

Hacia modelos de aprendizaje automático más robustos y equitativos en dermatología

Disparidades en representación con respecto al tono de piel



Celia Cintas PhD



JIS Go Live
2021



II SIMPOSIO
DE INTELIGENCIA
ARTIFICIAL EN SALUD

Photo: TOTTO RENNA

El equipo

Los trabajos aquí presentados fueron desarrollados gracias a la excelente colaboración entre lxs investigadorxs, médicxs, estudiantes de maestría y doctorandxs.



IBM Research



Stanford
University



Memorial Sloan Kettering
Cancer Center

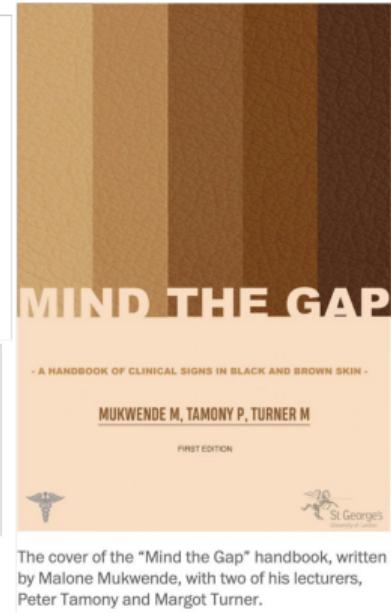
Carnegie Mellon University
Africa College of Engineering

IBM Research | Africa

Disparidades en Dermatología

- En poblaciones afrodescendientes, el melanoma es comúnmente diagnosticado en etapas tardías [MSL⁺17, WEK⁺11].
- La escasez de imágenes de las manifestaciones cutáneas de COVID-19 en pacientes latinxs y afrodescendientes es un problema, ya que dificulta la identificación de COVID-19 tanto para los dermatólogos como para el público [LJZ⁺20].

The screenshot shows a news article from The New York Times. At the top, there is a navigation bar with three horizontal lines, the text "The New York Times", and an "Account" dropdown menu. Below the header, the main title is "Dermatology Has a Problem With Skin Color". A subtext below the title reads: "Common conditions often manifest differently on dark skin. Yet physicians are trained mostly to diagnose them on white skin." The article is categorized under "HEALTH". The author's name is "By USHA LEE MCFARLING @ushamcfarling / JULY 21, 2020". There is also a "Report this" link. The background of the article area features vertical stripes of different skin tones.



The cover of the "Mind the Gap" handbook, written by Malone Mukwende, with two of his lecturers, Peter Tamony and Margot Turner.

Algunas preguntas...

- ❑ ? Las imágenes que usamos para **entrenar modelos** están **sesgados respecto a tonos de piel?**
- ❑ ? Podemos **cuantificar** de forma automática cuál es la **representación** de distintos tonos de piel en **materiales educativos** utilizado para formar dermatólogos y médicos de atención primaria?
- ❑ ? Nuestros modelos son **robustos** frente a cambios en los datos de entrada ?

IBM Research | Africa

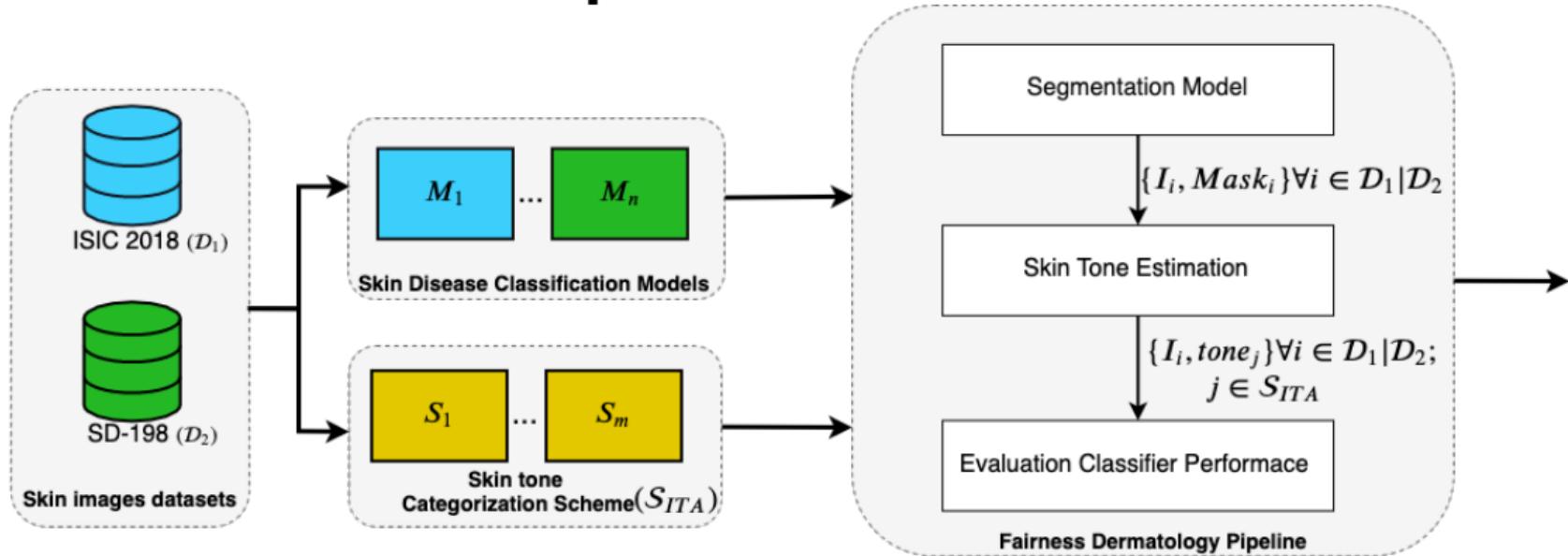
Trabajos relacionados

- Diagnóstico de enfermedades de la piel mediante aprendizaje automático
 - 1 Aprendizaje Automático para el diagnóstico de melanoma supera la performance de dermatólogos [CNP⁺16]
 - 2 ISIC challenges (<https://www.isic-archive.com/>)
- Estudios de inequidad en varias aplicaciones de visión computacional con respecto al tono de piel
 - 1 Análisis automatizado de imágenes faciales para la clasificación de género [BG18]
 - 2 Sistemas de detección de peatones en vehículos autónomos [WHM19]



IBM Research | Africa

Cómo evaluamos estos modelos respecto a distintos tonos de piel?



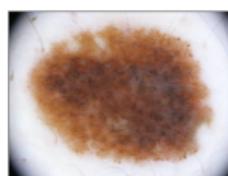
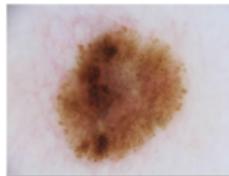
Kinyanjui, N.M., Odonga, T., Cintas, C., Codella, N.C., Panda, R., Sattigeri, P. and Varshney, K.R. Fairness of Classifiers Across Skin Tones in Dermatology. In MICCAI 2020 - International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 320-329).

IBM Research | Africa

Qué datos usamos para estudiar los modelos entrenados?

ISIC 2018 [Int18]

- 10015 imágenes dermatoscópicas
- 7 clases de enfermedades
- 2594 imágenes con máscaras de segmentación manuales



SD-198 [LGL⁺19]

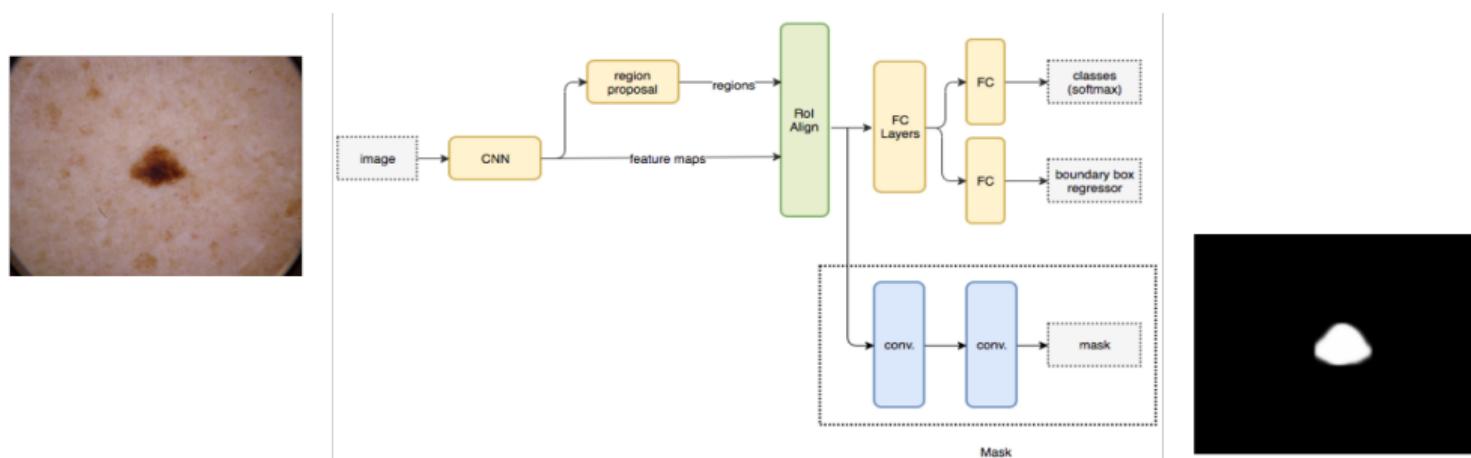
- 6548 imágenes clínicas
- 198 clases de enfermedades
- Sin máscaras de segmentación



IBM Research | Africa

Segmentación Automática para obtener la región de la piel sin lesiones

- 1 Adaptamos un modelo pre-entrenado llamado Mask R-CNN [HGDG17].
- 2 Aplicamos umbrales para obtener máscaras binarias y extraer el contorno para ambos conjuntos de datos.



IBM Research | Africa

Cómo estimamos el tono de piel de la región sin lesiones?

- 1 Dados los pixeles de regiones sanas, le asignamos un valor asociado al tono de piel.
 - 1 Usamos *Individual Typology Angle* (ITA) [WWdPR15]. Este valor está altamente correlacionado con el índice de melanina.
 - 2 $\text{ITA} = \tan^{-1} \left(\frac{L-50}{b} \right) \times \frac{180}{\pi}$, donde L indica la luminancia y b cuantifica la cantidad de amarillo.
- 2 Finalmente, agrupamos estos valores en categorías [CSD⁺15].

ITA Range	Skin Tone Category	Abbreviation
$\text{ITA} > 55^\circ$	Very Light	very_lt
$48^\circ < \text{ITA} \leq 55^\circ$	Light 2	lt2
$41^\circ < \text{ITA} \leq 48^\circ$	Light 1	lt1
$34.5^\circ < \text{ITA} \leq 41^\circ$	Intermediate 2	int2
$28^\circ < \text{ITA} \leq 34.5^\circ$	Intermediate 1	int1
$19^\circ < \text{ITA} \leq 28^\circ$	Tanned 2	tan2
$10^\circ < \text{ITA} \leq 19^\circ$	Tanned 1	tan1
$\text{ITA} \leq 10^\circ$	Dark	dark

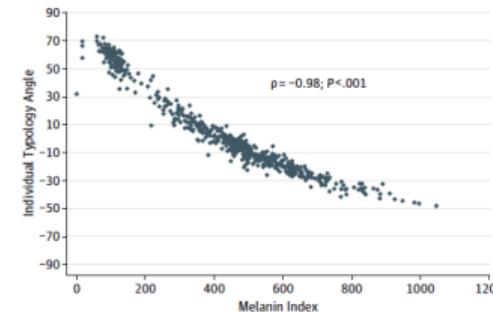


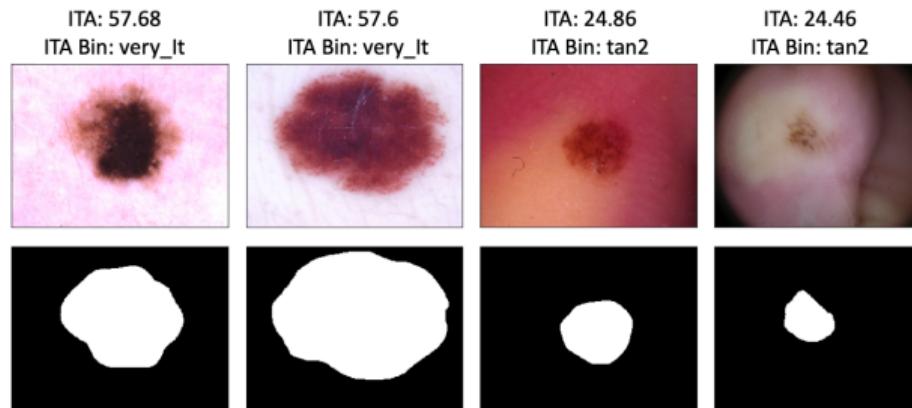
Figura de [WWdPR15].

IBM Research | Africa

Resultados

Modelo de segmentación y estimación de tono de piel en ISIC 2018

El modelo Mask R-CNN tiene una precisión **0.956**, una tasa de falsos negativos de **0.024**, y un error absoluto medio en la estimación de los valores ITA de **0.428** grados [KOC⁺19].

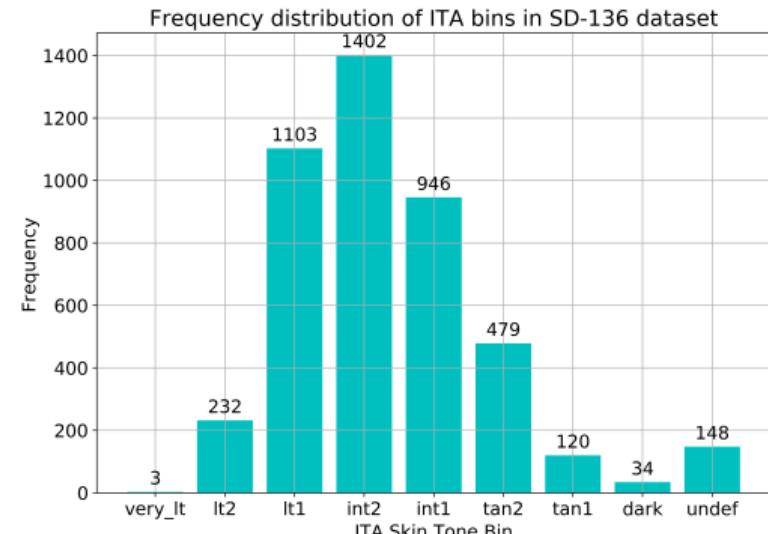
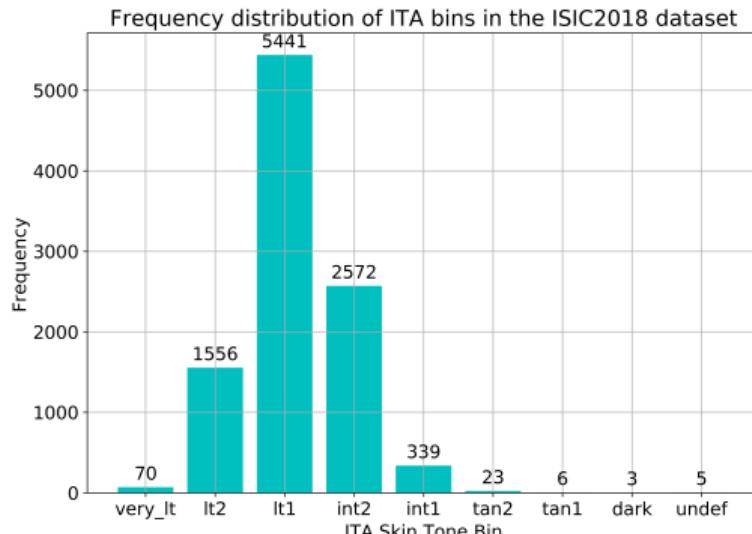


Resultados (Cont.)



Distribución con respecto a tonos de piel

En los conjuntos de datos explorados para ML y dermatología, los tonos de piel oscuros están subrepresentados ya que la mayor parte de los datos corresponden a poblaciones con tonos de piel claros.





Sabemos con que datos las computadoras están entrenadas .. pero con qué se preparan a los profesionales?

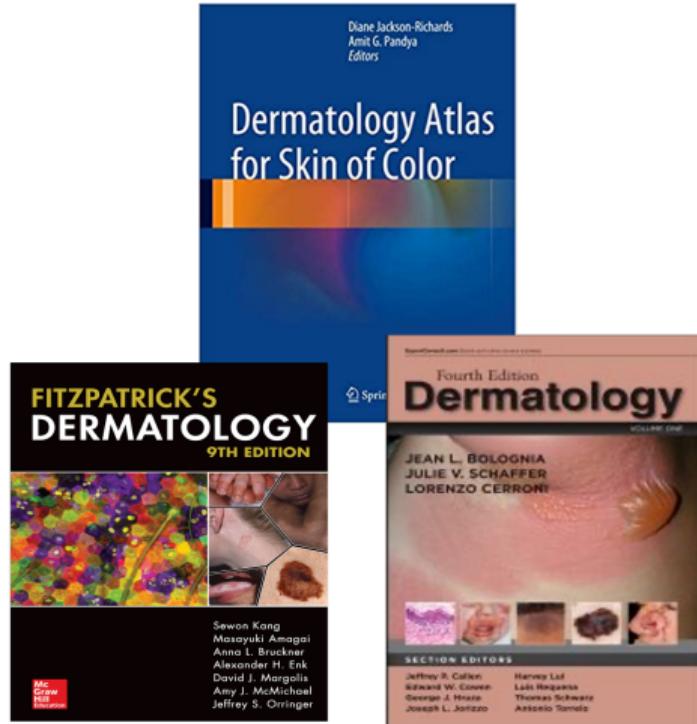


Ahora sabemos que los datos utilizados para **entrenar modelos** en ML están formados **mayoritariamente por población caucásica**.

Podemos **cuantificar** de forma automática cuál es la **representación** de distintos tonos de piel en **materiales educativos** utilizados para formar dermatólogos y médicos de atención primaria?

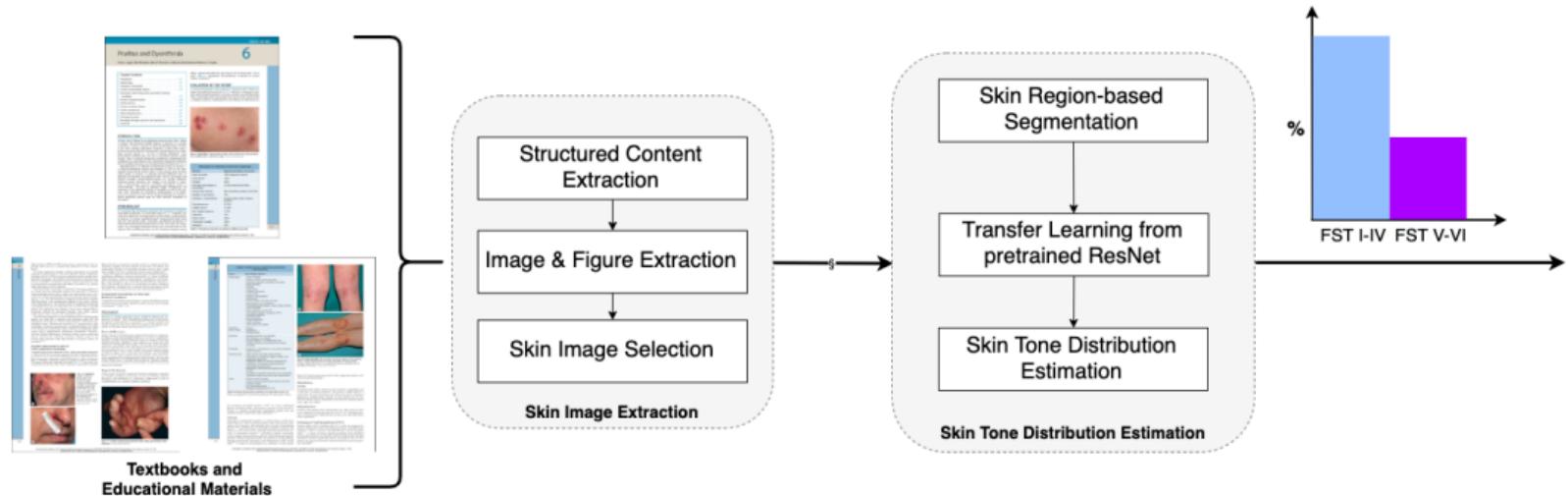
Sesgos con respecto al tono de piel en material educativo

- 1 La anotación manual del tono de la piel no es factible para un gran corpus de materiales educativos en dermatología debido a su subjetividad y naturaleza laboriosa.
- 2 Imágenes de diferentes tipos (diagramas, dibujos, imágenes de microscopios, etc.)



IBM Research | Africa

Cómo evaluamos los materiales académicos respecto a distintos tonos de piel?



- **Entrada:** Libros, artículos o capítulos de dermatología.
- **Salida:** Representación de tonos de piel binaria (FST I-IV) y (FST V-VI).

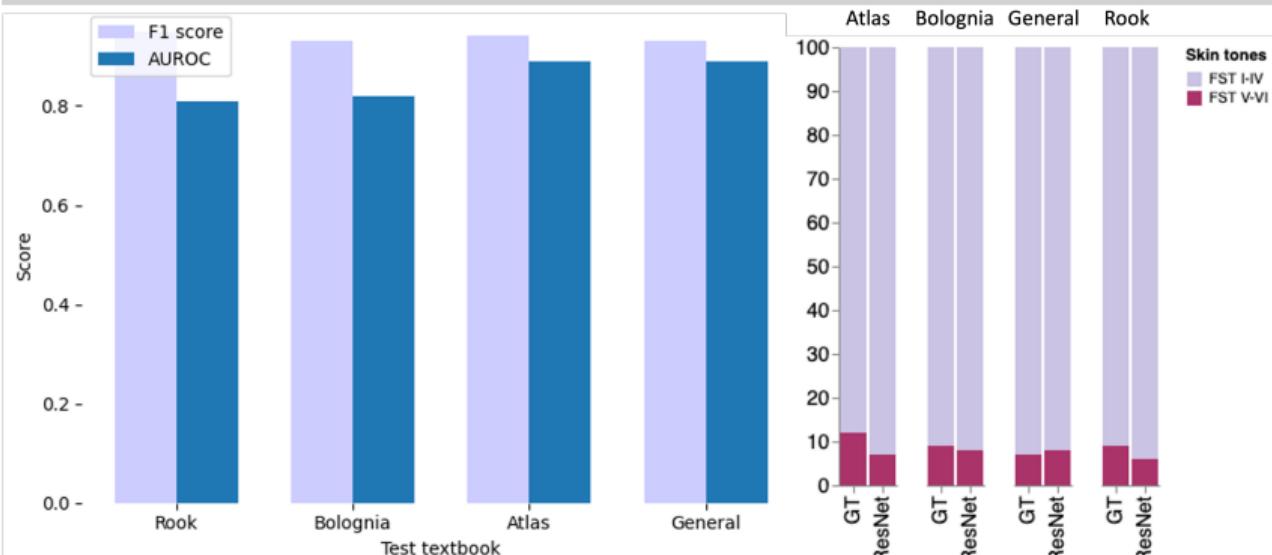
Tadesse ,A. G, Cintas, C., Daneshjou R., Varshney ,K., Staar, P., Speakman, S. Andrews, K., Agunwa, C., Jia, J. Bailey, E., Adelekun, A., Lipoff , B.J., Onyekaba, G., Rottemberg, V., Zou, J. Skin-tone Representation Analysis in Dermatology Academic Materials. AMIA 2021. Versión extendida en revisión.

IBM Research | Africa

Resultados Preliminares

Distribución de tonos de piel en libros de dermatología

En cada libro utilizado para validación de la extracción automática de distribuciones de tonos de piel observamos una sub-representación de tonos oscuros (FST V-VI), en todos los casos $\approx 10\%$.



Detección de valores atípicos en modelos de aprendizaje profundo

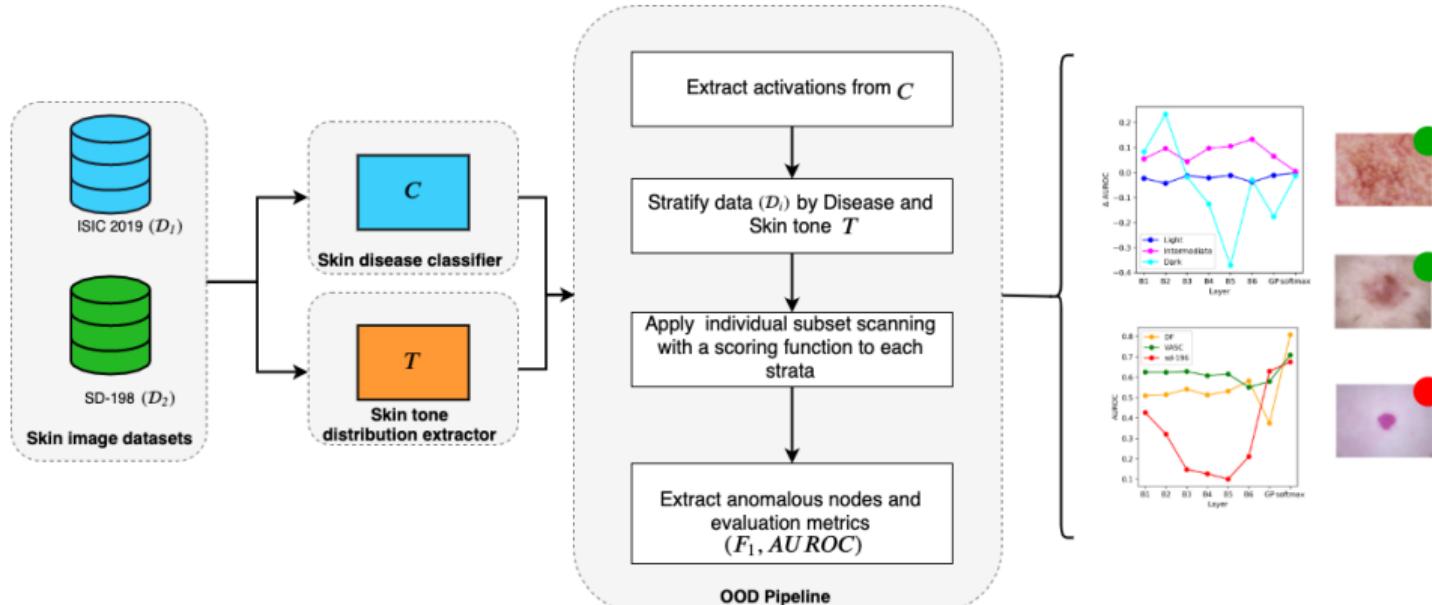
Los recientes avances en el aprendizaje profundo han dado lugar a mejoras en el desarrollo de la clasificación automatizada de condiciones de la piel. A medida que observamos un **creciente interés** en estos modelos en **dermatología**, es crucial abordar aspectos como **robustez** y **equidad** de estas soluciones.

En este trabajo evaluamos dos aspectos:

- 1 Cambios en el ambiente clínico (distinto hardware, distinto protocolo de toma de datos, etc.)
- 2 Condiciones de piel desconocidas.



Como detectamos entradas atípicas en modelos de aprendizaje profundo?



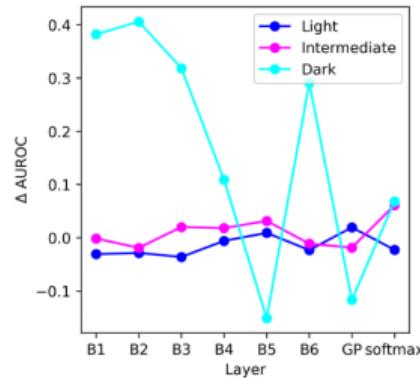
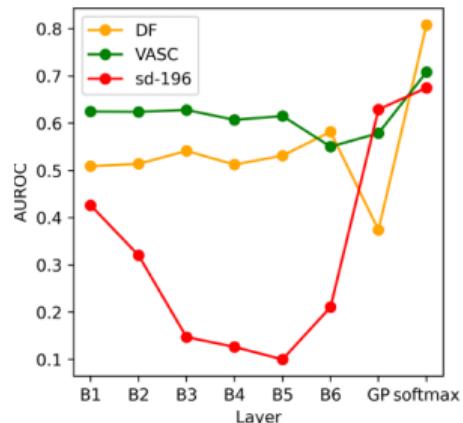
Kim, H., Tadesse, G.A., Cintas, C., Speakman, S. and Varshney, K., 2021. Out-of-Distribution Detection in Dermatology using Input Perturbation and Subset Scanning. In 6th Outlier Detection Description Workshop - KDD 2021. [KTC⁺21] (Trabajo extendido en revisión)

IBM Research | Africa

Resultados Preliminares

Resultados en ambos escenarios

Los patrones entre capas son **distintos** cuando queremos **detectar una clase nueva** a un **cambio en el protocolo** de la toma de dato [KTC⁺21].



Equidad en detectores de OOD

La performance varía con muestras de tonos FST V-VI. Esta inestabilidad puede deberse en parte a que la red está entrenada con datos que **carece en gran medida de** muestras de **tonos de piel oscuros**.

Conclusiones y trabajo futuro

- 1 Los datos utilizados para entrenar modelos están **mayoritariamente** conformadas por ejemplos de **tonos de piel claros** (valores ITA [34.5°, 48°]).
- 2 En todos los materiales académicos evaluados automáticamente encontramos una **sub-representación de tonos FST V-VI ≈ 10%**.
- 3 **Limitaciones:** Segmentación no excluye completamente áreas de la piel con lesiones. Binarización de tonos de piel **no captura toda la diversidad de tonos de piel**.
- 4 Siguientes pasos: **Piloto** de la aplicación para estimar representación en materiales académicos.

IBM Research | Africa

Gracias, Asante, Thanks!



IBM Research



Memorial Sloan Kettering
Cancer Center.

Carnegie Mellon University
Africa College of Engineering

celia.cintas@ibm.com  @RTFMCelia

IBM Research | Africa

References I

-  Joy Buolamwini and Timnit Gebru, *Gender shades: Intersectional accuracy disparities in commercial gender classification*, Proc. Conf. Fair. Account. Transp., February 2018, pp. 77–91.
-  Noel C. F. Codella, Quoc-Bao Nguyen, Sharath Pankanti, David A. Gutman, Brian Helba, Allan C. Halpern, and John R. Smith, *Deep learning ensembles for melanoma recognition in dermoscopy images*, IBM J. Res. Dev. 61 (2016), no. 4/5, 5.
-  Giuseppe R. Casale, Anna Maria Siani, Henri Diémoz, Giovanni Agnesod, Alfio V. Parisi, and Alfredo Colosimo, *Extreme UV index and solar exposures at Plateau Rosà (3500 m a.s.l.) in Valle d'Aosta Region, Italy*, Sci. Total Environ. 512–513 (2015), 622–630.
-  Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross B Girshick, *Mask r-cnn. corr abs/1703.06870 (2017)*, arXiv preprint arXiv:1703.06870 (2017).
IBM Research-Africa

References II

-  International Skin Imaging Collaboration, *ISIC 2018: Skin lesion analysis towards melanoma detection*, 2018, Available:
<https://challenge2018.isic-archive.com/>.
-  Newton M Kinyanjui, Timothy Odonga, Celia Cintas, Noel CF Codella, Rameswar Panda, Prasanna Sattigeri, and Kush R Varshney, *Estimating skin tone and effects on classification performance in dermatology datasets*, arXiv preprint arXiv:1910.13268 (2019).
-  Hannah Kim, Girmaw Abebe Tadesse, Celia Cintas, Skyler Speakman, and Kush Varshney, *Out-of-distribution detection in dermatology using input perturbation and subset scanning*, arXiv preprint arXiv:2105.11160 (2021).

References III

-  J. Lin, Z. Guo, D. Li, X. Hu, and Y. Zhang, *Automatic classification of clinical skin disease images with additional high-level position information*, 2019 Chinese Control Conference (CCC), 2019, pp. 8606–8610.
-  JC Lester, JL Jia, L Zhang, GA Okoye, and E Linos, *Absence of skin of colour images in publications of covid-19 skin manifestations*, British Journal of Dermatology (2020).
-  Krishnaraj Mahendararaj, Komal Sidhu, Christine S. M. Lau, Georgia J. McRoy, Ronald S. Chamberlain, and Franz O. Smith, *Malignant melanoma in African–Americans: A population-based clinical outcomes study involving 1106 African–American patients from the surveillance, epidemiology, and end result (SEER) database (1988–2011)*, Medicine 96 (2017), no. 15, e6258.

References IV

-  Xiao-Cheng Wu, Melody J. Eide, Jessica King, Mona Saraiya, Youjie Huang, Charles Wiggins, Jill S. Barnholtz-Sloan, Nicolle Martin, Vilma Cokkinides, Jacqueline Miller, Pragna Patel, Donatus U. Ekwueme, and Julian Kim, *Racial and ethnic variations in incidence and survival of cutaneous melanoma in the United States, 1999-2006*, J. Am. Acad. Dermatol. 65 (2011), no. 5, S26.e1–S26.e13.
-  Benjamin Wilson, Judy Hoffman, and Jamie Morgenstern, *Predictive inequity in object detection*, arXiv:1902.11097, February 2019.
-  Marcus Wilkes, Caradee Y. Wright, Johan L. du Plessis, and Anthony Reeder, *Fitzpatrick skin type, individual typology angle, and melanin index in an African population*, JAMA Dermatol. 151 (2015), no. 8, 902–903.