Received: 21 October 2023 | Revised: 13 May 2024 | Accepted: 23 May 2024

DOI: 10.1002/jvc2.477

JEADV CLINICAL PRACTICE
OPEN ACCESS JOURNAL OF THE EUROPEAN ACADEMY OF DERMATOLOGY & VENEREOLOGY

ORIGINAL ARTICLE

An insight into racial bias in dermoscopy repositories: A HAM10000 data set analysis

Andres Morales-Forero[1] | **Lili Rueda Jaime**[2] | **Sebastian Ramiro Gil-Quiñones**[2,3] | **Marlon Y. Barrera Montañez**[3] | **Samuel Bassetto**[1] | **Eric Coatanea**[4]

[1]Mathematics and Industrial Engineering (MAGI) Department, Université de Montréal (Polytechnique), Montreal, Canada

[2]Infectious and Clinical Dermatology Research Group, Universidad El Bosque, Bogotá, Colombia

[3]Dermatology Resident, Universidad El Bosque, Bogotá, Colombia

[4]Faculty of Engineering and Natural Sciences (ENS), Tampere University, Tampere, Finland

Correspondence
Sebastian Ramiro Gil-Quiñones, Infectious and Clinical Dermatology Research Group, Universidad El Bosque. Ak. 9 #131a-20, Bogotá, Colombia.
Email: sgil9602@gmail.com

Funding information
None

Abstract

Background: Studies have revealed a lack of representation of skin of colour patients in academic sources of dermatologic diseases, including databases. This visual racism has consequently generated less comfort and confidence among the specialists in the care and attention of this ethnic group, including the opportunity of being correctly diagnosed.

Objectives: To investigate and uncover potential racial biases in the HAM10000 data set through an exploratory analysis of the dark skin tones representation, the identification of inaccuracies in its documentation, the recognition of relevant skin conditions absent for darker skin and the lack of ethnic diversity variables crucial for validating diagnosis across different skin tones.

Methods: An exploratory examination was conducted to investigate the occurrence of dark skin within the HAM10000 database (housed in a Harvard Dataverse repository), consisting of 10,015 dermoscopic images of skin lesions. A visual depiction encompassing the whole skin tones was generated by sampling four crucial data points from each image and applying the Gray World Algorithm for colour normalization. To confirm the accuracy of the graphical representation, dermatologists validated the pixel sampling process by analysing a randomly selected 10% of the images for each type of skin lesion. This visual representation was produced for the entire data set as well as for each skin lesion type. The study was further enhanced by comparing the skin lesion representation within the HAM10000 data set against documented prevalences of relevant conditions affecting dark skin.

Results: Less than 5% of the images came from dark-skinned patients. Nevertheless, in about 4.9% of cases, our pixel sampling method might inadvertently capture shadows or dark spots resulting from the imaging device or the lesion itself rather than the individual's actual skin tone. In addition, there are inaccuracies in the data set's claims of diversity and comprehensive

coverage, notably the underrepresentation of conditions prevalent in darker skin and the absence of ethnic diversity variables.

Conclusions: Visual racism is an issue that needs to be addressed in medical sources of information and education. Image databases and artificial intelligence models need to be nourished with information, including all skin types, to guarantee equal access to opportunities. Furthermore, any instances where conditions affecting people of colour are underrepresented must be meticulously documented and reported to highlight and address these disparities effectively. This is particularly important in dermoscopy imaging, where solely relying on image-based racial bias analysis is limited. The alteration of the patient's actual skin tone by the dermatoscope's lighting complicates the accurate assessment of racial bias.

KEYWORDS

dermatology data sets, diagnostic equity, fairness-aware machine learning, inclusive artificial intelligence, racial bias

INTRODUCTION

In the medical area of dermatology, studies reveal suboptimal teaching and representation of skin diseases in the skin of colour to professionals in training, not only due to the lack of specialists and residents belonging to ethnic minorities but also due to 'visual racism'. This may be present in the learning material[1] alongside the lack of inclusion of chairs within academic curricula that allow academic enrichment in dermatological diseases, specifically in the afro-descendant population. This has consequently generated less comfort and confidence within the clinician in the care and attention of this ethnic group.[2]

The Harvard Dataverse repository serves as a valuable data source for research spanning a range of disciplines, including social sciences, arts and humanities, medicine and law.[3] Researchers can utilize this repository to upload, publish and receive proper recognition for their data. Moreover, it provides an opportunity for Artificial Intelligence (AI) developers to test their algorithms and utilize these images as a benchmark data set. While the repository undoubtedly benefits the scientific community, there have been concerns raised regarding the quality of certain published data. One such data set, named HAM10000,[4,5] contains a vast collection of dermatoscopic images of common pigmented skin lesions. Dermoscopy, an essential tool in dermatology, enhances diagnostic precision by providing detailed images of the skin's primary structure and its lesions. It has gained widespread use in screening for skin cancer and pigmented lesions.[6]

However, it is important to note that the documentation accompanying the HAM10000 data set claims that the images are sourced from diverse populations and encompass 'all important diagnostic categories in the realm of pigmented lesions'.[4,5] This assertion may lead researchers to assume that the data adequately represents populations, including Black or female individuals and their corresponding skin lesions, thus enabling the generalization of research findings. Nonetheless, a closer examination reveals a concerning scarcity or absence of images depicting lesions in Black-skinned patients, indicating a potential limitation.

MATERIALS AND METHODS

The HAM10000 data set contains 10,015 images of seven types of pigmented skin lesions: Actinic keratoses and intraepithelial carcinoma/Bowen's disease (akiec), basal cell carcinoma (bcc), benign keratosis-like lesions (solar lentigines/seborrheic keratoses and lichen-planus like keratoses, bkl), dermatofibroma (df), melanoma (mel), melanocytic nevi (nv) and vascular lesions (angiomas, angiokeratomas, pyogenic granulomas and haemorrhage, vasc). This data set is divided into two sets: training and test set. Both sets were analysed.

In numerous instances, the lesions within the images appear dark and are positioned at the centre, while some images exhibit black areas at the edges, attributable to the capture system rather than the skin colour. Therefore, we sampled pixels from four distinct corners of each image to accurately identify the patient's original skin, as shown in Figure 1. Specifically, pixels were displayed within pixels located between 8% and 12% of the length and width of each image (the four squares on each image).

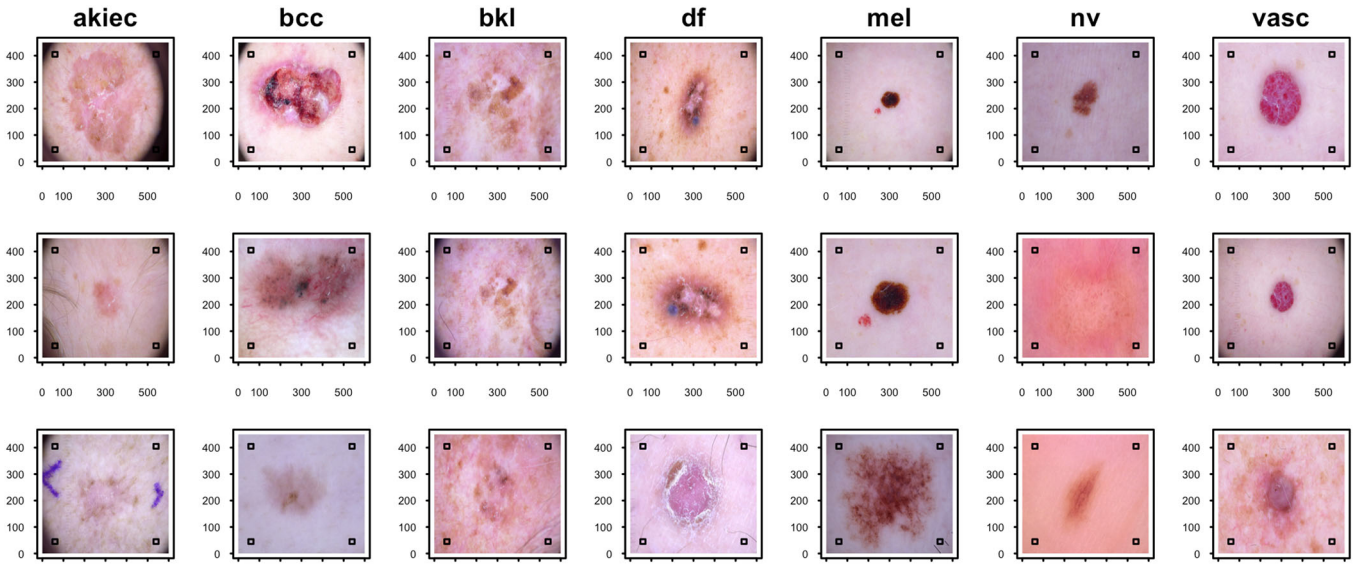

FIGURE 1 Visual overview of corner pixel sampling (the four squares at each corner) in the HAM10000 by lesion type (columns).

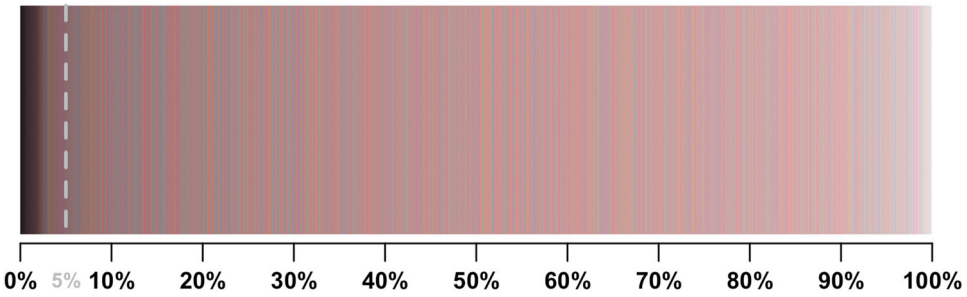

FIGURE 2 Proportional comparison of dark versus light skin images using the 10,015 pictures in the HAM10000 database.

This approach enabled us to avoid sampling parts of the lesion or dark regions of the image. However, some samples were expected to exclusively consist of dark pixels, as shown in the upper left image in Figure 1. We mitigate this by computing the median RGB representation of the four samples within each image, helping to balance out anomalies caused by shadows or variations in lighting. Further enhancing our methodology, we adjust the colour balance of the median RGB representation through the Gray World Algorithm (GWA). The GWA application helps neutralize colour biases arising from differences in white balance and colour calibration, ensuring a more consistent colour representation across the data set. We limit calibration factors to a tight range of 0.95–1.05 for subtle and controlled adjustments. Postcorrection, we sequence the adjusted median RGB values from the darkest to the lightest and visually represent this data on a graph, e, as depicted in Figure 2. Five dermatologists conducted a validation of the pixel sampling method using a randomly selected 10% of the images by the skin lesion type. The resulting graph provides an insight into the prevalence of darker versus lighter skin tones.

Furthermore, to analyse skin tone variations, we implement this method across each of these conditions. We also assess the skin lesions targeted by HAM10000 against those commonly found in dark-skinned populations, comparing their representation and prevalence. We reference existing literature to support the importance of incorporating certain lesions typical of dark skin into benchmark databases such as HAM10000.

RESULTS

The results show the proportion of fair- and dark-skinned individuals in the pictures sampled (see Figure 2). Less than 5% of the graph represents high phototypes (dark skin tones), as more than 90% exhibit fair-skinned phototypes.

By scrutinizing the distribution of various skin lesion types in the HAM10000 data set and their prevalence in

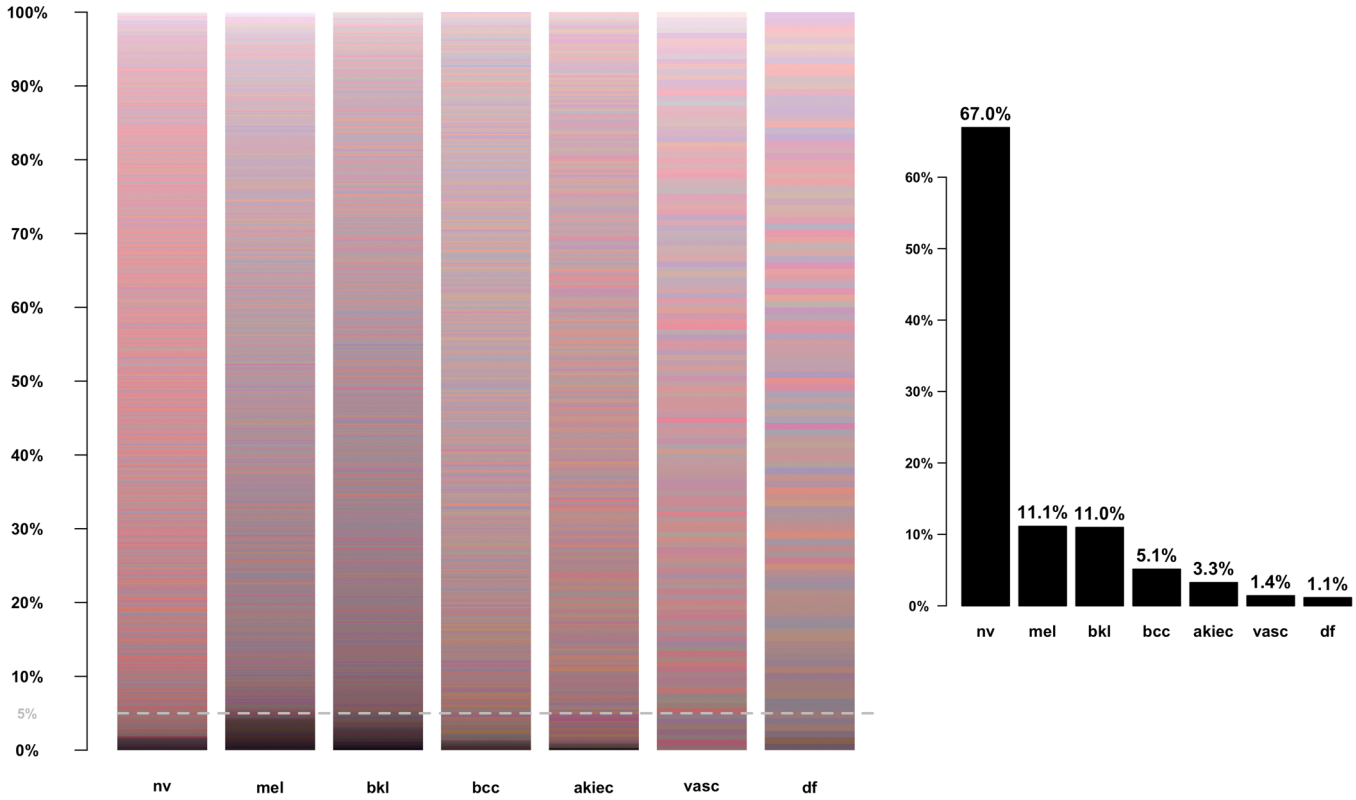

FIGURE 3 (Left side) Proportional comparison of dark versus light skin images across lesion type. (Right side) Proportion of images by lesion type.

darker skin tones, the data set predominantly features lesions commonly found and studied in lighter skin tones (see Figure 3). When aggregating nevi, melanoma and BCC, these conditions account for 83.2% of the HAM10000 data set, suggesting a bias towards conditions more frequently observed in lighter skin tones. In contrast, squamous cell carcinoma (SCC), the most common cancer in Black populations,[7] and dermatosis papulosa nigra (DPN), more prevalent or manifest distinctly in darker skin,[8] are notably absent.

Despite nevi being the most frequently documented condition in the HAM10000 data set (67%), the data set might inadequately reflect the wide range of appearances that nevi can exhibit on darker skin. The distribution of nevi colours, ranging from dark to light, closely reflects the overall composition of the whole data (see Figures 2 and 3), showing a similar proportion of dark pixels.

Conversely, melanoma, accounting for 11.1% of HAM10000, is the data set's second most common lesion type and shows the highest concentration of dark pixels. However, the data set overlooks acral lentiginous melanoma. This specific type of melanoma, more prevalent in individuals with darker skin,[9] is not adequately distinguished, which complicates the analysis of results by skin tone. This oversight is even more critical given the higher mortality rates associated with acral lentiginous melanoma in people of colour, primarily due to delayed diagnoses.[9]

While BCC is less common in darker-skinned individuals, it remains the second most common skin cancer among Black populations.[10] The significance of BCC in these individuals should not be overlooked. However, BCC is represented in just 5.1% of the data set, suggesting a potential shortfall in representation that may impair the ability of algorithmic diagnostics to accurately detect BCC in individuals with darker skin.

DISCUSSION

This research highlights an imbalance in the HAM10000 data set's representation of skin lesions, particularly noting a lack of diversity in images of individuals with darker skin tones. Compounded by the complete lack of metadata or any variables that would permit a detailed analysis of dermatological manifestations across various ethnic groups, this disparity poses a considerable barrier to comprehensively understanding and addressing the dermatological needs of a racially diverse population. It is known that skin cancer incidence is certainly lower in

the dark skin patients compared to White skin patients (0.9 vs. 22 per 100.000) which could be understood as a limitation.[11–13] However, the HAM10000 data set demonstrates a bias towards skin conditions more commonly associated with lighter skin tones. This skew overlooks significant conditions like SCC and DPN, which are either most prevalent in or uniquely manifest in Black populations.

The data set's failure to adequately represent the diversity of nevi presentations in darker skin, alongside the lack of detailed categorization for skin tones, undermines its diagnostic utility for these populations. Diagnostic criteria and warning signs can vary, as certain types or changes in nevi that are more concerning in White skin may not appear the same or pose the same level or risk in Black skin. Variations in colour, size and borders of nevi might not be adequately represented in the data set. As there is no variable that allows accountability between types of skin tones, these potential biases will remain unknown, which could hinder diagnosis and misguide clinicians.

The likely underrepresentation of Black skin in nevi cases suggests that the data set might not adequately represent the diagnostic complexities and unique appearance patterns of nevi in Black skin. Without explicit data categorizing skin tones, the data set lacks the depth needed to address these disparities effectively. Additionally, BCC in darker skin often manifests in areas not typically exposed to the sun, differing from the more common presentation in lighter skin tones. This variation in manifestation may make diagnosis more challenging, as the traditional indicators of BCC, like pearly nodules with distinct edges and telangiectasias, might not be as apparent on darker skin.[10] Furthermore, if the data set lacks diverse examples of this condition, algorithms trained on it might not perform well across all skin tones, potentially causing misdiagnosis or healthcare disparities.

The absence of acral lentiginous melanoma further exacerbates healthcare disparities. The lack of proper differentiation highlights a significant shortfall in the data set's approach to skin cancer diagnosis, contributing to ongoing disparities and discriminatory practices in healthcare. These gaps highlight the critical need for a more inclusive approach in dermatological research and diagnostics to ensure equitable healthcare outcomes across all skin tones.

The lack of ethnic metadata limits the ability of researchers to identify and address specific healthcare needs within various racial and ethnic groups. It also raises concerns about the equitable distribution of research attention and resources. The medical and machine learning community must prioritize the collection of such data moving forward, ensuring that all individuals are represented and that the research derived from such data sets truly advances the cause of global health equity. These representations are not merely mirroring population prevalence, but rather on ensuring the sample size is sufficient to diagnose adequately across ethnicities.

The implications of this bias extend beyond academic interest and enter the field of clinical significance. They manifest in the risk of perpetuating diagnostic inaccuracies, as healthcare professionals may be inadequately trained to recognize conditions as they present in darker-skinned individuals. The breadth and depth of the HAM10000 data set must be expanded to provide an equitable resource that serves the dermatological needs of a diverse global population. This expansion is not solely a quantitative increase in images of darker-skinned individuals but also a qualitative improvement that ensures a wide range of skin conditions are represented according to their unique clinical presentation in all skin tones.

The results in this article were potentially expected; many of these imbalances could have been foreseen before the images were collected. The absence of a more multiracial perspective or education for healthcare specialists involved in the data collection process, or the omission of healthcare specialists in the conception of algorithms based on data sets like HAM10000, indicates a siloed approach that neglects the diverse clinical realities faced by different ethnic groups. We consider that a multidisciplinary collaboration can foster an environment where diverse perspectives are valued and integrated into the development of medical databases and tools. For instance, it is documented that doctors of colour are more likely to serve minorities as well as provide care to underserved populations; therefore, increasing their representation within dermatology can reduce gaps and quality of care in general.[14–16]

In the wake of this study's findings, it is evident that the creation of medical data sets like HAM10000 requires not just an expansion in data quantity but a strategic overhaul in data curation practices. Future efforts must prioritize the inclusion of a broader spectrum of skin conditions, particularly those prevalent among darker-skinned populations, and ensure that such conditions are accurately represented in terms of severity, morphology and progression.

The results also go beyond the cautious use of HAM10000 or improvement of ML practices; data-based studies like this one can just reflect discriminations already presented in society. The HAM10000 analysis in this article contributes to unmasking the already reported problematic in dermatology research and education. Visual bias or visual racism has already been documented as an issue in the area of dermatology; the term has been

addressed in scientific literature as the underrepresentation of skin colour in educational material and in noneducational resources.[17] Perlman et al. studied the representation of images of skin of colour among important dermatology training materials in Medical School, such as the AAD curriculum, USMLE First Aid, Pathoma and so forth. They found that only 14.9% of 1123 images corresponded to skin of colour.[18] This fact is of paramount relevance since, as it is expected, the lack of inclusion of this type of patients in medical literature and learning material could affect dermatological practice, diagnostic accuracy and reliability in the context of caring for this group of patients.[19] Although there are genetic and phenotypic differences in human morphology (including skin colour), it is a fact that underrepresentation probably stems from a foundation of socioeconomic inequality and structural barriers that limit inclusivity in dermatology.[14]

Recognizing and addressing the existence of racial bias is the starting point towards a justified change that encompasses the paradigms of inclusion, diversity and representation, especially focused on the afro-descendant population so that it is possible to generate open spaces for discussions regarding restructuring of dermatology academic curricula. This could potentially be achieved by introducing elements that guarantee adequate clinical and social preparation in the context of medical care for patients belonging to ethnic and/or racial minorities.

The HAM10000 data set aimed to tackle 'the small size and lack of diversity of available data sets of dermatoscopic images' to train neural networks. However, even if these images are only used for academic purposes, the inclusion of Black minorities in studies is imperative, and it cannot be ignored. Moreover, if these populations are excluded, this fact should at least be mentioned in the database description. These clarifications are missing from this database's documentation. Despite previous studies highlighting similar disparities,[20] HAM10000's documentation does not address these concerns.

Our study further highlights the essential role of AI and regulatory organizations in devising and deploying technological healthcare solutions. It stresses the importance of ensuring these innovations are developed and applied with a keen awareness of their ethical implications, particularly in maintaining fairness and preventing bias in medical diagnosis. Accountability and transparency in their data and algorithms are imperative. This involves rigorous evaluation and validation processes that are open and accessible, allowing for the identification and correction of any biases that may exist, thereby promoting fairness and inclusivity in AI-driven dermatological diagnosis.

LIMITATIONS

The study employs median RGB analysis from image corners to mitigate the effect of dark pixels due to lesions or capture devices, alongside the GWA to correct for photographic discrepancies such as white balance and colour calibration, aiming to improve skin colour representation accuracy. However, randomly inspecting the 10% images, we estimate that around 4.9% of the time, our sampling method could mistakenly identify shadows or dark areas created by the dermatoscope or the lesion itself as skin tone, not accurately capturing the person's true colour (see Annex). Since lesioned or shadowed skin is often darker than the actual skin, there may be an overestimation of the number of images featuring dark skin. Nevertheless, despite our efforts to precisely sample healthy skin, our methodology—as any method relying solely on the HAM10000 data set—might not capture the full range of skin tones, especially in darker-skinned individuals. In these cases, lesions are often found in areas with less sun exposure, showing lighter pigmentation than the general skin tone. Additionally, the illumination from dermoscopes can make the skin colour seem lighter than it actually is. This highlights the limitation of using the HAM10000 data set alone for inferring patient ethnicity, emphasizing the importance of incorporating detailed demographic variables like ethnicity into dermatological data sets for more comprehensive analysis.

CONCLUSION

Visual racism is an issue that needs to be addressed in medical sources of information and education. Image databases and AI models need to be nourished with information, including all skin types, to guarantee equal access to opportunities. Incorporating exploratory data analysis into the AI developer practice manual is of utmost importance; this underrepresentation could have been identified since the HAM10000's inception in 2018. This article uses a straightforward and interpretable approach to dermoscopic image exploration. However, numerous techniques and methods are available to explore image data comprehensively. Neglecting this type of analysis can have far-reaching consequences, ranging from impeding the effective implementation of machine learning algorithms and perpetuating unfairness to posing risks to human lives. As the field advances, proactive efforts are essential to address and rectify these biases for the ethical and equitable evolution of AI technologies in dermatology.

AUTHOR CONTRIBUTIONS

Conceptualization, methodology and original draft preparation: Andres Morales-Forero and Lili Rueda Jaime. *Coding and computational analysis*: Andres Morales-Forero. *Methodology and results validation*: Lili Rueda Jaime, Sebastian Ramiro Gil-Quiñones and Marlon Y. Barrera Montañez. *Review and editing*: Andres Morales-Forero, Lili Rueda Jaime, Sebastian Ramiro Gil-Quiñones and Marlon Y. Barrera Montañez. *Supervision*: Samuel Bassetto and Eric Coatanea. This study was a multidisciplinary effort, reflecting a comprehensive approach to examining racial bias in dermoscopy repositories. All authors have read and agreed to the published version of the manuscript.

ACKNOWLEDGEMENTS

This article was funded by the authors themselves.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available in DermDataFairness at https://github.com/jamorafo/DermDataFairness. These data were derived from the following resources available in the public domain: Tschandl, Philipp. (2018). The HAM10000 data set, a large collection of multisource dermatoscopic images of common pigmented skin lesions. Harvard Dataverse, V4. https://doi.org/10.7910/DVN/DBW86T.

ETHICS STATEMENT

Ethical Approval: institutional ethics board approval (University of Queensland, Protocol-No. 2017001223). The HAM10000 data set used in this study provided figures that are anonymous aggregated data and are openly accessible to be used for academic purposes. We did not include human subjects; we just used unidentifiable data from the data set. We declare that the ethical concerns do not apply to this study.

ORCID

Sebastian Ramiro Gil-Quiñones https://orcid.org/0000-0002-5974-1431

Marlon Y. Barrera Montañez https://orcid.org/0000-0001-6483-5224

REFERENCES

1. Kamath P, Sundaram N, Morillo-Hernandez C, Barry F, James AJ. Visual racism in internet searches and dermatology textbooks. J Am Acad Dermatol. 2021;85(5):1348–9.
2. Lester JC, Taylor SC, Chren MM. Under-representation of skin of colour in dermatology images: not just an educational issue. Br J Dermatol. 2019;180(6):1521–2.
3. The President & Fellows of Harvard College. HARVARD Dataverse [Internet]. 2022 [cited 2022 Mar 6]. Available from: https://dataverse.harvard.edu/
4. Tschandl P. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions [Internet]. 2018. V3 ed. Group V, editor. Harvard Dataverse. Available from https://doi.org/10.7910/DVN/DBW86T
5. Tschandl P, Rosendahl C, Kittler H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Sci Data. 2018;5:180161.
6. Saini AD, Prodanovic E. Standard dermoscopy and videodermoscopy as tools for medical school education. Dermatol Pract Concept. 2019;9(3):214.
7. Kundu RV, Patterson S. Dermatologic conditions in skin of color: part II. Disorders occurring predominately in skin of color. Am Fam Physician. 2013;87(12):859–65.
8. Gloster Jr. HM, Neal K. Skin cancer in skin of color. J Am Acad Dermatol. 2006;55(5):741–60.
9. Mpofana N, Chibi B, Gqaleni N, Hussein A, Finlayson AJ, Kgarosi K, et al. Melasma in people with darker skin types: a scoping review protocol on prevalence, treatment options for melasma and impact on quality of life. Syst Rev. 2023;12(1):139.
10. Kundu RV, Patterson S. Dermatologic conditions in skin of color: part I. Special considerations for common skin disorders. Am Fam Physician. 2013;87(12):850–6.
11. Alvarado SM, Feng H. Representation of dark skin images of common dermatologic conditions in educational resources: a cross-sectional analysis. J Am Acad Dermatol. 2021;84(5):1427–31.
12. Henderson MD, Abboud J, Cogan CM, Poisson LM, Eide MJ, Shwayder TA, et al. Skin-of-color epidemiology: a report of the most common skin conditions by race. Pediatr Dermatol. 2012;29(5):584–9.
13. Alexis AF, Sergay AB, Taylor SC. Common dermatologic disorders in skin of color: a comparative practice survey. Cutis. 2007;80(5):387–94.
14. Akhiyat S, Cardwell L, Sokumbi O. Why dermatology is the second least diverse specialty in medicine: how did we get here? Clin Dermatol. 2020;38(3):310–5.
15. Cooper LA, Beach MC, Johnson RL, Inui TS. Delving below the surface: understanding how race and ethnicity influence relationships in health care. J Gen Intern Med. 2006;21(Suppl 1):S21–7.
16. Vasquez R, Jeong H, Florez-Pollack S, Rubinos LH, Lee SC, Pandya AG. What are the barriers faced by under-represented minorities applying to dermatology? A qualitative cross-sectional study of applicants applying to a large dermatology residency program. J Am Acad Dermatol. 2020;83(6):1770–3.
17. Davis SA, Narahari S, Feldman SR, Huang W, Pichardo-Geisinger RO, McMichael AJ. Top dermatologic conditions in patients of color: an analysis of nationally representative data. J Drugs Dermatol. 2012;11(4):466–73.
18. Perlman KL, Williams NM, Egbeto IA, Gao DX, Siddiquee N, Park JH. Skin of color lacks representation in medical student resources: a cross-sectional study. Int J Womens Dermatol. 2021;7(2):195–6.
19. Adelekun A, Onyekaba G, Lipoff JB. Skin color in dermatology textbooks: an updated evaluation and analysis. J Am Acad Dermatol. 2021;84(1):194–6.

20. Adamson AS, Smith A. Machine learning and health care disparities in dermatology. JAMA Dermatol. 2018;154(11): 1247–8. https://doi.org/10.1001/jamadermatol.2018.2348.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Morales-Forero A, Rueda Jaime L, Gil-Quiñones SR, Barrera Montañez MY, Bassetto S, Coatanea E. An insight into racial bias in dermoscopy repositories: a HAM10000 data set analysis. JEADV Clin Pract. 2024;1–8. https://doi.org/10.1002/jvc2.477