



ACTAS Derma-Sifiliográficas

Full English text available at
www.actasdermo.org



ORIGINAL

Uso del aprendizaje automático en el diagnóstico del melanoma. Limitaciones por superar



C. González-Cruz^a, M.A. Jofre^a, S. Podlipnik^{a,b}, M. Combalia^a, D. Gareau^d,
M. Gamboa^a, M.G. Vallone^a, Z. Faride Barragán-Estudillo^a, A.L. Tamez-Peña^a,
J. Montoya^a, M. América Jesús-Silva^a, C. Carrera^{a,b,c}, J. Malveyh^{a,b,c} y S. Puig^{a,b,c,*}

^a Servicio de Dermatología, Hospital Clínic de Barcelona, Barcelona, España

^b Institut d'Investigacions Biomediques August Pi I Sunyer (IDIBAPS), Barcelona, España

^c CIBER en Enfermedades raras, Instituto de Salud Carlos III, Barcelona, España

^d Laboratory of Investigative Dermatology, The Rockefeller University, Nueva York, EE. UU.

Recibido el 11 de agosto de 2019; aceptado el 16 de septiembre de 2019

Disponible en Internet el 2 de abril de 2020

PALABRAS CLAVE

Melanoma;
Cáncer de piel;
Dermatoscopia;
Clasificación de
imágenes;
Aprendizaje
automático;
Inteligencia artificial;
Redes neuronales
convolucionales

Resumen

Antecedentes: La clasificación automática de imágenes es una rama prometedora del aprendizaje automático (de sus siglas en inglés Machine Learning [ML]), y es una herramienta útil en el diagnóstico de cáncer de piel. Sin embargo, poco se ha estudiado acerca de las limitaciones de su uso en la práctica clínica diaria.

Objetivo: Determinar las limitaciones que existen en cuanto a la selección de imágenes usadas para el análisis por ML de las neoplasias cutáneas, en particular del melanoma.

Métodos: Se diseñó un estudio de cohorte retrospectivo, donde se incluyeron de forma consecutiva 2.849 imágenes dermatoscópicas de alta calidad de tumores cutáneos para su valoración por un sistema de ML, recogidas entre los años 2010 y 2014. Cada imagen dermatoscópica fue clasificada según las características de elegibilidad para el análisis por ML.

Resultados: De las 2.849 imágenes elegidas a partir de nuestra base de datos, 968 (34%) cumplieron los criterios de inclusión. De los 528 melanomas, 335 (63,4%) fueron excluidos. La ausencia de piel normal circundante (40,5% de todos los melanomas de nuestra base de datos) y la ausencia de pigmentación (14,2%) fueron las causas más frecuentes de exclusión para el análisis por ML.

Discusión: Solo el 36,6% de nuestros melanomas se consideraron aceptables para el análisis por sistemas de ML de última generación. Concluimos que los futuros sistemas de ML deberán ser entrenados a partir de bases de datos más grandes que incluyan imágenes representativas de la práctica clínica habitual. Afortunadamente, muchas de estas limitaciones están siendo superadas gracias a los avances realizados recientemente por la comunidad científica, como se ha demostrado en trabajos recientes.

© 2020 AEDV. Publicado por Elsevier España, S.L.U. Este es un artículo Open Access bajo la licencia CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

* Autor para correspondencia.

Correos electrónicos: susipuig@gmail.com, susipuig@gmail.com (S. Puig).

KEYWORDS

Melanoma;
Skin cancer;
Dermoscopy;
Image classification;
Machine learning;
Artificial Intelligence;
Convolutional neural networks

Machine Learning in Melanoma Diagnosis. Limitations About to be Overcome**Abstract**

Background: Automated image classification is a promising branch of machine learning (ML) useful for skin cancer diagnosis, but little has been determined about its limitations for general usability in current clinical practice.

Objective: To determine limitations in the selection of skin cancer images for ML analysis, particularly in melanoma.

Methods: Retrospective cohort study design, including 2,849 consecutive high-quality dermoscopy images of skin tumors from 2010 to 2014, for evaluation by a ML system. Each dermoscopy image was assorted according to its eligibility for ML analysis.

Results: Of the 2,849 images chosen from our database, 968 (34%) met the inclusion criteria for analysis by the ML system. Only 64.7% of nevi and 36.6% of melanoma met the inclusion criteria. Of the 528 melanomas, 335 (63.4%) were excluded. An absence of normal surrounding skin (40.5% of all melanomas from our database) and absence of pigmentation (14.2%) were the most common reasons for exclusion from ML analysis.

Discussion: Only 36.6% of our melanomas were admissible for analysis by state-of-the-art ML systems. We conclude that future ML systems should be trained on larger datasets which include relevant non-ideal images from lesions evaluated in real clinical practice. Fortunately, many of these limitations are being overcome by the scientific community as recent works show.

© 2020 AEDV. Published by Elsevier España, S.L.U. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introducción

La clasificación automatizada de imágenes por reconocimiento de patrones es una rama del aprendizaje automático (del inglés «Machine Learning» [ML]) que ofrece al dermatólogo una herramienta útil para diagnóstico de cáncer de piel¹. Las redes neuronales convolucionales profundas (del inglés «Deep convolutional neural networks» [DCNN]) han mejorado de manera extraordinaria la precisión en el aprendizaje de patrones y la clasificación de objetos², son utilizadas de manera satisfactoria en la clasificación de imágenes dermatoscópicas de lesiones cutáneas³. Sin embargo, la selección de imágenes mediante ML presenta en la actualidad ciertas restricciones que evitan su uso generalizado. En el presente estudio se evalúan algunos de los criterios de exclusión para la selección de imágenes de neoplasias cutáneas (con especial énfasis en el melanoma) por ML, mencionados en trabajos recientes^{1,4,5}.

Materiales y métodos

Este estudio se realizó en un hospital universitario de tercer nivel especializado en cáncer cutáneo, localizado en Barcelona, España. Se diseñó un estudio de cohorte retrospectivo donde se incluyeron de manera consecutiva 2.849 imágenes dermatoscópicas de alta calidad de tumores cutáneos, obtenidas a partir de la base de datos de la Unidad de Melanoma, recogidas entre el 2010 y 2014. Se utilizó el sistema de microscopía de epiluminiscencia fotográfica digital DermLite® 3 Gen con una conexión de rosca de 37 mm y una cámara Canon modelo G16. Se contó con el diagnóstico histológico en 2.429 de las imágenes. Finalmente, las imágenes se clasificaron según si cumplían o no los criterios de exclusión para el análisis por ML, según los mencionados en la bibliografía^{1,4,5}: dificultad en la detección del borde de

la lesión (ausencia de pigmentación, ausencia de piel normal circundante, presencia de pelo, ubicación en piel volar), metástasis cutánea o lesión ulcerada.

Este estudio fue aprobado por el comité de ética de nuestro centro. Todos los procedimientos con participantes humanos se realizaron de acuerdo con los estándares éticos del comité de investigación institucional y con la declaración de Helsinki de 1964 y sus enmiendas posteriores o estándares éticos comparables.

Resultados

De las 2.849 imágenes obtenidas a partir de nuestra base de datos, 968 (34%) se consideraron elegibles ya que no presentaron ningún criterio de exclusión para su análisis por ML. Nevus, melanomas y carcinomas basocelulares fueron las lesiones más frecuentes de nuestra base de datos. Solo el 64,7% de los nevus y el 36,6% de los melanomas no tenían ningún criterio de exclusión (tabla 1). De los 528 melanomas, 335 (63,4%) fueron excluidos. La ausencia de piel circundante normal (40,5% de todos los melanomas) y la ausencia de pigmentación (14,2%) fueron las causas más comunes de exclusión. Otros motivos de exclusión se muestran en la tabla 1.

Discusión

El melanoma representa la causa más frecuente de muerte por neoplasias cutáneas. El diagnóstico y el tratamiento precoz mejoran significativamente su pronóstico. Se requiere el desarrollo de un método de detección que sea eficaz. La clasificación automática de imágenes a partir del reconocimiento de patrones puede alcanzar una precisión diagnóstica similar a la de un dermatólogo experto⁶. Sin

Tabla 1 A. Imágenes seleccionadas para el análisis por ML. Localización y diagnóstico

	Presencia de algún criterio potencial de exclusión (% del total por localización o diagnóstico)		Ausencia de algún criterio potencial de exclusión (% del total por localización o diagnóstico)		Total
Localización					
Cabeza y cuello	633	(76,8%)	191	(23,2%)	824
Extremidades superiores	159	(62,1%)	97	(37,9%)	256
Extremidades inferiores	297	(60,4%)	195	(39,6%)	492
Piel volar	62	(100%)	0	(0%)	62
Tronco	538	(53,1%)	475	(46,9%)	1013
Mucosas	15	(83,3%)	3	(16,7%)	18
Otro	149	(81%)	35	(19%)	184
Diagnóstico					
Carcinoma basocelular	295	(69,6%)	129	(30,4%)	424
Carcinoma epidermoide	59	(89,4%)	7	(10,6%)	66
Cicatriz	21	(77,8%)	6	(22,2%)	27
Dermatofibroma	17	(77,3%)	5	(22,7%)	22
Lentigo	26	(66,7%)	13	(33,3%)	39
Melanoma	335	(63,4%)	193	(36,6%)	528
Metástasis cutánea	9	(100%)	0	0	9
Nevus	256	(35,3%)	470	(64,7%)	726
Queratosis actínica	137	(78,3%)	38	(21,7%)	175
Queratosis seborreica	95	(67,9%)	45	(32,1%)	140
Otros	225	(82,4%)	48	(17,6%)	273
Diagnóstico patológico NA	-	-	-	-	420
B. Razones de exclusión.					
Melanoma					Número de excluidos (% del total de melanomas)
Razones de exclusión					
Ausencia de pigmentación					75 (14,2%)
Ausencia de piel circundante normal					214 (40,5%)
Presencia de pelo					28 (5,3%)
Metástasis					29 (5,5%)
Localización en piel volar					23 (4,4%)
Lesión ulcerada					19 (3,6%)

embargo, existen algunas limitaciones que tendrán que ser superadas. Entre ellas se destacan los criterios de exclusión utilizados en la selección de las imágenes de las neoplasias cutáneas. A pesar de que a partir de nuestra base de datos se seleccionaron únicamente imágenes dermatoscópicas de alta calidad, solo el 34% de ellas no tenía ningún criterio de exclusión que permitiera su clasificación con los algoritmos de última generación de ML. Este hecho disminuye considerablemente la utilidad diagnóstica en la práctica clínica diaria de algunos sistemas de ML. Por otro lado, las lesiones de gran tamaño representan un problema importante para la utilización de los algoritmos de ML, ya que estas no se ajustan al diámetro de la mayoría de las lentes dermatoscópicas. Esto afecta la clasificación mediante la mayoría de algoritmos de ML, que requieren de la segmentación de la imagen para su análisis¹. Por otro lado, aunque en algunos trabajos se han propuesto métodos de detección/eliminación del vello⁵, el rendimiento de la mayoría de los sistemas de ML se ve perjudicado por su presencia.

Por último, cabe destacar que las bases de datos empleadas para el entrenamiento de los algoritmos actuales tienen poca representación de imágenes de lesiones de piel volar, lo que dificulta la correcta clasificación en estas localizaciones. Afortunadamente se está avanzando rápidamente para superar estas limitaciones en la selección de imágenes para la inteligencia artificial. Como muestra de ello, Yu et al.⁷ publicaron recientemente un trabajo en el que se utilizó el DCNN para la clasificación de melanoma acral y de nevus en piel volar. En el presente trabajo se consideraron las limitaciones de la mayoría, pero no de todos los sistemas de ML existentes en la actualidad.

Nuestro estudio muestra que los principales criterios de exclusión de imágenes de melanoma para clasificación mediante ML, fueron la ausencia de piel normal circundante y la ausencia de pigmentación. Gran parte de los melanomas se desarrollan sobre piel con daño actínico, por lo que la piel circundante puede ser patológica, lo que dificulta su análisis por la mayoría de los sistemas de ML actuales, ya que el

borde de la lesión no está bien definido⁵. Además, el melanoma amelanótico, que representa del 2% al 8% de todos los melanomas⁸, aún no se puede diagnosticar correctamente por la mayoría de los sistemas actuales de ML. Consideramos que todas estas limitaciones podrían resolverse a partir del diseño de algoritmos de ML que puedan trabajar con imágenes incompletas, incrementando el tamaño de las bases de datos y seleccionando un mayor número de imágenes de dermatoscopia que sean representativas de la práctica clínica habitual.

En conclusión, consideramos que los sistemas de ML, especialmente aquellos basados en el «deep learning», no solo convertirán el ML en una herramienta valiosa para el dermatólogo, sino también para la población en general. Sin embargo, estos sistemas deberán superar algunas limitaciones que les permitirán ampliar el espectro de las imágenes clasificables. El avance en los últimos años ha sido rápido y evidente ya que, incluso algunos de los criterios de exclusión que hemos tenido en cuenta en este trabajo han sido recientemente resueltos por algoritmos presentados en el Simposio Internacional ISIC³.

Financiación

Este estudio de la Unidad de Melanoma, Hospital Clínic, Barcelona fue financiado en parte por subvenciones del Fondo de Investigaciones Sanitarias P.I. 12/00840, PI15/00956 y PI15/00716 España; por el CIBER de Enfermedades Raras del Instituto de Salud Carlos III, España, cofinanciado por el Fondo Europeo de Desarrollo Regional (FEDER). Unión Europea. Una manera de hacer Europa; por el AGAUR 2014_SGR_603 y 2017_SGR_1134 del Gobierno catalán, España; por una beca de la «Fundació La Marató de TV3, 201331-30», Cataluña, España; por la Comisión Europea bajo el 6.º Programa Marco, Contrato n.º: LSHC-CT-2006-018702 (GenoMEL); por el programa CERCA/Generalitat de Catalunya y por una beca de investigación de la Fundación Científica de la Asociación Española Contra el Cáncer GCB15152978SOEN, España. Parte del trabajo se desarrolló en el edificio Centro Esther Koplowitz, Barcelona.

Conflicto de intereses

Los autores declaran no tener ningún conflicto de intereses.

Agradecimientos

Gracias a nuestros pacientes y a sus familias, que son la razón principal de nuestros estudios; a las enfermeras de la Unidad de Melanoma del Hospital Clínic de Barcelona, Daniel Gabriel, Pablo Iglesias y María E. Moliner por ayudar a recopilar datos de pacientes y a Paul Hetherington por ayudar con la edición y la corrección al inglés del manuscrito.

Bibliografía

1. Gareau DS, Correa da Rosa J, Yagerman S, Carucci JA, Gulati N, Hueto F, et al. Digital imaging biomarkers feed machine learning for melanoma screening. *Exp Dermatol.* 2017;26:615–8, doi:10.1111/exd.13250.
2. Fujisawa Y, Otomo Y, Ogata Y, Nakamura Y, Fujita R, Ishitsuka Y, et al. Deep-learning-based, computer-aided classifier developed with a small dataset of clinical images surpasses board-certified dermatologists in skin tumour diagnosis. *Br J Dermatol.* 2018 [Publicación electrónica] doi:10.1111/bjd.16924.
3. Marchetti MA, Codella NCF, Dusza SW, Gutman DA, Helba B, Kalloo A, et al. Results of the 2016 International Skin Imaging Collaboration International Symposium on Biomedical Imaging challenge: Comparison of the accuracy of computer algorithms to dermatologists for the diagnosis of melanoma from dermoscopic images. *J Am Acad Dermatol.* 2018;78:270–7, <http://dx.doi.org/10.1016/j.jaad.2017.08.016>.
4. Tschandl P, Rosendahl C, Kittler H. The HAM10000 dataset, a large collection of multi-source dermoscopic images of common pigmented skin lesions. *Sci Data.* 2018;5:180161, doi:10.1038/sdata.2018.161.
5. Celebi M, Wen Q, Iyatomi H, Shimizu K, Zhou H, Schaefer G. A state-of-the-art survey on lesion border detection in dermoscopy images. In: Celebi ME, Mendonca T, Marques J, eds. *Dermoscopy image analysis*. Boca Raton, FL: CRC Press; 2015.
6. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature.* 2017;542:115–8, doi:10.1038/nature21056.
7. Yu C, Yang S, Kim W, Jung J, Chung KY, Lee SW, et al. Acral melanoma detection using a convolutional neural network for dermoscopy images. *PLoS One.* 2018;13:e0193321, <http://dx.doi.org/10.1371/journal.pone.0193321>.
8. Pizzichetta MA, Kittler H, Stanganelli I, Ghigliotti G, Corradin MT, Rubegni P, et al. Dermoscopic diagnosis of amelanotic/hypomelanotic melanoma. *Br J Dermatol.* 2017;177:538–40, doi:10.1111/bjd.15093.