Title
Expert Performance in Visual Differentiation of Bacterial and Fungal Keratitis

Journal
Ophthalmology, 129(2)

ISSN
0161-6420

Authors
Redd, Travis K
Prajna, N Venkatesh
Srinivasan, Muthiah
et al.

Publication Date
2022-02-01

DOI
10.1016/j.ophtha.2021.09.019

Peer reviewed

Published in final edited form as:
Ophthalmology. 2022 February ; 129(2): 227–230. doi:10.1016/j.ophtha.2021.09.019.

Expert Performance in Visual Differentiation of Bacterial and Fungal Keratitis

Travis K Redd, MD, MPH[1], **N Venkatesh Prajna, MD**[2], **Muthiah Srinivasan, MD**[2], **Prajna Lalitha, MD**[2], **Tiru Krishnan, MD**[2], **Revathi Rajaraman, MD**[2], **Anitha Venugopal, MD**[2], **Brandon Lujan, MD**[1], **Nisha Acharya, MD**[3], **Gerami D Seitzman, MD**[3], **Jennifer Rose-Nussbaumer, MD**[4], **Thomas M Lietman, MD**[3], **J Peter Campbell, MD, MPH**[1], **Jeremy D Keenan, MD, MPH**[3] **Corneal Ulcer Image Interpretation Study Group**

[1]Casey Eye Institute, Oregon Health & Science University, Portland, OR, USA.

[2]Aravind Eye Hospital, Madurai, India.

[3]Francis I. Proctor Foundation, University of California San Francisco, San Francisco, CA, USA.

[4]Kaiser Permanente, Redwood City, CA, USA.

Abstract

This study quantifies the performance of an international cohort of cornea specialists in image-based differentiation of bacterial and fungal keratitis, identifying significant regional variation and establishing a reference standard for comparison against machine learning models.

Keywords

Infectious Keratitis; Corneal Ulcer; Bacterial Keratitis; Fungal Keratitis

Prompt identification of the etiology of infectious keratitis is important to guide antimicrobial therapy, but culture results are not available immediately and are falsely negative in 40-60% of cases.[1] Novel diagnostic modalities including artificial intelligence (AI) models for image-based diagnosis of corneal ulcers have been proposed to address this gap, with promising preliminary results.[2,3] However, before these models can be implemented they must be compared against human clinical impression of the cause of infection, which is the current standard of care for determining empiric therapy in the absence of microbiologic data.

Prior studies have suggested even expert cornea specialists are only able to correctly distinguish bacterial from fungal keratitis 2/3 to 3/4 of the time based on clinical impression, but these were small surveys conducted among discrete populations of cornea specialists using only categorical analytic measures such as accuracy.[4,5] Quantifying respondents' estimated probability of their prediction allows determination of receiver

Corresponding author contact information: Travis Redd, MD, MPH, Casey Eye Institute, Oregon Health & Science University, 515 SW Campus Dr, Portland, OR, 97239, redd@ohsu.edu.

Conflict of interest statement: No conflicting relationships exists for any author operating characteristics (ROC) and the area under the ROC curve (AUC), which provide a more informative evaluation of predictive performance and enable direct comparison against AI models. Herein we measure human AUC using the largest international cohort of expert cornea clinicians yet assembled for image-based classification of corneal ulcers.

Several large clinical trials for infectious keratitis have been conducted at the Aravind Eye Care System in South India.[6,7] Each corneal ulcer in these trials was microbiologically proven to be either bacterial or filamentous fungal keratitis, and each subject underwent corneal photography at initial presentation using handheld Nikon (Tokyo, Japan) D-series digital single lens reflex cameras according to a standardized photography protocol, resulting in a large database of bacterial and fungal corneal ulcer images from South India. We obtained a testing set from this database consisting of 100 images from 100 ulcers using stratified random sampling to ensure balanced classes (50 bacterial images and 50 fungal images).

Cornea specialists were recruited from the Casey Eye Institute, the Proctor Foundation at University of California San Francisco, Aravind, and kera-net (https://corneasociety.org/discussions) via email correspondence. Subjects provided an estimated probability that each image in the testing set represented fungal rather than bacterial keratitis using a secure web-based image grading platform (https://tctc.ohsu.edu). Aravind physicians who cared for any of the subjects in the above randomized trials were excluded to ensure participant responses were based only on information presented in the photographs. No other clinical or historical information was provided. To account for geographic variability in the prevalence of fungal keratitis and resulting differences in pre-test probability assumed by respondents from varying regions, all graders were informed that in this image set 50% of cases represented culture-proven bacterial ulcers and 50% were from culture-proven fungal infections. Ten images (five fungal, five bacterial) were presented twice to each grader to allow measurement of test-retest reliability. This study was approved by the Institutional Review Board at Oregon Health & Science University and adhered to the tenets of the Declaration of Helsinki.

66 cornea specialists from 16 countries, the majority of whom practice primarily in the United States (50%) or India (18%) (Table S1, available at www.aaojournal.org). Individual expert AUCs were highly variable, ranging from 0.39 to 0.82 with a mean of 0.61. The mean individual AUC varied significantly according to practice location ($P < 0.001$ [one-way ANOVA comparing all 16 countries]), with experts practicing in India significantly outperforming their colleagues practicing in other countries on this testing set of ulcers from South India (AUC 0.72 vs. 0.59, $P < 0.001$; Figure S1, available at www.aaojournal.org). The intraclass correlation coefficient among all respondents was 0.71 (95% CI 0.67-0.75), indicating moderate test-retest reliability.

To estimate overall human performance and establish the benchmark for comparison against AI model performance, we determined the *ensemble estimated probability* (the mean predicted probability across multiple respondents). The AUC of the ensemble estimated probability among all 66 respondents was 0.72 (95% CI 0.63-0.82). The ensemble estimated probability among all Indian experts achieved an AUC of 0.81, which was statistically significantly higher than among non-Indian experts (0.68; P < 0.001 [DeLong method]; Figure 1). In this context the terms “Indian” and “non-Indian” are used to indicate a participant’s primary practice location, not their ethnic, racial, or cultural affiliations. Subgroup analysis demonstrated that the ensemble estimated probability among Indian experts was statistically significantly more accurate for identifying fungal ulcers (76%) compared to non-Indian graders (accuracy = 49%; P < 0.001 [McNemar’s test]; Figure S2, available at www.aaojournal.org). There was no difference between groups in the accuracy for identifying bacterial ulcers (71% vs. 71%; P = 1; McNemar’s test). This difference is likely attributable to Indian experts’ greater familiarity with fungal keratitis, which accounts for nearly half of corneal ulcers in South India but is rare in temperate regions including most of the United States and Europe.

In this study graders were only presented a single image, which likely contains less information than in-person slit lamp examination and clinical history would provide and may explain the relatively poor overall performance. However, this is identical to the amount of information available to computer vision models and thus allows direct comparison between the two modalities. Further, published evidence indicates human performance does not significantly improve when obtaining clinical history and performing slit lamp examination.[5] Nonetheless, future implementations of prediction models will ideally incorporate information obtained from the clinical history and other aspects of the examination in addition to imaging data into a multivariate risk model to maximize predictive accuracy. Evaluation of human and AI performance must also be investigated for other causes of infection, including viral and parasitic etiologies. Finally, other covariates including expert experience with infectious keratitis may influence performance; future studies may benefit from assessing this and other unmeasured factors.

This large international survey establishes the overall performance and regional variability among expert corneal specialists for image-based determination of the underlying etiology of corneal ulcers. These findings establish the benchmark against which AI models will be compared, and reinforces the importance of considering geographic variability in ulcer epidemiology and human performance when evaluating and implementing novel diagnostic models.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Financial Support:

This study was supported by the National Institutes of Health (NIH K12EY027720 and core grant P30EY10572) and unrestricted departmental funding provided by Research to Prevent Blindness. The funding organization had no role in the design or conduct of this research.

Appendix:

Study group members:

Drs. Diana Alvarez-Melloni, Menen Ayalew, Ashwin Balasubramanian, Elsie Chan, Matilda Chan, Meenu Chaudhary, Thomas Chia, James Chodosh, YY Choong, Joseph Christenbury, Josephine Christy, John Clements, John Dart, Mohammad Dastjerdi, Matthew Denny, Sathish Devarajan, Mohamed Elghobaier, Chris Estopinal, Preethika Gandhi, Nikhil Gokhale, Colleen Halfpenny, Rossen Hazarbassanov, Natalie Hernandez, Anna Hovakimyan, David Hwang, Frank Hwang, Tomas Jaeschke, Vishal Jhanji, Faris Karas, Divya Karthik, Camila Kase, Lakshmi Kattana, Tyson Kim, Aaleya Koreishi, David Liang, Christine Martinez, Rafael Martinez-Costa, Stephen McLeod, Jodhbir S Mehta, Michael Mimouni, Adam Moss, Afshan Nanji, Nathan Nataneli, Vasudha Panday, Sayali Pradhan, Ying Qian, Naveen Rao, Julie Schallhorn, Ruti Sella, Suvitha Selvaraj, David Spokes, Neha Shaik, Nakul Shekhawat, Alan Sugar, Audrey Talley Rostov, Napaporn Tananuvat, Chulaluck Tangmonkongvoragul, Tanya Trinh, Sonal Tuli, Phit Upaphong, Bart van Dooren, Manoj Vasudevan, Elizabeth Viriya, and Maria Woodward.

References

1. Mcleod SD, Kolahdouz-isfahani A, Rostamian K, Flowers CW, Lee PP, Mcdonnell P. The Role of Smears, Cultures, and Antibiotic Sensitivity Testing in the Management of Suspected Infectious Keratitis. Ophthalmology. 1996;103:23–28. [PubMed: 8628555]
2. Ghosh AK, Thammasudjarit R, Jongkhajornpong P, Attia J, Thakkinstian A. Deep Learning for Discrimination between Fungal Keratitis and Bacterial Keratitis: DeepKeratitis. Cornea. 2021; in press.
3. Kuo MT, Hsu BWY, Yin YK, et al. A deep learning approach in diagnosing fungal keratitis based on corneal photographs. Sci Rep. 2020;10(1):14424. [PubMed: 32879364]
4. Dalmon C, Porco TC, Lietman TM, et al. The clinical differentiation of bacterial and fungal keratitis: a photographic survey. Invest Ophthalmol Vis Sci. 2012;53(4):1787–1791. [PubMed: 22395880]
5. Dahlgren MA, Lingappan A, Wilhelmus KR. The clinical diagnosis of microbial keratitis. Am J Ophthalmol. 2007;143(6):940–944. [PubMed: 17408586]
6. Srinivasan M, Mascarenhas J, Rajaraman R, et al. The steroids for corneal ulcers trial (SCUT): Secondary 12-month clinical outcomes of a randomized controlled trial. Am J Ophthalmol. 2014;157(2):327–333. [PubMed: 24315294]
7. Prajna NV, Krishnan T, Mascarenhas J, et al. The Mycotic Ulcer Treatment Trial: a randomized trial comparing natamycin vs voriconazole. JAMA Ophthalmol. 2013;131(4):422–429. [PubMed: 23710492]

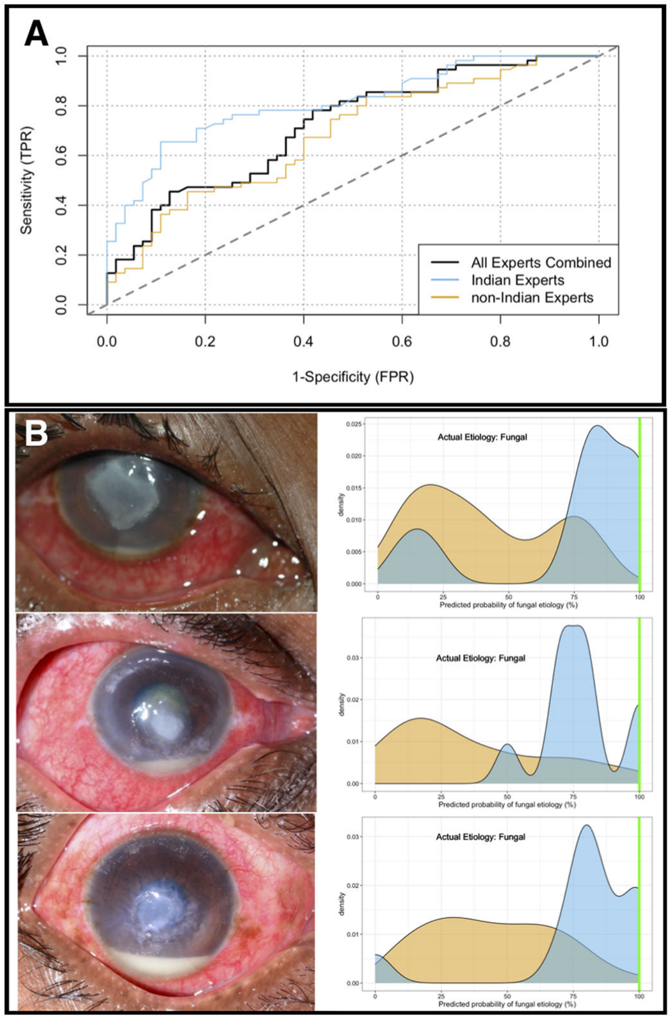

Figure 1: Comparison of Performance Among Indian and non-Indian Cornea Specialists

A) Empirical receiver operating characteristic (ROC) curves for the ensemble estimated probability of fungal keratitis among all cornea specialists combined (black), cornea specialists practicing in India (blue), and cornea specialists practicing outside India (orange). TPR = true positive rate. FPR = false positive rate. B) The three corneal ulcer images from the testing set which demonstrated the the largest difference in ensemble estimated probability between Indian (blue) and non-Indian (orange) expert image graders. Next to each image is the corresponding distribution of responses among both groups of experts. In all three examples the image was obtained from a case of culture-proven fungal keratitis, and in each case the ensemble prediction of Indian graders was closer to the ground truth than the ensemble prediction among non-Indian graders.