

Comparación de métodos de clasificación aplicados al diagnóstico de melanomas malignos mediante asimetría

Comparison of classification methods applied to the diagnosis of malignant melanoma using asymmetry

Juan Paulo Sánchez-Hernández,^{1*} Deny Lizbeth Hernández-Rabadán²

¹ Universidad Politécnica del Estado de Morelos.

Boulevard Cuauhnáhuac 566, col. Lomas del Texcal. Jiutepec, Morelos, México. CP 62550

² Instituto Tecnológico y de Estudios Superiores de Monterrey, campus Cuernavaca.

Autopista del Sol km 104, col. Real del Puente. Xochitepec, Morelos, México. CP 62790

* Correo-e: juan.paulosh@upemor.edu.mx

PALABRAS CLAVE:

melanoma maligno, redes neuronales, ID3, máquinas de soporte vectorial, Naive Bayes, momentos geométricos

RESUMEN

Las características físicas externas de un melanoma maligno, tales como su asimetría, color, diámetro y borde, permiten identificarlo y diferenciarlo de una lesión común o melanoma benigno sin necesidad de recurrir a una biopsia. En este trabajo se presentan una metodología y una comparación de resultados obtenidos mediante las redes neuronales artificiales, algoritmo ID3, máquinas de soporte vectorial y clasificadores naive Bayes. Además, se extrajeron los patrones de la imagen segmentada con momentos geométricos, los cuales se utilizaron para entrenar cada uno de los modelos presentados. Para comparar los resultados se empleó la validación cruzada, la cual arroja la precisión de cada uno de los modelos evaluados.

KEYWORDS:

melanoma, neural networks, naive Bayes, SVM, ID3, cross validation

ABSTRACT

The external physical characteristics of a melanoma, such as asymmetry, color, border and diameter, allow its identification and differentiation from a benign melanoma, without biopsy. This research presents a methodology and a comparison of results obtained using artificial neural networks, ID3, support vector machines and naive Bayes classifiers are presented. In addition, the patterns of the segmented image were extracted using geometric moments, which were used to train each of the models presented. To compare the results used cross-validation is applied to obtain the precision of the models tested.

1 INTRODUCCIÓN

El cáncer puede ser ocasionado por distintos factores, tanto internos como externos; dentro de los segundos se encuentran el consumo del tabaco, organismos infecciosos o la radiación, y en los primeros se consideran la herencia familiar, hormonas o mutaciones metabólicas. La combinación de dichos factores puede causar la muerte a los individuos, según otro estudio [1]. Sin embargo, si se realiza un diagnóstico temprano, el cáncer puede ser eliminado a través de una simple cirugía. Respecto a las características físicas externas utilizadas por los médicos para el diagnóstico, existe un consenso en cuanto a cuáles permiten identificar un melanoma: la asimetría (A), bordes (B), color (C) y diámetro (D) mayor a 6 mm. Por supuesto, dicho diagnóstico debe estar soportado por los antecedentes y sintomatología del paciente [2].

A pesar de que las características para el diagnóstico del melanoma son conocidas, surgen confusiones e inconsistencias entre los médicos al momento del diagnóstico.

En este trabajo se presenta una metodología para reconocer y extraer la información que caracteriza a un melanoma tomando en cuenta su asimetría. Mediante las características obtenidas de la asimetría se lleva a cabo el entrenamiento de métodos de clasificación, tales como las redes neuronales [3][4][5], algoritmo ID3 [6], máquinas de soporte vectorial [7] y naive Bayes [8]. Lo anterior se hace con el fin de realizar el diagnóstico preliminar de la clasificación de un melanoma maligno con respecto a uno benigno.

2 TRABAJOS RELACIONADOS

Durante los últimos años, la comunidad científica ha mostrado mucho interés por el desarrollo de sistemas computacionales para el diagnóstico de lesiones de piel, con el principal objetivo de detectar tempranamente aquellas lesiones clasificadas como melanomas malignos. Muchas de las metodologías desarrolladas por los investigadores para diagnosticar un melanoma se apoyan en las características visuales de la piel. El estudio realizado por Lee y Claridge [9] demuestra que la irregularidad del borde es una de las características de suma importancia clínica para el diagnóstico, por tal motivo, dicha característica fue elegida en esta investigación. Según Lee *et al.* [10] hay dos tipos de irregularidades de bordes: las irre-

gularidades de textura, que corresponden a pequeñas variaciones a lo largo del borde, y las irregularidades de estructura, que describen protuberancias y hendiduras globales. Se han desarrollado muchos índices para medir la irregularidad del borde, los más comunes utilizan descriptores de forma, como el índice de compacidad [11]. Desafortunadamente, el índice de compacidad es sensible a ruido, por lo que esta característica no es suficiente por sí sola para lograr discriminar un melanoma de tumores benignos. En esta investigación además del índice de compacidad, se utiliza el índice de alargamiento [12] con el objetivo de reforzar las decisiones en el diagnóstico.

Un trabajo precedente en materia de reconocimiento visual de melanomas malignos a partir de sus características de área y características radiométricas es el descrito en [3]. Algunos otros trabajos, además de las características de forma, toman en cuenta características de color y textura para su diagnóstico [13]. En [14] compararon el desempeño de métodos de clasificación simbólicos mediante características geométricas, color y características basadas en wavelets, con las que obtuvieron mejores resultados.

El estado del arte realizado por [15] reporta que los métodos de clasificación que destacan en este tema, no sólo por ser los más utilizados en las investigaciones sino por su nivel de desempeño en el reconocimiento de melanomas malignos, son aquellos basados en reglas: los estadísticos, las redes neuronales artificiales (RNA) y las máquinas de soporte vectorial (MSV).

Se han realizado algunos estudios comparativos de la eficiencia de algunos métodos de clasificación en el diagnóstico de melanomas malignos; por ejemplo, en [16] comparan las MSV, las RNA y el método de K-vecinos más cercanos y concluyen que las MSV superan el promedio de reconocimiento de los otros dos métodos.

3 METODOLOGÍA

Los esfuerzos se centran en la clasificación de los melanomas malignos, tomando en cuenta las características de asimetría de la lesión. Para alcanzar los objetivos se plantea la metodología (figura 1) dividida en tres etapas importantes: 1) captura, limpieza, extracción de las características de la imagen y discretización; 2) entrenamiento y reconocimiento de los distintos clasificadores, y 3) evaluación de cada uno

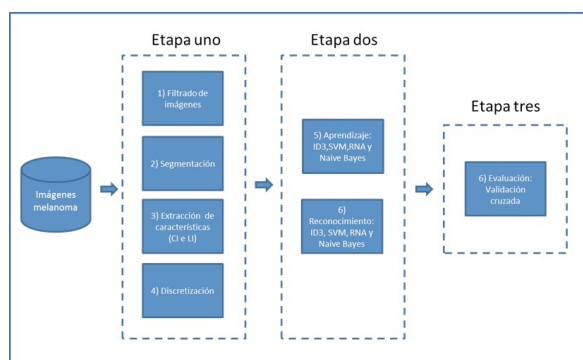


Figura 1. Metodología empleada

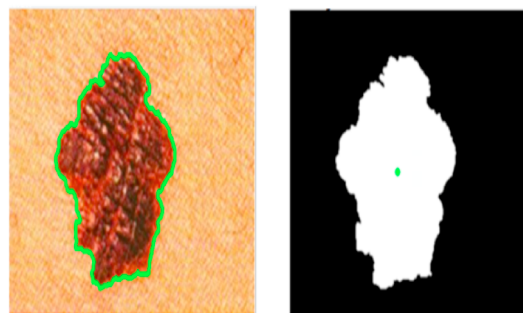


Figura 2. a) Imagen segmentada b) Imagen binarizada

de los clasificadores mediante la validación cruzada.

Como un paso preliminar, se realizó la recolección de las imágenes de melanomas malignos y no malignos de tal manera que se utilizó como base de datos de nuestra metodología. Una vez que se obtuvieron las imágenes de los melanomas, se procedió con el inicio de las tres etapas:

1) En la primera etapa se realizó el preprocesamiento y transformación de las imágenes; es decir, se aplicaron filtros para realzar los bordes de la imagen. Posteriormente se binarizaron, se segmentaron y se extrajeron sus bordes (figura 2). Una vez que se obtuvieron los bordes de la imagen, se calcularon los índices de compacidad (compactness index, CI) y de alargamiento (lengthening index, LI) [12], y se discretizaron las imágenes.

2) Posteriormente se realizaron los dos procesos importantes, se entrenó cada uno de los modelos utilizando las mismas características y se realizó el reconocimiento del melanoma.

3) Finalmente se evaluó el modelo mediante la validación cruzada tomando en cuenta el conjunto de 110 imágenes de melanoma maligno y no maligno.

3.1 Extracción de características y discretización

Para poder extraer las características de la imagen, es necesario segmentar y binarizar la imagen lesionada (figura 2). La característica que se tomó en cuenta en esta investigación fue la asimetría, de acuerdo con el cuadro nemotécnico ABCD. Los clasificadores se entrenaron utilizando dicha característica.

Para poder extraer la asimetría del melanoma, se utilizan los momentos geométricos. El primero de ellos es el momento de orden cero (M_{00}) (1), el cual es el

área (A) de la imagen $I(x, y)$:

$$M_{00} = \sum_x \sum_y I(x, y) \quad (1)$$

Calculando A , y tomando en cuenta que tenemos el perímetro (P) de la imagen podemos calcular el índice de compacidad (CI) [11] de la imagen utilizando (2):

$$CI = P^2 / 4\pi A \quad (2)$$

El otro parámetro que se toma en cuenta para entrenar nuestros modelos, es LI [12], el cual se calcula utilizando los momentos geométricos de orden 2, tal como se muestra en (3):

$$LI = \lambda_1 / \lambda_2,$$

$$\lambda_1 = \frac{m_{20}^c + m_{02}^c - \sqrt{(m_{20}^c - m_{02}^c)^2 + 4(m_{11}^c)^2}}{2},$$

$$\lambda_2 = \frac{m_{20}^c + m_{02}^c + \sqrt{(m_{20}^c - m_{02}^c)^2 + 4(m_{11}^c)^2}}{2}. \quad (3)$$

Cuando se extraen las características y se obtienen las medidas de asimetría, en esta fase se discretizan los valores, de tal manera que sea más fácil manejarlos.

En la discretización se generan cinco binas (range1, range2, range3, range4 y range5) por cada una de las características de IC, al igual que LI (figura 3).

Dicho proceso de discretización de valores es muy importante debido a que nos permite una mejor manejo y manipulación de los mismos. Una vez que se han discretizado las características, se toman dichos patrones como entrada a los clasificadores elegidos en esta investigación, con el objetivo de entrenarlos y evaluar su eficiencia.

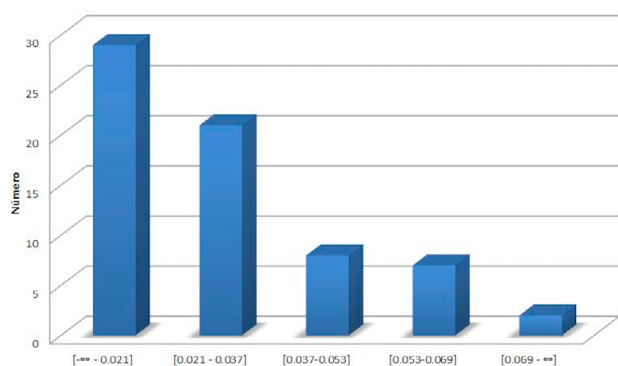


Figura 3. Discretización en cinco binas

3.2 Entrenamiento de clasificadores

Para la fase de aprendizaje y reconocimiento de los resultados, se utiliza la herramienta Rapid Miner, la cual es una de las herramientas para minería de datos muy utilizada en el ambiente científico [17]. Si observamos la metodología nuevamente (figura 1), en la etapa dos, es en donde entrenamos cada uno de los clasificadores elegidos. Como se ha mencionado previamente, la característica que se toma en cuenta para entrenar cada uno de los clasificadores es la asimetría de la lesión, representadas por medio de los dos índices previamente explicados (CI y LI). Con dichas características se construye el árbol de decisión utilizando el algoritmo ID3, en donde los nodos son los índices; los arcos, las características discretizadas y las hojas, la variable clase.

En las redes neuronales se utilizaron los parámetros para entrenar el modelo que se mencionan a continuación. El número de neuronas en la capa de entrada: se tienen 12 neuronas, las cuales son los parámetros discretizados de CI y LI; para el caso de la capa oculta, se configura la red neuronal con nueve neuronas y finalmente en la capa de salida se tienen dos neuronas. En la fase de entrenamiento se estableció que fueron 500 ciclos para entrenar nuestra red.

En lo que se refiere en los métodos de naive Bayes, se utilizan las mismas características de asimetría, en donde el nodo padre es la variable clase (cáncer y no cáncer), y los hijos los parámetros discretizados de los índices de asimetría (range1, range2, range3, range4 y range5). Finalmente, se entrenó con las máquinas de soporte vectorial utilizando un kernel polinomial de grado dos. Al igual que los clasificadores anteriores, se utilizaron los parámetros de asimetría de la señal CI e LI para entrenar el modelo.

4 RESULTADOS

Los resultados obtenidos con cada uno de los clasificadores previamente mencionados se presentan en la tabla 1. Al respecto, vale la pena mencionar que se utilizó la validación cruzada, la cual es uno de los métodos de validación más utilizados [18][19]. De los resultados que se presentan en la tabla 1, es importante mencionar que no son resultados tan útiles debido a que las características de asimetría para distinguir el melanoma maligno no son suficientes. Sin embargo, vale la pena mencionar que de las cuatro técnicas para clasificar que se presentan, naive Bayes e ID3 fueron las más eficaces.

Tabla 1. Resultados

MODELO	EFICIENCIA	ERROR DE CLASIFICACIÓN
Naive Bayes	70.37%	29.63%
ID3	74.07%	25.93%
Redes neuronales	66.67%	33.33%
Máquinas de soporte vectorial	66.67%	33.33%

5 CONCLUSIONES Y TRABAJOS FUTUROS

El objetivo de este trabajo fue realizar una comparación de cuatro clasificadores: naive Bayes, ID3, redes neuronales y máquinas de soporte vectorial utilizando dos características de asimetría (CI y LI). Como se mencionó, los resultados no fueron los esperados, probablemente debido a que las características de asimetría que se utilizan no fueron suficientes para apoyar el proceso de entrenamiento. Lo anterior puede significar que no son linealmente separables.

Como trabajo futuro, se pretende ampliar la base de datos de imágenes de melanoma maligno y no maligno, incluir más características de asimetría, además tomar en cuenta otras características para reforzar la eficiencia de clasificación, tales como el color, borde y diámetro mayor a 3 mm del melanoma. Además, se pretende mejorar el proceso de filtrado de la imagen de tal manera que se permita una eficiente extracción de características que impacte en el entrenamiento y evaluación de los métodos de clasificación.

REFERENCIAS

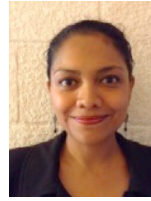
1. American Cancer Society. *Cancer Facts and Figures 2014*. (Consultado junio 2014). Disponible en: <http://www.cancer.org/acs/groups/content/research/documents/webcontent/acspc-042151.pdf>
2. Vega-González, M. T., Toussaint-Caire, S., Gutiérrez-Vidrio, R. M., Orozco-Topete, R. d. I. L., León, E., Martínez-Said, H., López-Graniel, C., et al. Manejo de Lesiones Pigmentadas. *Gaceta Mexicana de Oncología (GAMO)*. 2005, 4 (2), 14-17.
3. Abdul, J., Sibi, S., Aswin, R. B. Artificial neural network based detection of skin. *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*. 2012, 1.
4. Rumelhart, D. E., Hinton, G. E., Williams, R. J. Learning representation by back-propagation errors. *Nature*. 1986, 323, 533-536.
5. Werbos, P. J. Backpropagation through time: what it does and how to do it. *Proceeding of the IEEE*. 1990, 79 (10), 1550-1560.
6. Quinlan, J. R. Induction of decision trees. *Machine Learning*. 1986, 1 (1), 81-106.
7. Vapnik, V. *Estimation of dependences based on empirical data*. Secaucus, NJ, USA: Springer-Verlag New York, 1982.
8. Rish, I. An empirical study of the naive Bayes classifier. *In IJCAI 2001 workshop on empirical methods in artificial intelligence*. 2001, 3 (22), 41-46.
9. Lee, T. K., Claridge, E. Predictive power of irregular border shapes for malignant melanomas. *Skin Research and Technology*. 2005, 11 (1), 1-8.
10. Lee, T. K., McLean, D. I., Atkins, M. S. Irregularity index: a new border irregularity measure for cutaneous melanocytic lesions. *Medical image analysis*. 2003, 7 (1), 47-64.
11. Ercal, F., Chawla, A., Stoecker, W. V., Lee, H. C., Moss, R. H. Neural network diagnosis of malignant melanoma from color images. *Biomedical Engineering, IEEE Transactions*. 1994, 41 (9), 837-845.
12. Amaliah, B., Fatichah C., Widyanto, M. R. Abcd features extraction of image dermatoscopic based on morphology analysis for malanoma skin cancer diagnosis. *Journal Ilmu komputer dan informasi*. 2010, 3 (2), 82-90.
13. Celebi, M. E., Kingravi, H. A., Uddin, B., Iyatomi, H., Aslandogan, Y. A., Stoecker W. V., Moss, R. H. A methodological approach to the classification of dermoscopy images. *Computerized Medical Imaging and Graphics*. 2007, 31 (6), 362-373.
14. Ogorzałek, M., Nowak, L., Surówka, G., Alekseenko, A. Modern techniques for computer-aided melanoma diagnosis. En: Mandi Murph (Ed.). *Melanoma in the clinic - diagnosis, management and complications of malignancy*. ISBN: 978-953-307-571-6, InTech, 2011.
15. Maglogiannis, I., Doukas, C. N. Overview of advanced computer vision systems for skin lesions characterization. *Information Technology in Biomedicine, IEEE Transactions on*. 2009, 13 (5), 721-733.
16. Torre, E. L., Caputo, B., Tommasi, T. Learning methods for melanoma recognition. *International Journal of Imaging Systems and Technology*. 2010, 20 (4), 316-322.
17. Kaur, P., Khurm, S. S., Josan, G. S. Analysis for classification of similar documents among various websites using rapid miner. *Issues and Challenges in Intelligent Computing Techniques (ICICT), 2014 International Conference on*. 2014, 465-470. n° DOI: 10.1109/ICICT.2014.6781327
18. Statnikov, A., Aliferis, C. F., Tsamardinos, I., Hardin D., Levy, S. A comprehensive evaluation of multicategory classification methods for microarray gene expression cancer diagnosis. *Bioinformatics*. 2004, 21 (5), 631-643.
19. Labrarere, J., Bertrand, R., Fine, M. F. How to derive and validate clinical prediction models for use in intensive care medicine. *Springer-Verlag Berlin Heidelberg and ESICM*. 2014, 513-527. n° DOI 10.1007/s00134-014-3227-6

Acerca de los autores



Juan Paulo Sánchez Hernández es Maestro en Ciencias Computacionales por el Instituto Tecnológico y de Estudios Superiores de Monterrey (2008) y Doctor en Ciencias Computacionales por la misma institución (2014). Actualmente es profesor investigador en la Universidad

Politécnica del Estado de Morelos desde 2012. Sus áreas de interés incluyen aprendizaje, reconocimiento de patrones y optimización combinatoria.



Deny Lizbeth Hernández Rabadán es Maestra en Ciencias Computacionales por el Instituto Tecnológico y de Estudios Superiores de Monterrey (2009). Realiza sus estudios de Doctorado en Ciencias Computacionales por la misma institución.

Actualmente colabora en forma parte de proyectos de investigación con el centro de investigación en biotecnología (CEIB) de la UAEM. Sus áreas de interés son procesamiento de imágenes y reconocimiento de patrones.