

## Aplicación de métodos de aprendizaje-máquina para clasificación de imágenes ecográficas de ovario mediante Transformada de Fourier

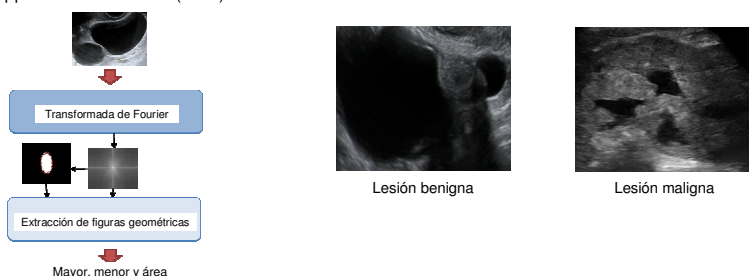
Martínez Más J\*, López Soto A\*, Rubio Ciudad M\*, Bueno Crespo A\*\*, Remezal Solano M\*\*\*, Martínez Cendán JP\*

\* Servicio de Obstetricia y Ginecología, Hospital Universitario Santa Lucía, Cartagena, Murcia.  
\*\* Profesor en Departamento de Informática, Universidad Católica San Antonio de Murcia (UCAM)  
\*\*\* Servicio de Obstetricia y Ginecología, Hospital Clínico Universitario Virgen de la Arrixaca.

**Introducción:** Los tumores de ovario son uno de los grupos diagnósticos más comunes al que se enfrenta un ginecólogo a diario, siendo el examen ecográfico la prueba diagnóstica más utilizada en su estudio, para determinar si se trata de una lesión benigna o maligna. Aún así, se trata de una prueba altamente examinador-dependiente, pudiendo haber una gran variabilidad entre dos diferentes especialistas al examinar el mismo caso.

**Objetivo:** El objetivo de este trabajo es la evaluación de diferentes sistemas bien conocidos de Aprendizaje-máquina para realizar la clasificación automática de tumores ováricos a partir de imágenes ecográficas.

**Material y Métodos:** Utilizamos una base de datos obtenida de 187 imágenes de pacientes reales, cuyo diagnóstico anatomopatológico fue obtenido en menos de 120 días tras la obtención de la imagen ecográfica. Cada una de las imágenes ha sido tratada mediante el método de Transformada de Fourier, focalizándose sobre la región de interés de la imagen, obteniendo para cada una, los Histogramas de Intensidad y Patrones Binarios Locales. Se ha aplicado tres técnicas de Aprendizaje-máquina clásicas para realizar la clasificación: K-Nearest Neighbours (KNN), Linear Discriminant (LD) y Support Vector Machine (SVM).



**KNN:** Este sistema asigna al vector de entrada la etiqueta más cercana a la mayoría de ejemplos de entrenamiento que se sitúan a su alrededor. Utilizamos las métricas Distancia Euclidiana y City block metric, ambas tipos especiales de la métrica Minkowski. Dados dos diferentes vectores de entrada  $X_n$  y  $X_m$ , su distancia utilizando la métrica Minkowski viene dada por:

$$d_{n,m} = p \sqrt[p]{\sum_{i=1}^d |x_{n,i} - x_{m,i}|^p}$$

siendo  $d$  igual a 3 en nuestro trabajo. City block metric se obtiene con  $p=1$ , mientras que con  $p=2$  obtenemos la distancia Euclidiana. La desventaja del método KNN es que el sistema debe retener todos los ejemplos de entrenamiento para poder clasificar nuevas imágenes.

**LD:** Este clasificador obtiene la etiqueta de  $x$  utilizando una combinación lineal ponderada de las imágenes de entrada:

$$y(x) = w_0 + w_1x_1 + w_2x_2 + \dots + w_dx_d$$

donde  $w=[w_1, w_2, \dots, w_d]$  es el vector de ponderación y  $w_0$  el sesgo. La magnitud de la ponderación  $w_i$  muestra la importancia de  $x_i$  y su signo indica si el efecto es positivo o negativo. En un sistema LC,  $x$  es asignado al grupo C1 si  $y(x) > 0$  y al grupo C2 si  $y(x) < 0$ .

**SVM:** Se trata de un método discriminativo de kernel basado en la teoría de aprendizaje estadístico. Un kernel es una función que transforma los datos de entrada en un espacio multidimensional que puede ser lineal o no lineal, escogiendo un kernel lineal para realizar este trabajo. Después de que la entrada sea transformada por la aplicación del kernel, el SVM determina el hiperplano marginal máximo para separar los dos diferentes grupos de clasificación en el espacio multidimensional resultante. Hemos utilizado dos procedimientos bien conocidos de entrenamiento: Least Squares (LS) y Sequential Minimal Optimization (SMO).

**Resultados:** Los experimentos han sido realizados en MATLAB 8.2 (R2013b), en el mismo ordenador. La Tabla 1 muestra los resultados obtenidos, en función de la Precisión (ACC) y Área Bajo la Curva (AUC). Se aprecia que el clasificador KNN ofrece unos resultados muy imprecisos para este problema, independientemente de la distancia elegida (Euclidiana o City block) y el tamaño de la aproximación local, por lo que este clasificador debería ser omitido del proceso de clasificación de nuestro Sistema de Ayuda al Diagnóstico Computerizado. Respecto a los otros dos sistemas, LD y SVM, sus resultados son significativamente mejores en ambos, y al compararlos entre ellos, el SVM aporta mejor precisión y AUC que el LD.

Método de clasificación		ACC (en %)	AUC
KNN	Euclidean distance, K=1	50.27	0.4836
	Euclidean distance, K=10	52.94	0.4522
	Euclidean distance, K=15	56.68	0.4377
	Euclidean distance, K=30	55.08	0.3907
	City block distance, K=1	53.48	0.5127
	City block distance, K=10	57.20	0.5169
	City block distance, K=15	58.29	0.4912
LD	Least Squares method	85.56	0.8514
	SVM	86.10	0.8545
(con Kernel lineal)	LS training	87.70	0.8740

**Conclusiones:** Los métodos de Aprendizaje-Máquina pueden ser eficientemente utilizados para el desarrollo de la fase de clasificación en Sistemas Computacionales de Ayuda al Diagnóstico para imágenes ecográficas de tumores ováricos. En particular, LD y SVM pueden ofrecer una clasificación automática con gran precisión. Por otro lado, y de acuerdo a los resultados obtenidos, los descriptores geométricos obtenidos mediante la función Transformada de Fourier aporta una información relevante y muy útil para la clasificación de tumores ováricos.

**Bibliografía:** [1]Berek JS, Berek & Novak's Gynecology, 14th edition. Lippincott Williams & Wilkins: p471.[2] Asociación Española Contra el Cáncer: Incidencia del cáncer de ovario (<https://www.aecc.es/sobrecancer/cancerporlocalizacion/cancerdeovario/paginas/incidencia.aspx>); [3] Roldán F. Cáncer de ovario. Boletín Oncológico 1998, 8 (1). (<http://www.bolonccl.com/boletin-8/cancer-de-ovario.html>); [4] Jacobs I, Oram D, Fairbanks J, Turner J, Frost C, Grudzinskas JG. A risk of malignancy index incorporating CA 125, ultrasound and menopausal status for the accurate preoperative diagnosis of ovarian cancer. Br J Obstet Gynaecol. 1990 Oct;97(10):922-9; [5] Timmerman D, Valentin L, Bourne TH, Collins WP, Verrelst H, Vergote I; International Ovarian Tumor Analysis (IOTA) Group. Terms, definitions and measurements to describe the sonographic features of adnexal tumors: a consensus opinion from the International Ovarian Tumor Analysis (IOTA) Group. Ultrasound Obstet Gynecol. 2000 Oct;16(5):500-5; [6] Timmerman D, Verrelst H, Bourne TH, De Moor B, Collins WP, Vergote I, Vandewalle J. Artificial neural network models for the preoperative discrimination between malignant and benign adnexal masses. Ultrasound Obstet Gynecol 1999; 13:17-25; [7] Timmerman D, Testa AC, Bourne T, Ameye L, Jurkovic D, Van Holsbeke C, Paladini D, Van Calster B, Vergote I, Van Huffel S, Valentin L. Simple ultrasound-based rules for the diagnosis of ovarian cancer. Ultrasound Obstet Gynecol 2008; 31: 681-690; [8] Timmerman D, Ameye L, Fischerova D, Epstein E, Melis GB, Guerriero S, Van Holsbeke C, Savelli L, Frusco R, Lissoni AA, Testa AC, Veldman J, Vergote I, Van Huffel S, Bourne T, Valentin L. Simple ultrasound rules to distinguish between benign and malignant adnexal masses before surgery: prospective validation by IOTA group. BMJ 2010;341:c6839; [9] Kaijser J, Sayasneh A, Van Hoorde K, Ghaem-Maghamsi S, Bourne T, Timmerman D, Van Calster B. Presurgical diagnosis of adnexal tumours using mathematical models and scoring systems: a systematic review and meta-analysis. Human Reproduction Update. 20 (3) 2014, pp. 449-462; [10] Garcia-Laencina PJ, Henriques Abreu P, Henriques Abreu M, Afonso N. Missing data imputation on the 5-year survival prediction of breast cancer patients with unknown discrete values. Computers in Biology and Medicine. 59, April 2015 pp. 125-133; [11] Garcia-Laencina PJ, Rodriguez-Bermúdez G, Roca-Dorda J. Feed-Forward Neural Network Architectures Based on Extreme Learning Machine for Parkinson's Disease Diagnosis. XIII Mediterranean Conference on Medical and Biological Engineering and Computing, IFMBE Proceedings, 41, 2014, pp. 746-749; [12] Khazendar S, Al-Assam H, Du H, Jassim S, Sayasneh A, Bourne T, Kaijser J, Timmerman D. Automated classification of static ultrasound images of ovarian tumours based on decision level fusion. 6th Computer Science and Electronic Engineering Conference (CEECE). 2014. University of Essex, UK; [13] Khazendar S, Sayasneh A, Al-Assam H, Du H, Kaijser J, Ferrara L, Timmerman D, Jassim S, Bourne T. Automated characterisation of ultrasound images of ovarian tumours: the diagnostic accuracy of a support vector machine and image processing with a local binary pattern operator. Facts Views Vis Obgyn., 7 (1), 2015, pp. 7-15; [14] Gonzales R, Woods R, Digital Image Processing, Addison-Wesley Publishing Company, 1992; [15] Alpaydm E. Introduction to Machine Learning, 2nd Edition, The MIT Press, 2010; [16] Bishop CM. Pattern Recognition and Machine Learning, Springer, 2010; [17] Rodríguez-Bermúdez G, García-Laencina PJ, Roca-Dorda J. Efficient automatic selection and combination of EEG features in least squares classifiers for motor-imagery brain computer interfaces. International Journal of Neural Systems, 23(4), 2013, 1350015 (17 pages).