

ESTUDIO DE TUMORES DE ESTIRPE VASCULAR MEDIANTE TÉCNICAS DE RECONOCIMIENTO DE PATRONES

César Vidal¹, Juan Miguel García-Gómez¹, Luis Martí-Bonmatí^{2,4},
Montserrat Robles^{1,4}, Joaquín Galant³, José Millet¹

¹Bioingeniería, Electrónica y Telemedicina(BET).Universidad Politécnica de Valencia

²Clínica Quirón de Valencia

³Hospital Universitario San Juan de Alicante

⁴Asociación para el Desarrollo y la Investigación de la Resonancia Magnética(ADIRM)

INTRODUCCIÓN

Este trabajo forma parte de la línea investigadora que tiene como objetivo el diseño, implementación y validación de un sistema de ayuda al diagnóstico radiológico de los tumores de partes blandas a partir de Imagen de Resonancia Magnética (IRM) [1][2]. En esta fase del desarrollo se ha afrontado la clasificación de los tumores agrupados en estirpes o agrupaciones histológicas. En concreto, se pretende estudiar la estirpe *vascular*, discriminándola frente al resto.

La mayoría de estos tumores son benignos y se conocen como «hemangiomas». Su apariencia en RM es bastante característica: lesiones de bordes bien definidos, de márgenes polilobulados, apariencia serpigínosa, con tractos internos, muy intensas en imágenes potenciadas en T1 e hipo-isointensas en las potenciadas en T2. No tienen edema periférico, necrosis ni hemorragia interna. Pueden, cuando son intramusculares, tener grasa interna y asociarse a atrofia muscular. También es típico que tengan calcificaciones esféricas en su interior (las llamamos flebolitos).

La importancia de clasificar correctamente estos tumores se debe a que son lesiones que sólo deben tratarse si producen alteraciones estéticas o del desarrollo de las estructuras mesenquimales próximas.

METODOLOGÍA

A continuación se describe el proceso seguido, describiendo la bases de datos, su origen y parámetros, así como las diferentes técnicas de reconocimiento de patrones utilizadas.

1 Base de datos

La base de datos utilizada consta de 346 casos de diversos hospitales: Hospital Universitario San Juan (Alicante), Hospital Universitario Dr. Pesset (Valencia) y Clínica Quirón (Valencia). Los parámetros fueron extraídos independientemente por diferentes radiólogos a partir de imágenes de Resonancia Magnética potenciadas en SE-T1 y SE-T2/STIR y datos epidemiológicos, obteniendo 22 parámetros por registro.

2 Técnicas

Se utilizaron diferentes técnicas de reconocimiento de patrones, aplicadas a la discriminación entre la estirpe *vascular* y las demás, consideradas éstas como una única clase. Igualmente se realizaron experimentos para discriminar entre la estirpe bajo estudio y un conjunto reducido de estirpes, en concreto aquellas de las que se disponía de menor número de casos (*quiste, grasa, nerviosa, sinovial y fibrosa*), así como frente a esas estirpes una a una. En todos los casos se dividió la base de datos en dos conjuntos diferentes, uno de entrenamiento, con el 70% de los casos, y uno de test, con el restante 30%. El primero de los conjuntos se utilizó para la estimación/aprendizaje del modelo, mientras que el segundo se utilizó para la validación de esos modelos aprendidos por las diferentes técnicas.

Así mismo, se aplicó la técnica de Análisis de Componentes Principales (PCA) para extraer nuevos parámetros relevantes útiles para comparar los resultados y representar las diferentes estirpes en un espacio tridimensional, con el objeto de apreciar visualmente las agrupaciones de casos y el solapamiento entre las estirpes.

Las técnicas de reconocimiento de patrones utilizadas son descritas brevemente a continuación, así como la técnica de extracción de características PCA:

Clasificadores lineales [3]: la clasificación se realiza en función de un hiperplano óptimo de separación entre dos clases. Se ha utilizado SVM (Support Vector Machines), en concreto su implementación *SVMlight*. El uso de kernels puede mejorar la clasificación en los casos en los que no se puedan separar linealmente.

Clasificadores basados en distancias [4]: métodos sencillos e intuitivos ampliamente utilizados en clasificación, donde se definen distancias entre muestras en el espacio d-dimensional de variables. En concreto se ha utilizado la técnica de clasificación por los k vecinos más cercanos (*k-NN*).

Clasificadores paramétricos [5]: se basan en la estimación de parámetros de las distribuciones de probabilidad que representan las clases de estudio, mediante distribuciones multinomiales, mixtura de multinomiales o una combinación de modelos (Bernouilli, gaussiana y multinomial).

Árboles de decisión [4]: realizan una clasificación mediante la aproximación intuitiva de realizar preguntas sucesivas sobre los valores de las variables, generándose un árbol según las posibles contestaciones, en cuyas hojas se encuentran las etiquetas correspondientes a la clasificación. La técnica utilizada es *CART* (Classification And Regression Trees).

PCA [6][7]: El objetivo de PCA es generar unas nuevas variables de tal forma que estas últimas puedan expresar de una forma más estructurada la variabilidad de la primera matriz. En concreto la matriz de variables resultantes está ahora ordenada, de forma que las nuevas variables son ortogonales y la primera es la que contiene más variabilidad, la segunda variable es la segunda con más variabilidad y así sucesivamente.

RESULTADOS

A continuación se muestran los resultados de aplicar las diferentes técnicas a la discriminación entre los casos de la estirpe *vascular* y el resto considerados como una sola estirpe, entre *vascular* y otras estirpes simultáneamente y entre *vascular* y otras cinco estirpes una a una.

Así mismo se presentan unas gráficas con la representación tridimensional de diferentes estirpes frente a *vascular*, con los datos obtenidos tras aplicar la técnica PCA a las seis estirpes representadas.

1 *Vascular* frente a todas las estirpes

En la tabla 1 se muestran los resultados de clasificar el conjunto de test con tres técnicas diferentes, agrupando las estirpes diferentes de *vascular* como una clase única:

Técnica	Exa	Sen	Esp	VPP	VPN
k-NN	0.93	0.75	1	0.85	0.94
CART	0.97	0.96	0.96	0.88	0.99
SVM	0.95	0.94	0.97	0.84	0.98

Tabla 1. Resultados sobre test de la discriminación entre la estirpe vascular y las demás agrupadas. (Exa: exactitud, Sen: sensibilidad, Esp: especificidad, VPP: valor predictivo positivo, VPN: valor predictivo negativo).

2 Vascular frente a varias estirpes

En la tabla 2 se describen los resultados, expresados mediante la exactitud, obtenidos en la discriminación simultánea entre varias estirpes, partiendo de *vascular*, *quiste* y *grasa*, y añadiendo progresivamente estirpes con vistas a comprobar el comportamiento de los clasificadores al aumentar la complejidad del problema:

Técnica	V-Q-G	V-Q-G-N	V-Q-G-N-S	V-Q-G-N-S-F
MNom	0.98	0.92	0.87	0.83
SVM	0.97	0.94	-	-
CART	0.94	0.87	0.82	0.76
KNN	0.95	0.89	0.84	0.78

Tabla 2. Resultados (exactitud) sobre test de la discriminación entre varias estirpes. (MNom: Multinomial; V: Vascular; Q: Quiste; G: Grasa; N: Nerviosa; S: Sinovial; F: Fibrosa).

3 Vascular frente a otras estirpes una a una

En la tabla 3 se observan los resultados evaluados sobre el conjunto test, de la clasificación entre la estirpe bajo estudio y otras cinco estirpes, *quiste*, *grasa*, *nerviosa*, *sinovial* y *fibrosa*, una a una. La técnica que proporcionó mejores resultados fue k-NN.

Estirpe	Exa	Sen	Esp	VPP	VPN
Quiste	0.93	0.96	0.80	0.96	0.80
Grasa	0.84	0.83	0.85	0.83	0.85
Nerviosa	0.95	0.96	0.94	0.96	0.94
Sinovial	0.81	0.87	0.66	0.84	0.73
Fibrosa	0.85	0.87	0.82	0.87	0.82

Tabla 3. Resultados de la discriminación entre la estirpe vascular y otras estirpes una a una. (Exa: exactitud, Sen: sensibilidad, Esp: especificidad, VPP: valor predictivo positivo, VPN: valor predictivo negativo).

4 Representación con PCA

En las gráficas siguientes se muestra la representación tridimensional de los casos correspondientes a la estirpe *vascular*, frente a los casos de las otras cinco estirpes: *quiste*, *grasa*, *nerviosa*, *sinovial* y *fibrosa*. Las coordenadas representadas corresponden a las tres primeras componentes obtenidas tras aplicar PCA a los datos. Se incluye también una gráfica con la representación de los casos de la estirpe vascular frente al resto, con el fin de dar una visión completa del problema.

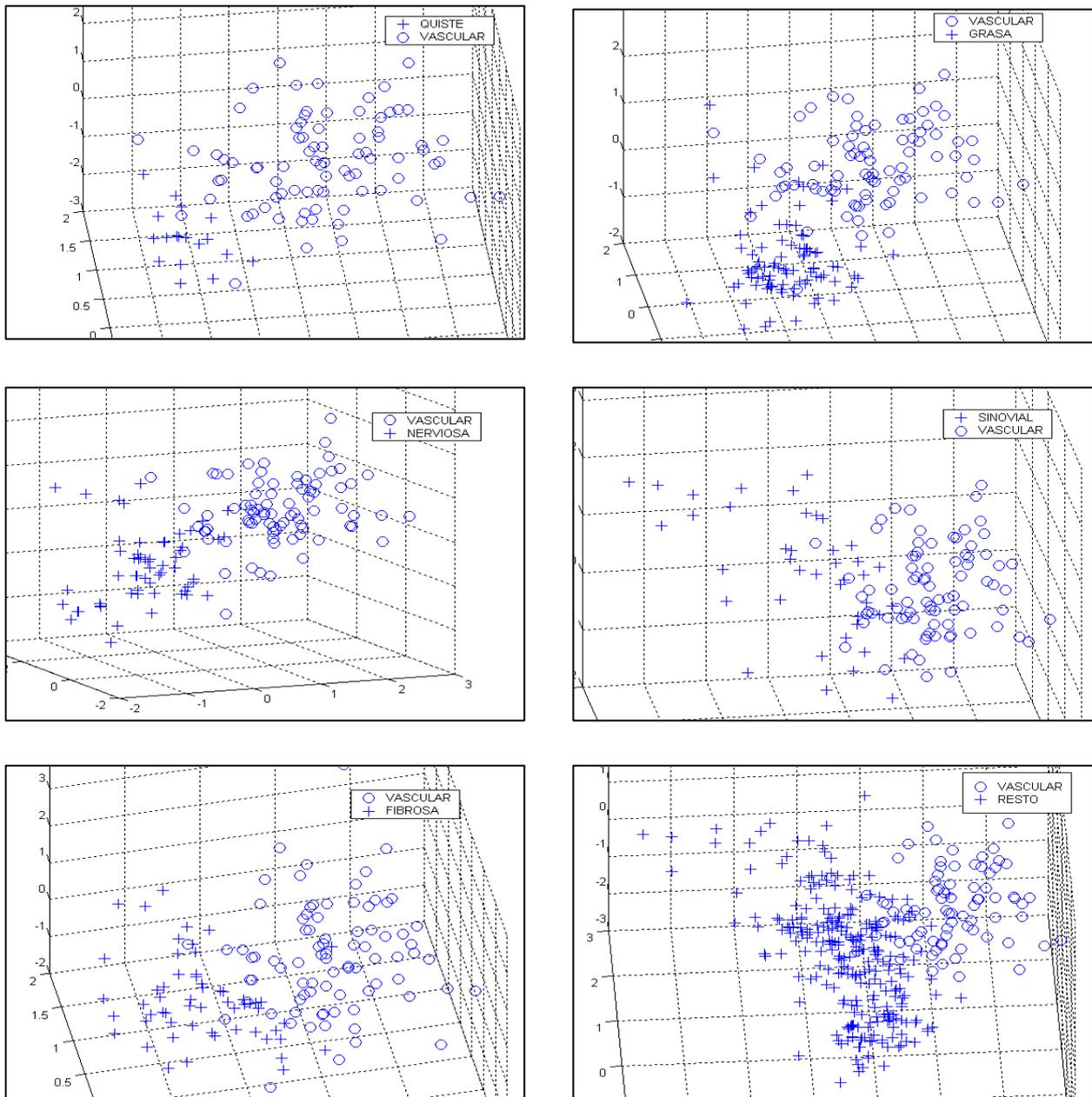


Figura 1. Representación tridimensional de los casos de estirpe vascular frente a las estirpes quiste, grasa, vascular, nerviosa, sinovial y fibrosa tras la aplicación de PCA.

DISCUSIÓN Y CONCLUSIONES

Los resultados obtenidos al diferenciar entre *vascular* y las demás estirpes agrupadas han sido muy buenos, con valores altos de sensibilidad, aún teniendo en cuenta la lógica diferencia entre el número de casos de ambas clases.

La discriminación entre *vascular* y varias estirpes simultáneamente proporciona también exactitud elevada, aunque los resultados empeoran conforme se aumenta el número de las mismas, debido al aumento en complejidad de los modelos. Al discriminar entre la estirpe *vascular* y otras cinco estirpes una a una, se han obtenido también muy buenos resultados, demostrándose pues que la clasificación automática de la estirpe vascular es posible.

A la vista de los resultados obtenidos tras aplicar la técnica PCA, donde se aprecia que las estirpes ocupan zonas bien definidas del espacio tridimensional generado a partir de las nuevas variables representativas, se demuestra que esta técnica proporciona una herramienta gráfica útil e intuitiva para la representación de los datos y el estudio de las agrupaciones de los mismos, así como para la reducción de parámetros sin pérdida apreciable de información.

BIBLIOGRAFÍA

1. J. Galant Herrero. *Aportación de la resonancia Magnética a la caracterización de las lesiones de partes blandas*. Tesis Doctoral. Facultad de Medicina, Universidad Miguel Hernández de Alicante. Noviembre 1998.
2. J.M. García-Gómez, C. Vidal, A. Juan, L. Martí-Bonmatí, J. Galant, M. Robles. *Pattern Classification Techniques applied to Soft Tissue Tumours using MR and epidemiological data*. MAGMA (Official journal of the European Society for Magnetic Resonance in Medicine and Biology, Vol 15, Suplemento 1, p.p. 274-275. 2002.
3. N. Cristianini, J. Shawe-Taylor. *An introduction to support vector machines and other kernel-based learning methods*. Cambridge University Press, Cambridge, 2000
4. R. Duda, P. Hart, D. Stork. *Pattern Classification*. John Wiley & Sons, Inc., New York, 2001.
5. A. Webb. *Statistical Pattern Recognition*. Arnold, London 1999.
6. T. Aluja, A. Morineau. *Aprender de los datos: El análisis de componentes principales-Una aproximación desde el Data Mining*. EUB, 1999.
7. S. Wold, K. Esbensen, P. Geladi, *Principal Component Analysis*. Chemometrics and Intelligent Laboratory Systems. Elsevier Science Publishers B. V., 1987.