto, a medida que es posible obtener más información, se tiende a adquirirla, es el momento de plantearse el ir caminando hacia la disponibilidad de herramientas que faciliten estas tareas de compartición e integración de la información con el fin de facilitar los procesos de investigación y formación.

5. GRID COMPUTING

Surge por la amplia difusión de Internet y de la Web como nuevo paradigma de la computación distribuida a gran escala. La computación GRID es un concepto de entorno distribuido que permite la compartición, selección y agregación de cada uno de los recursos computacionales distribuidos en una área geográfica extensa (supercomputadores, clusters de computación, sistemas de almacenamiento, fuentes de datos, aplicaciones, instrumentos, gente) y presenta todos estos recursos desde una perspectiva global, unificándolos para solucionar la computación a gran escala y el uso intensivo de información por parte de las aplicaciones como, por ejemplo, el análisis de la actividad cerebral, etc. El concepto de GRID nace a partir de la computación bajo demanda. El objetivo es poder suministrar recursos informáticos físicos y lógicos como se hace con otros bienes de consumo, ya sea la electricidad, el agua, etc.

De la colaboración entre varios centros en el desarrollo de un proyecto tipo 'gran reto' nace la necesidad de interconectar los recursos. Cuando el recurso básico que se comparte lo forman unidades de proceso, se habla de grids computacionales. Si el objetivo es poner a disposición grandes cantidades de datos y sistemas de almacenamiento se habla de grids de datos o Datagrids. Existen otras aproximaciones, como los grids de acceso, orientados a permitir la interconexión simultánea de sesiones de videoconferencia, en los que el recurso que se comparte es más específico.

El Grid se basa en el despliegue de un software intermedio o de base (middleware) que ofrece los servicios específicos para establecer conexiones, compartir recursos, transmitir datos o gestionar trabajos. Desde su comienzo se han desarrollado numerosas implementaciones de middlewares Grid, siendo las más extensamente aceptadas el Globus Toolkit y el Unicore.

Actualmente, las arquitecturas grid están tendiendo a una completa virtualización de los recursos y a una progresiva incorporación de ontologías e información semántica con el objetivo de alcanzar una mayor capacidad de interoperabilidad. La particularización de los entornos Grid en el ámbito Biomédico ha dado lugar a lo que se denominan 'Healthgrids', cuyo principal objetivo es poner a disposición de todos los actores de la salud (desde la epidemiología hasta la ingeniería biomolecular), datos médicos, aplicaciones y recursos con el objetivo de mejorar los

análisis clínicos a nivel de población y fomentar una provisión sanitaria más personalizada. Con este objetivo ha nacido la asociación internacional para la investigación 'Healthgrid', a la que la UPV pertenece y cuyo principal objetivo es fomentar el uso y aprovechamiento de las tecnologías Grid por el entorno de la asistencia sanitaria.

En el acceso a estos datos es fundamental garantizar la seguridad y la confidencialidad, por lo que los healthgrids han adoptado técnicas de mejora de la privacidad (Privacy Enhancement Techniques) que eviten el uso malintencionado de la información biomédica. Además, los healthgrids consideran problemas como el almacenamiento de datos por periodos muy largos o la heterogeneidad de la información, más allá incluso de la adaptación de formatos. Los healthgrids suponen una evolución de las bases de datos distribuidas al permitir la ejecución de servicios de forma ubicua sobre los datos y al incrementar la seguridad sobre la privacidad de la información^{21,22}.

Existen numerosos ejemplos de aplicaciones Grid en salud. Las áreas de aplicación más relevantes son:

- Procesamiento de Imágenes Médicas. Destacan en este ámbito las aplicaciones en cribado Mamográfico, donde se comparten estudios de mamografía y se proporcionan herramientas para el diagnóstico asistido por computador o la búsqueda basada en contenido (MAMMOGRID, eDIAMOND, GP-CALMA).
- Simulación mediante modelos biomédicos. Con aplicaciones muy diversas, existen ejemplos en la simulación de la administración de fármacos por vía respiratoria, la simulación de la cirugía craneofacial o la propagación de potenciales eléctricos (DATAGRID, GEMSS, CAMAEC-GRID).
- Genómica y Proteómica. Con aplicaciones orientadas a la integración de bases de datos de proteínas y genes y a la simulación de los procesos de interacción (BIOGRID).

En el marco de la red IM3, las tecnologías Grid se están utilizando para combinar diferentes fuentes de información (como diferentes bases de datos de imágenes DICOM) y servicios (como clasificadores automáticos) permitiendo un acceso global y transparente a los datos y a su proceso.

6. OBTENCIÓN DE CONOCIMIENTO

La gran cantidad de datos que se pueden acumular en determinados proyectos de investigación en los que además de la información de imagen médica y demás información clínica se extrae información genética hacen que sea muy difícil el análisis y la búsqueda de conocimiento por un ser humano con los métodos tradicionales, haciéndose cada vez más necesaria la ayuda de herramientas

Especial: Imagen Médica Molecular y Multimodalidad

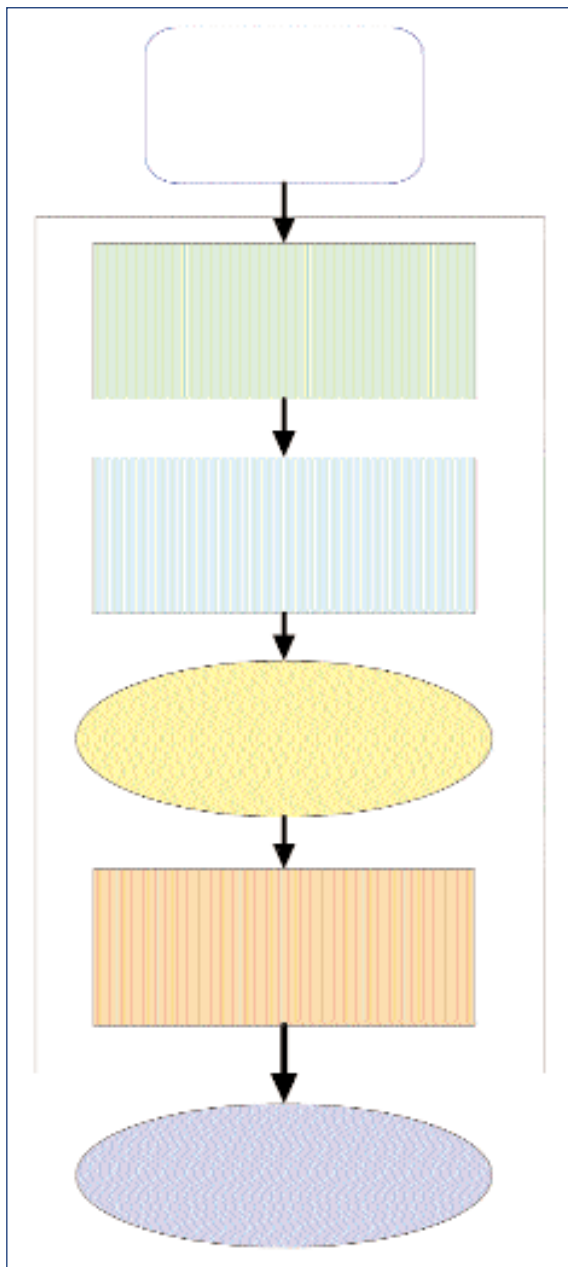


Figura 10: Modelo clásico de DCBD

automáticas que, manejando la componente de información más científica y objetiva, ayude en la toma de decisiones y en el descubrimiento de patrones y/o tendencias contenidos en los datos.

En 1996 Fayyad y col. (23) definieron el Descubrimiento de Conocimiento en Bases de Datos (DCBD) como el proceso **no trivial de identificar patrones válidos, novedosos, potencialmente útiles y en última instancia comprensibles a partir de los datos**. Este proceso reúne un importante número de disciplinas relacionadas con el análisis y gestión de la información como la estadística, los sistemas de información, el aprendizaje automático, la minería de

datos, la visualización de datos, la computación paralela, etc. El fin último es procesar automáticamente grandes cantidades de datos y encontrar información relevante o conocimiento a partir de ellos.

En la figura 10 se muestra el proceso clásico de DCBD a partir de Sistemas de Información (SI). En la primera fase, se requiere una interfaz con el SI, este interfaz debe estar bien planificado, debido al masivo uso que realizarán los procesos de minería de datos sobre el mismo, es por ello que generalmente se realiza de forma desconectada, en la cual una sola vez se realiza la consulta de toda la información histórica sobre la que se desea realizar minería y se almacena en un sistema secundario. La fase principal del proceso es la Minería de Datos que será la que consiga extraer el conocimiento en forma de modelo, reglas o patrones.

Uno de los pasos importantes en el proceso de DCBD es la Minería de Datos, en la cual se incluyen aquellos métodos de aprendizaje, estadística y representación que nos permiten obtener hipótesis de patrones y modelos a partir de los conjuntos de datos médicos utilizados. Las técnicas de minería de datos nos permiten observar las características de grandes volúmenes de datos desde un punto de vista conjunto. La minería de datos aporta una visión global que permite la obtención de tendencias y patrones que a simple vista resultarían difíciles de distinguir debido a la alta dimensionalidad en el número de variables y registros.

Tras evaluar los resultados de métodos de minería de datos sobre un sistema de información, el modelo o el clasificador resultante se puede considerar como una nueva pieza del sistema de información.

En nuestro grupo se está desarrollando una herramienta denominada SOC (Sistema de Orientación Clínica) que permita realizar procesos de DCBD y sirva de sistema de ayuda a la decisión médica. En general, parte de bases de datos con casos ya diagnosticados que sirvan para el aprendizaje de los algoritmos de clasificación. Estos clasificadores están basados tanto en técnicas de reconocimiento de patrones (redes neuronales, gaussianas, cálculo de distancias,...) como en fórmulas o funciones discriminantes empleadas habitualmente en el ámbito médico. Finalmente, a partir de la incorporación de nuevos casos por parte del personal médico, se podrán mejorar los modelos de clasificación desarrollados. Por otro lado, una vez seleccionada la base de datos de interés, la aplicación ofrece la posibilidad de llevar a cabo un completo estudio estadístico. Dicho estudio se compone de un informe detallado de cada uno de los parámetros de la base de datos y de información visual en forma de gráficos de tartas y gráficos de distribución 3D.

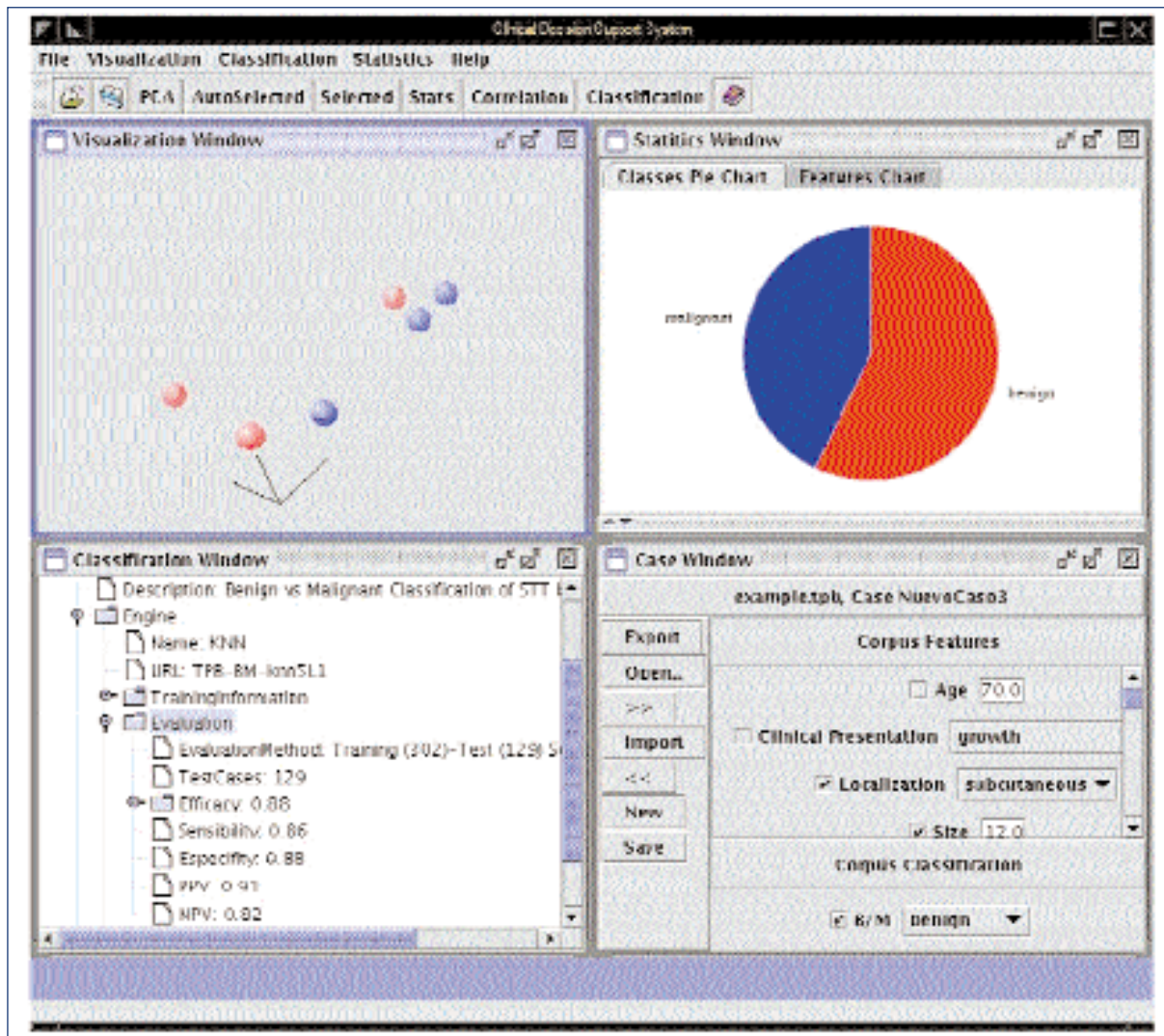


Figura 11. Vista de la interfaz de usuario de SOC

6.1 Arquitectura de SOC

Los usuarios de SOC se pueden conectar al servidor SOC a través de un interfaz multiplataforma. Este interfaz permite acceder a las funcionalidades de acceso a la base de datos clínica, a los motores de clasificación, a los informes estadísticos y a un módulo de representación 3D de la información.

La versión instalable de la aplicación reside en un servidor centralizado basado en Java Web Star, lo que permite un mantenimiento centralizado de todas las copias distribuidas por los hospitales o centros de investigación. El interfaz SOC sirve de punto de conexión único al sistema de minería de datos, disponiendo de las interfaces de acceso a datos, informes estadísticos, visualización de datos y acceso a motores de Minería de datos (típicamente clasificadores). El acceso a datos se realizará de forma distribuida utilizando los estándares de base de datos, estando preparado el diseño para incorporar el acceso a

Sistemas de Información. Como se muestra en la figura 11, mediante la interfaz se pretende ofrecer la información disponible de manera global y ordenada. En el cuadrante inferior derecho de la pantalla se muestran los casos de la base de datos seleccionada, en el inferior izquierdo se tiene acceso a un listado de los clasificadores disponibles para dicha base de datos. Finalmente, en la mitad de arriba de la pantalla se ofrece una ventana con gráficas estadísticas y una ventana de visualización que permitirá observar la clasificación de varias muestras en función de 1, 2 ó 3 parámetros.

El módulo estadístico de SOC ofrece un detallado informe estadístico del fichero de muestras con el que se está trabajando, incluyendo la funcionalidad de impresión del estudio completo. También incluye una herramienta para trabajar con curvas ROC ayudando al análisis de los resultados así como a la comparación de ficheros de muestras.

Especial: Imagen Médica Molecular y Multimodalidad

El Módulo de Clasificación tiene como objetivo ofrecer ayuda al diagnóstico médico. Cada clasificador tiene asociada una información científica que permite la reproducibilidad del experimento (número de muestras empleadas para el entrenamiento y test, método de validación escogido, valores de los parámetros de aprendizaje, ...), así como también una evaluación y auditoria del propio clasificador (tasa de acierto, sensibilidad y especificidad, etc.).

La comunicación entre los objetos de clasificación y la aplicación SOC se realiza mediante XML. Esto permite una comunicación eficaz tanto de las consultas como de los resultados y además facilita la actualización de los clasificadores, pues no residen físicamente en el mismo lugar que SOC.

AGRADECIMIENTOS

Este trabajo ha sido financiado por la Red Temática de Imagen médica, molecular y multimodalidad IM3 financiada por el FIS.

Referencias

- (1) Red temática IM3. <http://im3.rediris.es>
- (2) Red temática INBIOMED. <http://www.INBIOMED.retics.net>
- (3) Especial: INBIOMED, red de investigación en Informática Biomédica (2004) I+S Informática y Salud, 46:7-47.
- (4) Manjón Herrera JV, Martínez-Bisbal MC, Celda Muñoz B, Martí-Bonmatí L, Robles Viejo M. (2003). SIVIEW 2.0. Software para la generación y análisis de imágenes moleculares de espectroscopía de RMN. Libro de Actas Inforsalud 2003:101-105. ISBN: 84-930487-6-3.
- (5) Manjón JV, Martí-Bonmatí L, Robles M, Celda B. (2004). Postproceso en imagen médica: morfología, funcional y molecular. I+S Informática y Salud, 45:62-70
- (6) Statistical Parametric Mapping (SPM). <http://www.fil.ion.ucl.ac.uk/spm/>
- (7) Meier P, Zierler KL (1954). On the theory of the indicator-dilution method for measurement of blood flow and volume. *Appl. Physiol.* 6:731-744
- (8) Ostergaard L, Weisskoff RM, Chesler DA, Gyldensted C, Rosen BR (1996). High resolution measurement of cerebral blood flow using intravascular tracer bolus passages. Part I: Mathematical approach and statistical analysis. *Magn. Reson. Med.* 36:715-725
- (9) Bammer R. (2003). Basic principles of diffusion-weighted imaging. *Eur J Radiol.* 45(3):169-84.
- (10) Bammer R, Acar B, Moseley ME. (2003). In vivo MR tractography using diffusion imaging. *Eur J Radiol.* 2003 45(3):223-34.
- (11) Hu X, Parish T (1994). Reduction of field of view for dynamic imaging. *Magn Reson Med* 31:691-694
- (12) Madore B, Fredrickson JO, Alley MT, Pelc NJ (2000). A reduced field-of-view method to increase temporal resolution or reduce scan time in cine MRI. *Magn Reson Med* 43:549-558.
- (13) Brummer ME, Moratal-Pérez D, Millet-Roig J, Pettigrew RI, Dixon WT. Quantitative Flow Imaging With "NoQuist" Reduced Field of View Acquisition. 10th Scientific Meeting of ISMRM, Toronto, 2003.
- (14) Brummer ME, Moratal-Pérez D, Hong CY, Pettigrew RI, Millet-Roig J, Dixon WT. Noquist: Reduced Field-of-View Imaging by Direct Fourier Inversion. *Magn Reson Med.* 51(2). 331-42
- (15) Maldonado JA, Robles M, Crespo P. (2003). Integration of distributed healthcare records: publishing legacy data as XML documents compliant with CEN/TC251. Proceedings of the 16th IEEE Symposium on Computer Based Medical Systems. pp:213- 218. IEEE Computer Society Press. ISBN 0-7695-1901- 6; ISSN 1063-7125
- (16) Crespo P, Maldonado JA, Robles M, Chavarría M. (2003). Tecnologías de la información al servicio de la historia clínica electrónica. En: Carnicero J. (Coordinador). De la historia clínica a la historia de salud electrónica. Informes SEIS (5). Pamplona: Sociedad Española de Informática de la Salud. pp: 147-192. ISBN 84-930487-7-1
- (17) VPN accesible en: <http://www.cpd.upv.es/proyectos.html>
- (18) Monday PB. Web Service Patterns: Java edition. The Author Press (2003)
- (19) Newcomer E. Understanding Web Services: XML, WSDL, SOAP and UDDI. Independent Technologies Guide. Addison Wesley 2002.
- (20) Wiederhold, G. (1992). Mediators in the Architecture of Future Information Systems. *Computer* 25(3), pp. 38-49.
- (21) Grid Computing Info Centre. Disponible en: <http://www.gridcomputing.com/>
- (22) Open Grid Services Architecture Working Group. Disponible en: <https://forge.gridforum.org/projects/ogsa-wg>
- (23) Fayyad, U.; Piatetsky-Shapiro, G., and Smith, P., From data mining to knowledge discovery: An overview, in Fayyad, U.; Piatetsky-Shapiro, G., Smith, P. and Uthurusamy, R. (eds.) *Advances in Knowledge Discovery and Data Mining*, AAAI/MIT Press, Cambridge Mass. 1996.