ORIGINAL ARTICLE

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

Michael A. Marchetti, MD,[a] Noel C. F. Codella, PhD,[b] Stephen W. Dusza, DrPH,[a] David A. Gutman, MD, PhD,[c] Brian Helba,[d] Aadi Kalloo, MHS,[a] Nabin Mishra, PhD,[e] Cristina Carrera, MD, PhD,[f] M. Emre Celebi, PhD,[g] Jennifer L. DeFazio, MD,[a] Natalia Jaimes, MD,[h,i] Ashfaq A. Marghoob, MD,[a] Elizabeth Quigley, MD,[a] Alon Scope, MD,[a,j] Oriol Yélamos, MD,[a] and Allan C. Halpern, MD,[a] for the International Skin Imaging Collaboration

New York, Yorktown Heights, and Clifton Park, New York; Atlanta, Georgia; Rolla, Missouri; Barcelona, Spain; Conway, Arkansas; Medellín, Colombia; Miami, Florida; and Tel Aviv, Israel

Background: Computer vision may aid in melanoma detection.

Objective: We sought to compare melanoma diagnostic accuracy of computer algorithms to dermatologists using dermoscopic images.

Methods: We conducted a cross-sectional study using 100 randomly selected dermoscopic images (50 melanomas, 44 nevi, and 6 lentigines) from an international computer vision melanoma challenge dataset (n = 379), along with individual algorithm results from 25 teams. We used 5 methods (nonlearned and machine learning) to combine individual automated predictions into "fusion" algorithms. In a companion study, 8 dermatologists classified the lesions in the 100 images as either benign or malignant.

Results: The average sensitivity and specificity of dermatologists in classification was 82% and 59%. At 82% sensitivity, dermatologist specificity was similar to the top challenge algorithm (59% vs. 62%, $P = .68$) but lower than the best-performing fusion algorithm (59% vs. 76%, $P = .02$). Receiver operating characteristic

From the Dermatology Service,[a] Department of Medicine, Memorial Sloan Kettering Cancer Center, New York; IBM Research Division,[b] Thomas J. Watson Research Center, Yorktown Heights; Departments of Neurology, Psychiatry, and Biomedical Informatics,[c] Emory University School of Medicine, Atlanta; Kitware Inc,[d] Clifton Park; Stoecker & Associates,[e] Rolla; Melanoma Unit,[f] Department of Dermatology, Hospital Clinic, Institut d'Investigacions Biomèdiques August Pi i Sunyer, CIBER de Enfermedades Raras, Instituto de Salud Carlos III, University of Barcelona; Department of Computer Science,[g] University of Central Arkansas, Conway; Dermatology Service,[h] Aurora Centro Especializado en Cáncer de Piel, Medellín; Department of Dermatology and Cutaneous Surgery,[i] University of Miami Miller School of Medicine, Miami; and the Department of Dermatology,[j] Sheba Medical Center, Sackler School of Medicine, Tel Aviv University.

Drs Marchetti and Codella contributed equally to this article.

Supported in part by the National Institutes of Health/National Cancer Institute Cancer Center Support Grant P30 CA008748.

Disclosure: Dr Codella is an employee of IBM and an IBM stockholder. Dr Scope is a consultant to Emerald Inc. Drs Marchetti, Dusza, Halpern, Marghoob, DeFazio, Yélamos, Carrera, Jaimes, Mishra, Kalloo, Quigley, Gutman, Helba, and Celebi have no conflicts of interest to declare.

Preliminary data of this study were presented at the Annual Meeting of the Society for Melanoma Research Congress, November 8, 2016, Boston, Massachusetts.

Accepted for publication August 7, 2017.

Correspondence to: Allan C. Halpern, MD, Dermatology Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, 16 E 60th St, New York, NY 10022. E-mail: halperna@mskcc.org.

Published online September 29, 2017.

0190-9622/$36.00

© 2017 by the American Academy of Dermatology, Inc.

http://dx.doi.org/10.1016/j.jaad.2017.08.016

area of the top fusion algorithm was greater than the mean receiver operating characteristic area of dermatologists (0.86 vs. 0.71, $P = .001$).

Limitations: The dataset lacked the full spectrum of skin lesions encountered in clinical practice, particularly banal lesions. Readers and algorithms were not provided clinical data (eg, age or lesion history/symptoms). Results obtained using our study design cannot be extrapolated to clinical practice.

Conclusion: Deep learning computer vision systems classified melanoma dermoscopy images with accuracy that exceeded some but not all dermatologists. (J Am Acad Dermatol http://dx.doi.org/10.1016/j.jaad.2017.08.016.)

Key words: computer algorithm; computer vision; dermatologist; International Skin Imaging Collaboration; International Symposium on Biomedical Imaging; machine learning; melanoma; reader study; skin cancer.

The early diagnosis of melanoma remains challenging.[1] Estimates of the sensitivity of dermatologists for melanoma in reader studies were 70% for the Nevisense trial[2] and 78% for the MelaFind trial.[3] In addition, because nonphysicians detect the majority of melanomas[4] and because population-based melanoma screening by clinicians is not recommended in the United States,[5] there is not only interest in the development of automated image analysis algorithms to help dermatologists classify dermoscopic images, but also to aid laypersons or nondermatology physicians in melanoma detection.[6-13] To date, the lack of a large, public dataset of skin images has limited the ability to directly compare the diagnostic performance of competing automated image analysis approaches against clinicians.

To address this limitation, the International Skin Imaging Collaboration (ISIC) Melanoma Project created an open-access archive of dermoscopic images of skin lesions for education and research.[14] We describe the melanoma classification results from a challenge conducted by the ISIC Archive[15] at the 2016 International Symposium on Biomedical Imaging (ISBI) involving 25 competing teams.[16] We further performed a companion reader study with 8 experienced dermatologists on a subset of images; these results served as a reference comparator to the automated algorithm approaches.

CAPSULE SUMMARY

- Computer vision has shown promise in medical diagnosis.
- A machine learning fusion algorithm using predictions from 16 algorithms exceeded the performance of most dermatologists in the classification of 100 dermoscopic images of melanomas and nevi.
- These results should not be extrapolated to clinical practice until validation in prospective studies.

MATERIALS AND METHODS

Institutional review board approval

Institutional review board approval was obtained at Memorial Sloan Kettering and the study was conducted in accordance with the Helsinki Declaration.

ISBI 2016 melanoma detection challenge dataset

Details of the challenge tasks, evaluation criteria, timeline, and participation have been previously described.[15,17,18] In December 2015, 1552 lesions were chosen from ~12,000 dermoscopic images in the ISIC Archive; after excluding 273 for inadequate image quality, 1279 lesions (248 [19.3%] melanomas and 1031 [80.7%] nevi or lentigines) were included. Images were excluded because of poor focus or if they included multiple lesions or lesions that encompassed the entire field of view. The dataset was randomly divided into training (n = 900 [19.2% melanomas]) and test (n = 379 [19.8% melanomas]) sets. All melanomas and a majority of the nevi/lentigines (n = 869, 84%) had been histopathologically examined. Nonhistopathologically examined nevi (n = 162) originated from a longitudinal study of children; selection from this dataset was biased to include lesions with the largest diameters, and all images were reviewed by ≥2 dermatologists to confirm their benign nature.[19] Images used in this challenge were obtained with multiple camera/dermatoscope combinations and originated from >12 dermatology clinics around the world.

Twenty-five teams participated in the challenge, all of which used deep learning, a form of machine learning that uses multiple processing layers to automatically identify increasingly abstract concepts present in data. Computer algorithms were ranked using average precision, which corresponds to the

Download English Version:

https://daneshyari.com/en/article/8715304

Download Persian Version:

https://daneshyari.com/article/8715304

Daneshyari.com