NAPCRG 52nd Annual Meeting — Abstracts of Completed Research 2024.

Submission Id: 6186

Title

Crowdsourcing Dermatology Images with Google Search Ads: Creating a Real-World Skin Condition Dataset for AI Development

Priority 1 (Research Category) Research Capacity Building

Presenters

Yejin Jeong, BA, Mike Schaekermann, Steven Lin, MD

Abstract

Context:

Traditional clinical health datasets often fail to represent the diversity and breadth of real-world disease, influencing research and AI tools developed on those datasets. Dermatology, easily documentable through patient-contributed images, presents an ideal case for innovative dataset creation.

Objective:

To develop a scalable, representative dermatology dataset using crowdsourcing via internet search ads, enhancing medical education and AI application development.

Study Design and Analysis:

We employed Google Search ads to gather dermatology images from the public, with informed consent, between March and November 2023. Images were curated, de-identified, and labeled by dermatologists, providing a dataset of 10,408 images from 5,033 contributors across the United States.

Setting or Dataset:

The dataset aggregated includes diverse skin conditions and demographic data, now publicly accessible on GitHub.

Population Studied:

Internet users across the United States, representing various demographics and skin types, contributed to the dataset.

Intervention/Instrument:

Google Search ads targeted individuals searching for skin-related terms, inviting them to contribute images and related information via a web platform.

Outcome Measures:

Measures included the volume of contributions, demographic diversity, and the diagnostic usability of images as evaluated by dermatologists.

Results:

The study achieved a median of 22 submissions per day, with a significant representation of diverse skin conditions and demographics. Over 97.5% of contributions were usable, with high dermatologist confidence correlated with the completeness of accompanying data.

Conclusions:

Crowdsourcing via search ads effectively generates diverse, representative dermatological datasets. The SCIN dataset can significantly enhance dermatological research, education, and AI tool accuracy, particularly in underrepresented communities.

Downloaded from the Annals of Family Medicine website at www.AnnFamMed.org.Copyright © 2024 Annals of Family Medicine, Inc. For the private, noncommercial use of one individual user of the Web site. All other rights reserved. Contact copyrights@aafp.org for copyright questions and/or permission requests.