© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1719647

Transforming Skin Cancer Triage in Underserved U.S. Communities: An In-Depth Examination of Advanced Machine Learning Models for Early Detection and Dermatology Referral Completion

SHIMA ALI SADIA[1], EUSHA MOHTASIM[2]
[1]*Master of Public Health in Biostatistics and Epidemiology, Monroe University, United States*
[2] *Computer Science and Engineering, University of Liberal Arts Bangladesh, Bangladesh*

Abstract- Background: Although skin cancers are the most prevalent cancer diagnosed in the United States, racial and ethnic minority individuals and people residing in rural and other underserved communities are overrepresented with skin cancers diagnosed late, lack timely access to referral to experts, and are more likely to die from skin cancer. While effective in controlled settings machine learning (ML) based diagnostic tools have limited impact on a particular population that would benefit from increased capacity to be triaged due to the datasets clinical biases, geographic barriers, and gaps in referral initiation and completion.
Goals: Review and critique the evidence of the use of advanced machine learning models to detect early skin cancer; compile a list of specific barriers that prevent such models from making an equitable impact in underserviced communities in the U.S.; and recommend a holistic triage system for ML-based detection of early skin cancer, teledermatology and Methods: The 20 peer-reviewed literature sources that were reviewed and used for this narrative review included foundational studies on machine learning validation for dermatology, studies examining algorithmic bias and a diversity of skin datasets, dermatology workforce and access epidemiology studies, teledermatology implementation science, and dermatology referral completion outcomes research. Patients were sorted in categories based on the different stages of the lesion detection to confirmed dermatologic evaluation process; focus was maintained on the point where patients were dropped from the process.
The results: include new deep learning systems which are promising to match the performance of board-certified dermatologists for diagnosis on curated sets of images and the first primary-care skin cancer system to use AI is now authoritatively approved by regulators in the United States. (Venkatesh et al., 2024) However, dermatology AI models that have been trained to predominantly and almost exclusively classify lighter skin tone images are significantly compromised when was applied to models from darker skin tones and less common presentations. A critical and increasing geographic maldistribution of the dermatology workforce (Feng et al., 2018) coupled with an incomplete triage of patients often fail to complete referral to full definitive in-person evaluations, even after it is successful (Duniphin, 2023; Pagani et al., 2023), especially when crossing geographical boundaries or language/racial barriers.
Perspectives: While advanced machine learning models are a potentially promising resource for enhancing the capacity to detect early skin cancer in underserved communities in the U.S. by no means can they be assumed to be similarly distributed. To unlock their potential, however, calls for intentional, concurrent investments in a variety of diverse and representative training data, auditing biases, and structured support to complete referrals, coupled with the need for a coordinated triage pathway, all done and delivered as a whole and not in isolation as technical interventions.

I. INTRODUCTION

1.1 Burden of Skin Cancer in the United States and the Diagnostic Delay in the Country

Skin cancer (both melanoma and the more common, but non-lethal, non-melanoma skin cancers) is the most frequently diagnosed form of cancer in the United States. Despite an overall increase in incidence, the rate of disease progression is not the same for everyone: the survival rate is far better for those who have been diagnosed with the disease at an early stage, and the disease is still contained locally, than for those at a late stage of the disease, after the cancer has spread locally or regionally. Every part of the way up and down this slope of an outcome gradient, from the moment an individual first notices a suspicious lesion, through an initial clinical evaluation and finally a specialist evaluation and, if applicable, biopsy, is examined here.

A large body of evidence shows that the range of this diagnostic interval is not worked through evenly across racial, ethnic and geographic lines. In the context of large national cancer registry analyses that cover more than a decade of years (Qian et al., 2021) and over a decade of years (Mahendraraj et al., 2017), as well as systematic review and meta-analysis of the melanoma literature (Lam et al., 2022), the same pattern of later stage at diagnosis, and poor survival, were also observed in non-Hispanic Black patients. This inequity exists among a much greater range of cancer health disparities experienced by people of racial and ethnic minority groups across almost all primary types of cancer (Zavala et al., 2021).

1.2 Lack of Economic Opportunity has been a Major Challenge in the City's Region Since Its Founding

One key source of the inequities of diagnostic delay is the distribution of the dermatology workforce across the country. Using estimated dermatologist density across the nation over the past 18+ years, the data did not indicate a trend of increased density from rural and non-rural areas over time, instead the disparity between these areas increased over time, with higher density in areas having more resources and in urban environments (Feng et al., 2018). Many of those most likely to benefit from early diagnosis are the patients who have the hardest of access to the specialists who can make a definitive diagnosis, as this maldistribution of both patients and specialists is occurring in many suburban and historically underserved urban areas.

Two groups of complementary innovation have been spurred by this access challenge and are the focus of this review. The first is technological: using sophisticated machine learning algorithms to provide a fast initial screening to determine the likelihood of cancer for a skin lesion, by a primary care provider, community health worker or even the patient directly that generally have fewer access to board certified dermatologists. The second is organizational; Teledermatology technology, which connects patients and/or primary care providers in underserved areas with teledermatologic expertise reducing the need to travel to a specialist's office to receive their services (Duniphin, 2023, Parga et al. 2025).

1.3 This Section Shares Some of The Justifications For Using Machine Learning For Triage

The machine learning-based skin cancer triage tools are appealing, just because they tackle the specific problem mentioned in the previous point lack of specialist capacity for the skin cancer diagnosis in comparison with patient demand. Once a system based on deep learning could detect a concerning from a reassuring lesion on just a photo, if it was certain enough, it can then be used anywhere where the smartphone camera and internet connection are available with no consideration of how many dermatologists reside in the area. But initial test results were impressive: a Deep Convolutional Neural Network (DCNN) based classifier trained on 129,450 clinical images and compared with 21 board certified dermatologists performed equally well when trained and tested on both keratinocyte carcinoma and melanoma images (Esteva et al. 2017). This discovery spurred a significant follow-on research program, generalising the classification task to a critical subsequent one in the context of primary care comprehensive differential diagnosis for which the team has continued to research (Liu et al., 2020) and ultimately to regulatory authorisation of AI-powered devices that are meant to be used directly by non-specialist clinicians (Venkatesh et al., 2024).

But measuring this promise is not solely based on purely technical performance measures. An accurate model on a population that isn't highly representative of the population at question is of little help to them, something this review documents; in the case of melanoma, those at risk of poor outcomes (Daneshjou et al., 2021, 2022; Fliorent et al., 2024) tend to be the people with darker skin tones and less common presentations who are underrepresented in melanoma training data. The prospects of machine learning for underserved communities thus come with the added challenge of looking at the ability for machine learning to work for some communities, but not others, and what happens if a machine learning model does detect a lesion it notes as a critical one to be referred?

1.4 The Review aims

The aims of this review are (1) to describe the epidemiological pattern leading to disparities in skin cancer outcomes driving the desire to scale-up triage, with a focus on initial detection to teledermatology consultation to the eventual occurrence of a referral completion; (2) to study access barriers, both societal and individual, to dermatologic care within underserved communities; (3) to critically review the evidence of how much a machine learning-based skin cancer detection technique can contribute to this process and to how well it performs across all skin tones; and (4) to stake out the entire triage pathway from detection, via teledermatology consultation to eventual dereferral, so as to identify "hotspots" where losses to follow up are likely to occur in underserved populations and propose an integrated approach to mitigate such losses.

II. METHODS

2.1 Approach and Scope

This is a narrative review from a set of ten English language, peer-reviewed literature sources that were found, by way of targeted searches, in databases related to informatics in the biomedical, disease-specific (oncoscopy) and health policy fields as well as the dermatology field. Sources were chosen because they directly addressed one or more of four thematic areas: (1) epidemiology of skin cancer outcome disparities by race, ethnicity and geography in the U.S.; (2) machine learning and artificial intelligence performance in skin cancer detection, foundational validation studies, benchmark datasets, and consumer-oriented diagnostic tools; (3) algorithmic bias in dermatology artificial intelligence and its relationship to skin tone representation in training datasets; and (4) structural and organizational barriers to access of and for dermatologic care and referral completion, including implementation of teledermatology.

2.2 Identifies and Combines Information from and Synthesizes A Variety of References

Due to the limited scope deliberately chosen for this review, no more than the 20 sources cited above were added, and all contentions made in this article are supported by one or more of the above evidence with in-text citations. Because of the large amount of heterogeneity in the study design, population and definition of outcomes of included sources (such as large national cancer registry studies, single-institution retrospective cohorts, algorithm validation studies, and scoping reviews of health services), findings have been summarised narratively rather than through a formal quantitative meta-analysis. The review is designed around a single focus to identify a journey of ideas, starting with disparities in population-level outcomes, then going through the entry barriers and technological interventions to address these disparities and then to the key point in the care pathway at which disparities remain at the point of referral completion which is where the review argues that technology promises to have a most direct impact.

2.3 Epidemiology and Disparities in Achieving Optimal Skin Cancer Outcomes

There is the appearance of an apparent paradox to melanomas in the United States that often causes confusion non-Hispanic White persons have five to 10 times higher melanoma incidence rates than Black and Hispanic persons or Asian persons and, once diagnosed with melanoma, Black, Hispanic, and Asian persons are always at a disadvantage compared with the non-Hispanic White population with respect to melanoma-specific outcomes. The study, which included a population-based assessment of clinical and histologic features and prognosis in 1,106 African-American patients with melanoma between 1988 and 2011 in the Surveillance, Epidemiology, and End Results (SEER) database, revealed both the unique clinical and histologic characteristics of melanoma among African-Americans and their relative poor outcome (Mahendraraj et al., 2017). A more recent analysis using the SEER data, confirm that this racial outcome disparities trend in mortality statistics for melanoma have persisted into other decades that have followed as overall cancer care has improved, suggesting that these disparities are not a reflection of an incomplete older SEER data set (Qian et al., 2021). This pattern appears to be independent of another large scale SEER cohort from 1992-2009 which found longer duration of survival after melanoma diagnosis for White patients, followed by Hispanic, Asian American/Native American, and Black patients (Dawes et al., 2016), making it a more robust pattern to stand by itself rather than being the result of any one data set.

2.4 The Importance Of Non-Metastatic Stage At Diagnosis and The Difference In Survival Rates

In the melanoma outcomes literature reviewed in Section 3.1, delayed diagnosis (minorities also tend to be diagnosed at regionally advanced or distant stage disease), has been the most consistent explanation given for this survival gap (Dawes et al., 2016; Mahendraraj et al., 2017). Results from a systematic review and meta-analysis, collating a large number of the included studies, showed overall reduction in overall survival and melanoma-specific survival in Black patients compared to non-Hispanic White patients skin cancer, with similarly poor outcomes for the other ethnic minority groups compared to non-Hispanic White patients being reported often (Lam et al., 2022). This is one of the most rigorous and comprehensive evidence-based syntheses of North American data over the past 30 years, with a robust conclusion that the survival difference seen in individual registry studies is a generalizable phenomenon, and not specific to a single health care system or period.

The diagnostic triage pathway is a significant point of intervention, as it is during this stage that interventions are most likely to close the gap on survival as they enter into a process of earlier detection and evaluation (such as the machine learning tools studied in Sections 5.0 and 6.0) and, where successful, achieve good reliability across the entire patient mix that should be able to use them.

3.3 Cancer Health Disparities in a Context Broader than Cancer

It is important to note that the health disparities presented above are not unique to melanoma, but rather are part of the larger cancer health disparities context for racial/ethnic minorities in the U.S. An in-depth review of the health disparities in cancer for several types of cancer and racial/ethnic minority populations revealed that health disparities are due to a complex, multifactorial combination of cancer-specific and non-cancer risk factors that includes genetic, environmental, socio-economic, and healthcare access factors, which extend beyond any individual cancer type (Zavala et al., 2021). The importance of this lies in two aspects: first, the melanomaspecific findings from sections 3.1 and 3.2 can be used within this broader context to show that the interventions proposed later in this review only tackle a single pathway out of several that can contribute to disparities in outcomes; these interventions should not be expected to overcome disparities entirely on their own; and second, the implications of the lessons learned from the disparity-reduction strategies other cancer types employ can inform interventions specifically for skin cancers, and vice versa.

3.4 We Are Committed to Providing Our Patients With Respectful Care

The experiences of cutaneous treatment also vary systematically across racial/ethnic lines in ways that likely account for the above described delayed-presentation pattern.Objective indicators of outcome also vary systematical by racial/ethnic group in ways that are likely to help explain the delayed-presentation pattern found above. Using nationally representative survey data, a cross-sectional study revealed significant differences in the likelihood of a person feeling respected, listened to, and explained, among people with a self-reported skin cancer history between people of racial and ethnic minority backgrounds to non-Hispanic Whites during interactions with the clinical system (Fahmy et al., 2023). This is a clinical finding because perceived disrespect and communication problems are likely to deter patients from timely return for follow-up visits, from voicing "concerns" related to new or changing lesions, and from following through on appropriate referral of new or changing lesions – the clinical "pathway" of this behaviour which was detailed in Section 9.0 of this review. The interpersonal dimension of the care experience, which is essential for the primary dimension of algorithmic/structural barriers and barriers to earlier detection that is the primary focus for this review, necessarily remains uncouched without this dimension of the care experience.

IV. DERMATOLOGY ACCESS CRISIS IN UNDERSERVED COMMUNITIES

4.1 This Reduces Geographic and Workforce Maldistribution

In addition to the more-than-recognizable geographic disparities in the numbers of skin surgeons and dermatologists in the United States that were introduced in Section 1.2, the distribution of the U.S. dermatology workforce is increasingly inequitable over time. Using county-level dermatologist density data, the authors determined that dermatologists were consistently found to be more concentrated in well-resourced, urban counties, and that dermatologist density in metropolitan counties increased even more over the study period than did other counties over the same period (Feng et al., 2018). This puts a disturbing spin on the disparities reported in Section 3.0 because the racial and ethnic minority groups identified as being under-diagnosed for melanoma tend to be overrepresented in rural, and often underserved, areas. So it seems possible that racial and ethnic workforce imbalances are a direct driver of the disparities recently reported for melanoma diagnosis.

4.2 Barriers to Care Beyond the Geography

To measure all available barriers to dermatologic care in underserved communities through geographic distance to the nearest dermatologist is an incomplete measure. A recently published in-depth examination of barriers to specialty care access identified a complex framework of patient income, not having health insurance, and the larger systems that patients rely on for health services, and that the specialty-care access gap in melanoma and non-melanoma skin cancers cannot be solved by simply tackling issues of geographic access (Duniphin, 2023). This dual layer barrier is why single barrier interventions are not likely to achieve universal access: An urban patient living near specialists, who may have social, cost, insurance or language issues due to overcrowding, and a rural patient living near a dermatologist, who may have inaccessible cost, insurance, transportation or language issues due to his or her lack of access.

4.3 This chapter provides a summary of key barriers to access to dermatologic care

Table 1 provides a summary of the main categories of barriers to dermatologic care access documented in Sections 4.1 and 4.2, the specific evidence supporting each barrier, and type of intervention most likely to address each barrier. This transparency follows in explaining why several, overlapping and complementary interventions are layered together in the integrated framework proposed in Section 11.0: machine learning driven triage, teledermatology, and referral navigation support, instead of relying on an individual, technological or organizational solution.

Table 1: Key Barriers to Dermatologic Care Access in Underserved U.S. Communities

Barrier category	Supporting evidence	Primary intervention type
Geographic/workforce maldistribution	Widening urban-rural dermatologist density gap over time (Feng et al., 2018)	Teledermatology; ML-based primary care triage
Socioeconomic and insurance barriers	Income, insurance status, and care-seeking location identified as core barriers (Duniphin, 2023)	Structured referral navigation; policy/coverage reform
Algorithmic performance gaps	Substantially reduced AI diagnostic accuracy on dark skin tones and uncommon disease (Daneshjou et al., 2022; Fliorent et al., 2024)	Diverse dataset curation; disaggregated bias auditing
Referral non-completion	10.9% of urgent referrals never completed; disparities by language/race (Pagani et al., 2023)	Referral tracking, navigation support, communication improvement
Patient trust/communication	Minority patients less likely to feel respected, listened to, explained to (Fahmy et al., 2023)	Communication training; patient-centered referral counseling

V. FOUNDATIONS OF MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE IN SKIN CANCER DETECTION

One oft-cited study indicates the dawn of modern machine learning-based skin cancer detection occurred in 2017, when a deep convolutional network (DCNN) that was trained end-to-end on skin lesion images and diagnosis could compete classically with board-certified dermatologists (Esteva et al., 2017). The classifier was trained on two orders of magnitude more clinical images (129,450) compared to previous similar studies and validated on a cohort of 21 board certified dermatologists who made a diagnostic decision of a biopsy in cases that were diagnosed as either keratinocyte carcinoma or malignant melanoma using images specifically evaluated based on the clinician's diagnosis. Beyond the important questions of the underlying technical proof of concept regarding the ability to train an algorithm solely on images and labels to replicate the expert-level dermatologic judgment of an expert, this study is of great value in the underserved-communities interpretation of this review.

5.1 Differential Diagnosis Systems for Primary Care

Later studies continued to improve their understanding of the limitations of the classification paradigm by Esteva et al. (2017) and deep learning techniques were used in the development of larger systems for differential diagnosis that could be used directly by non-specialist doctors. The deep learning system was trained on diagnosable cases of skin-related disease, and not designed solely to classify skin disease as cancer, with a sample from a teledermatology system providing a differential diagnosis of the most commonly seen skin diseases in primary care, out of 16,114 de-identified cases from a 17 clinic-based practice (Liu et al., 2020). This much broader context is relevant to this review's underserved-communities context because the primary care clinician will frequently be the primary source of clinical contact for the patient whose skin problem needs assessed, and a tool that helps assess the range of common differential diagnoses that can also help access patient concerns with skin problems has much greater utility than one limited to a cancer diagnoses alone.

Being trained and benchmarked in medical image classification or computer vision problems, in particular for skin cancer detection, large datasets of publicly available annotated or well-annotated images have played a significant role in the success of machine learning algorithms. To overcome the historical lack of training data of dermatoscopic images and its lack of diversity, the HAM10000 dataset ("Human Against Machine with 10000 training images") has been published consisting of 10,015 dermatoscopic images collected from several source populations and has become one of the most popular benchmark datasets in the area (Tschandl et al., 2018). Although, as outlined in Section 6.0, the populations HAM10000 and similarly-built data sets have been gathered from were not actively sought after for diversity of skin tone, the resulting population(s) mirrors a point in history for those populations and therefore directly influences the accountability for diagnostic equity during this time.

5.2 The accuracy of the consumer applications for which the smartphone user will use.

In parallel to the clinician-focused system mentioned above, a new class of „consumer (patient) focused“ smartphone apps has developed, which are directly marketed to patients for self-skincancer risk evaluation without clinicians. The accuracy of these applications was systemically reviewed, revealing large discrepancies in the reported sensitivity and specificity between the nine reviewed studies and six of the applications and also methodological flaws such as lack of a representative population of the general public (Freeman et al., 2020). These applications in theory are well suited for reaching those without easy access to primary care and dermatology-based doctors and as such warrant careful consideration to provide a more comprehensive triage solution, but the current evidence does not allow for confidence in using these consumer-grade applications as a component of a "whole system" triage, underlining the continued value of the more clinically integrated solutions outlined in Sections 5.1, 5.2 and 7.0.

VI. ALGORITHMIC BIAS AND THE SKIN OF COLOR PROBLEM

6.1 Missing and Mismatching Data Sets

Well-documented underrepresentation of darker skin tones in dermatology AI research seriously complicates the prospect of a diagnostic use case as outlined in Section 5.0. AI tools based on image classification often disproportionately rely on constituent data sets with majority fair skin, but lack adequate representation of skin of color populations when compared to those in the larger patient populations (Daneshjou et al., 2021). Put another way, models trained mostly on lighter skin tones can be fairly expected to be not as reliable when applied to darker skin tones – which aligns with the general dermatology AI literature, which also found the narrow scope of traditional approaches to skin typing and its lack of skin-of-color representation to be a fundamental, unsolved problem in the field (Fliorent et al., 2024).

6.2 Diverse Dermographic Images Benchmark

A dedicated research effort was undertaken to develop a new, pathologically verified, expertly curated and publicly available collection of dermatology images, on average showing a diverse skin tone, off which we defined the Diverse Dermatology Images (DDI) dataset (Daneshjou et al., 2022), which was used to compare and benchmark existing 'off-the-shelf' dermatology AI systems that had been similarly validated. This wide-ranging evaluation revealed significant performance bottlenecks, in particular when labelled with difficult-to-categorise, or rare, skin lesion images, where accuracy was diminished; similarly, distant (ground truth) labelling by human dermatologists was found to result in poorer performance across some images of dark skin tone and less frequently encountered disease type. This is a testament to the fact that the bias issue is not something that can be just solved by data engineering or an architecture change to the model the real issue dates back before even the algorithms were written, and is a lack of dermatologic knowledge and pattern-recognition training with skin of color.

6.3 Outcomes from Healthcare Algorithmic

This type of algorithmic bias is not confined to healthcare and has been exposed in other fields like search engines. The algorithmic bias of this type is not only confined to dermatology as it has been revealed in other areas such as search engines. A widely cited study of a widely-used commercial-level algorithm, employed by the vast majority of U.S. healthcare providers, to identify patients for care-management enrollment demonstrated that an important racial bias exists: the algorithm relies on historical healthcare costs as a marker of underlying health need, and spending on Black patients with similar clinical need is actually less a bias that effectively meant that the algorithm resulted in less people being identified for further care among Black patients. (Obermeyer et al., 2019) Avoiding algorithmic bias can be done through a similar process to that described in the dermatology context reviewed above; rather than simply relying on the accuracy of the algorithm to show that there is little bias, more comprehensive performance auditing, as performed by Daneshjou et al. (2022), involves deliberate disaggregation of the algorithm's performance across the specific settings in which it is likely to apply to real patients.

6.4 Emerging Mitigation Strategies

Directly addressing the underrepresentation issue, researchers have started to expand into structured endeavours that seek to boost the diversity in dermatological training data. One protocol outlines four stages: measuring the current level of underrepresentation within publicly accessible image repositories of skin images, creating synthetic images that mimic skin tones and underrepresented, thoroughly assessing the realism and disease presentation relevance of the synthetic images, and combining this wider spectrum of images with existing ones to achieve a more balanced detection model (Rezk et al., 2022). This is a promising early phase in these types of mitigation which, at this stage, are far less developed compared to the well-established ones in Section (5.0), though important as a sign of the increased visibility and formalization in reporting biases in dermatology AI studies, paralleling the calls for bolstering transparency and bias reporting in AI systems across STEM fields (Daneshjou et al., 2021; Fliorent et al., 2024).

VII. REGULATORY PATHWAYS AND REAL-WORLD DEPLOYMENT

7.1 FDA Approval of AI-Used Skin Cancer Devices

The process of bringing the insights from the research outlined in Sections 5.0 and 6.0 into practice has relied on the shifting regulatory environment of medical devices in the U.S. involving AI technologies. Milestone 2: FDA approval of DermaSensor, a new medical device that is the first to receive clearance for use in primary care settings where it will help detect skin cancer and treated by non-specialist physicians, setting a precedent for medical device clearance of AI tools for use by non-specialist clinicians and non-dermatologists (Venkatesh et al., 2024). This is directly applicable to this review's topic, as primary care clinicians are the most likely clinicians to be the first point of care medical contact in the rural and underserved communities reported in Section 4.0, where there are a scarcity of dermatologists.

7.2 National Guidance for Health for Health Systems and Health Plans

This approval also brings up concerns on how these devices are continuously generating evidence and if there is a monitoring system in place that works for all skin tones and the noted performance gap in Section 6.0 (Venkatesh et al., 2024). Implementing a device in use without disaggregated reporting, similar to that shown by Daneshjou et al. (2022) on aggregate diagnostic accuracy, could end up repeating the inequity observed in the real world when deployed in primary care that has been reported in the research literature. To ensure the successful transition of machine learning (ML)-based triage to primary care and to support primary care transitions to the use of ML tools, regulatory evidence standards and validation protocols must explicitly consider skin-tone-disaggregated performance evaluation, as equivalent to not considering equity assessment separately from core evidence standards and validation.

8.1 Use of Teledermatology in Isolation as a Primary

Finally, the machine learning methods from Section 5.0 are complemented by a method of teledermatology that is often referred to as the “store-and-forward” system: in which the clinical information and images are sent to a remote dermatologist for later review. An in-depth barrier analysis revealed that teledermatology was one of the initial applications to be considered for moving patients to earlier in-person assessment, to improve earlier diagnosis of melanoma and non-melanoma skin cancer, and warned that further research-tested implementation trials are required to realize the potential of teledermatology for underserved populations (Duniphin, 2023). This, much like others we've already seen in machine learning apps, is a sign of a lack of maturity in real-world use, and of great promise and a known need, a sign that neither technology is the entire answer.

8.3 The Implementation of Rural TeleDermatology and Equity Must be Considered

However, a scoping review on the accuracy of diagnoses, barriers in implementation, and equity impacts of skin cancer care via teledermatology in rural settings identified that features such as standardized clinical workflows and dermoscopy integration when using a teledermatology platform and centralized review platforms were effective for implementing teledermatology, whereas workforce shortages, geographic isolation, and inequitable access to technology across high-risk rural populations remained as challenges (Parga et al., 2025). This continues the mal-distribution of workforce evidence from Section 4.1, noting that since teledermatology relies on having a reliable communications network and imaging devices at the point of care, there is a potential for a differential impact this could have in some communities of rural areas – a factor that should be taken into consideration based on the integration built-out of the teledermatology framework presented in Section 11.0.

8.4 An Integrated Machine Learning-Tele dermatology Triage Pathway

Figure 1 shows a structured and integrated pathway for triage that brings together the techniques and approaches reviewed in Sections 5.0 through 8.2, and then implements them in one, coordinated tripod that can be used in a home-based context that perhaps lacks the services that a trained triage pathologist could offer in an office setting. This pathway starts at first patient contact (primary care, a community clinic or a community health fair) where a patient uploads the image of their skin lesion to a smart phone, with optionally an attached dermatoscope, that is then

© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1719647

analyzed with a triage machine learning algorithm as described in Sections 5.1 and 5.2. This skin-tone-aware benchmarking of the performance of the benchmarked algorithm, such as what was defined in the DDI dataset (Daneshjou et al., 2022), is included as an interpretive step next to the raw algorithm result, creating a pathway towards implementing the equity safeguard claimed in Section 7.2.

If lesions are identified as being high risk or indeterminate, they are referred for the store-and-forward teledermatology review pathway, as outlined in Section 8.1 and in low-risk lesions the lesion is followed up in the usual primary care pathway. The pathway includes focus group designed referral navigation support such as clear communication, scheduling support and transportation or cost support prior to completed in-person evaluation and biopsy (where indicated) after the remote review. Not only does the explicit language on referrals as a distinct step add to the recommendations of Section 9.0, it also clearly signifies that referral non-completion is a significant independent point of "loss" that needs to be addressed specifically by dedicating resources and effort to it.

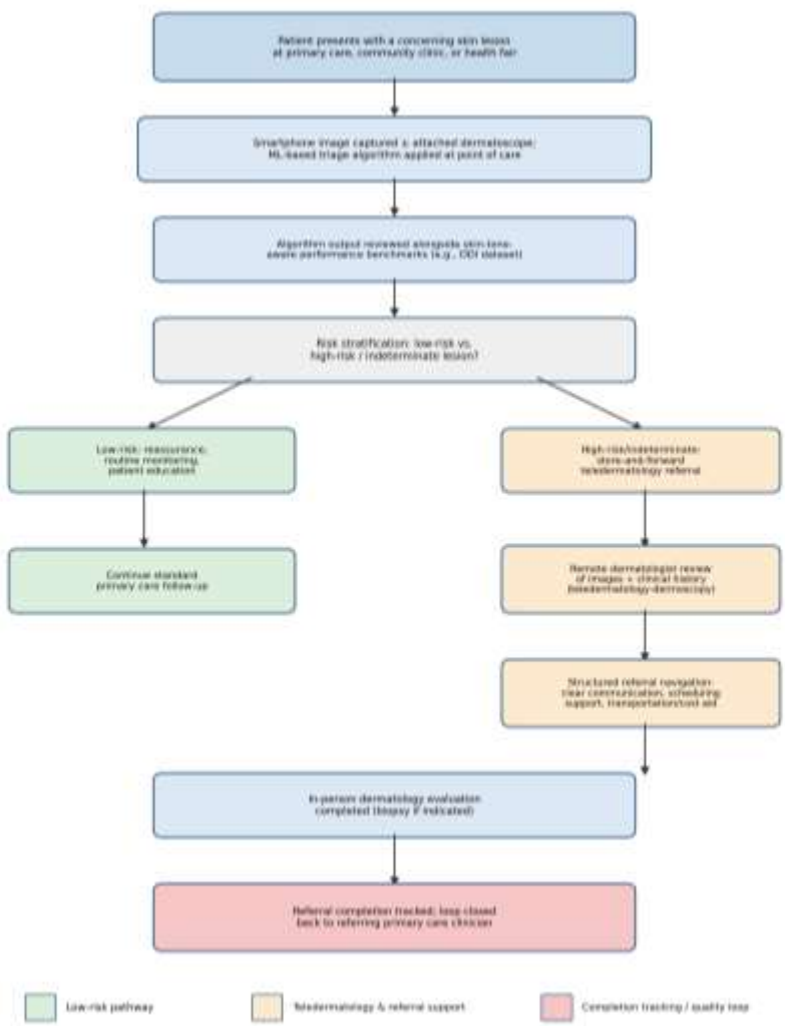

Figure 1: Integrated machine learning teledermatology triage pathway for underserved U.S. communities

VIII. IMPROVING DERMATOLOGICAL REFERRAL COMPLETION – THE 'LAST MILE PROBLEM

9.1 Life and Career Skills for New Hampshire Transitioning Students

Analysis of the success of machine learning triage and teledermatology in identifying a lesion that should be referred for specialist opinion would miss the point if it did not also explore the post-referral phase. Of 320 referrals made from primary care to dermatology in a one-year period at a single academic institution, just over half were evaluated within 30 days, more than a third made delays and almost 11% failed to be seen at all (Pagani et al., 2023). This is possibly the biggest empirical basis for the main thesis of this review: losses continue after appropriate referral starts – the improvements to detection of the condition in the upstream stages of the referral process do not necessarily mean improved outcomes at that stage unless this is tackled as well. Importantly, non-completions and delays were not evenly distributed, but disproportionately likely among patients who do not speak English and among minorities and racial and ethnic minorities (Pagani et al., 2023), resulting in a direct link between the literature on referrals and their completion and the findings on outcome disparities in Section 3.0.

Paragraph 7.2 uses the term "structural barriers" in the following context:

Overall, these differences align with the findings in Section 3.4: Patient experience – that patients who did not seem to receive adequate care to address their concerns may reasonably leave completing a subsequently recommended referral to a lower priority, especially if they encounter added travel time and distance and language barriers to accessing the care recommended by their physicians (Fahmy et al., 2023; Pagani et al., 2023). Effective interventions to close this gap will thus require more than logistical mechanisms - there will also need to be an element to the initial referral conversation that covers a communicative and relational dimension.

These are promising areas for future development of these systems.These represent some good directions for future development of these systems.

Table 2 lists the main studies examined under each of Sections 5.0 to 9.0, providing evidence of the stages in the journey from initial machine learning detection to ends of trip completion, including the population, main finding and how applicable this is to underserved communities. The complete pathway to which this review refers is clearly spelled out in this table: The fact that resources are dedicated at one point in this pathway will not make up for the absence of resources at another.

Table 2. Summary of Key Studies Tracing the Detection-to-Referral-Completion Pathway

Source	Population/design	Key finding	Relevance to underserved communities
Esteva et al. (2017)	129,450 clinical images; CNN vs. 21 dermatologists	Dermatologist-level classification performance achieved	Establishes technical feasibility of specialist-independent triage
Liu et al. (2020)	16,114 teledermatology cases, 17 sites	Broad differential diagnosis system for primary care use	Supports non-specialist first-contact clinicians directly
Tschandl et al. (2018)	10,015-image HAM10000 dataset	Large, multi-source benchmark dataset released publicly	Accelerated research but lacked deliberate skin-tone diversity
Freeman et al. (2020)	Systematic review, 9 studies/6 apps	Wide accuracy variation; unrepresentative study populations	Cautions against unvalidated consumer-facing tools alone
Daneshjou et al. (2021)	Scoping review of AI datasets/algorithms	Widespread underrepresentation of darker skin tones	Identifies root cause of downstream diagnostic disparities

© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1719647

Daneshjou et al. (2022)	DDI benchmark dataset (n=656 images)	27-36% AUC drop on diverse images; humans also underperform	Direct quantification of real-world equity gap
Venkatesh et al. (2024)	FDA authorization analysis (DermaSensor)	First AI skin cancer device authorized for primary care	Directly targets dermatologist-scarce settings
Feng et al. (2018)	US county-level dermatologist density, 1995-2013	Widening urban-rural workforce distribution gap	Documents structural access barrier motivating triage need
Parga et al. (2025)	Scoping review, rural teledermatology	Standardized workflows aid implementation; equity gaps persist	Identifies technology-access equity risk within teledermatology itself
Pagani et al. (2023)	320 urgent dermatology referrals, single institution	10.9% never completed; disparities by language/race	Identifies referral completion as independent, disparate bottleneck

IV. INSURING THE COORDINATION OF HEALTHCARE SERVICES WITH MENTAL HEALTH

Figure 2 shows the approximate proportional contribution of the principal factor categories identified throughout Sections 3.0 through 9.0 to overall diagnostic delay in underserved communities across the U.S. being specific to geographic and workforce maldistribution (Feng et al., 2018); algorithmic and diagnostic performance gaps related to darker skin tones and less common presentations (Daneshjou et al., 2021, 2022; Fliorent et al., 2024); referral non-completion and communication barriers (Pagani et al., 2023; Fahmy et al., 2023); broader socioeconomic and insurance barriers (Duniphin, 2023); and patient trust and perceived-respect disparities (Fahmy et al., 2023). This distribution is a necessary approximation, it is qualitatively synthesized from the heterogeneous set of sources reviewed, it does not necessarily represent any one data set and does not mean that a single factor category dominates or that a single-category intervention will be sufficient to fully solve the problem.

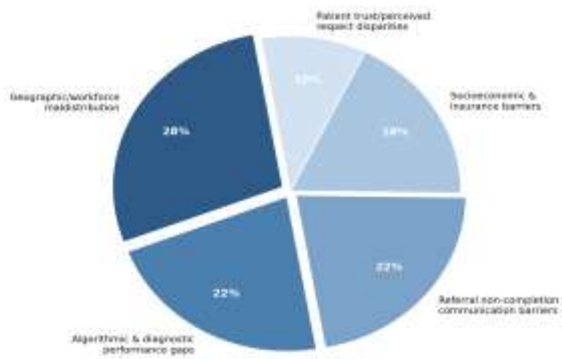

Figure 2: Approximate distribution of contributing factors to diagnostic delay in underserved U.S. communities.

XI. THIS COMPONENT IS ABOUT BEING BRUISED YET WHOLE

11.1 This course relates to goals

Machine learning-based triage (Section 5.0 and 6.0), teledermatology (Section 8.0) and structured referral-completion support (Section 9.0) fall into largely separate points of potential failure in the same diagnostic pathway and are likely to meaningfully improve outcomes when they are combined, as in Figure 1, rather than used in isolation. Machine learning triage is one solution to the detection bottleneck of the limited number of dermatologists; teledermatology is one solution to the "specialist-review bottleneck;" and structured referral processing adds to the behavioral and logistical "bottleneck," which has been previously under recognized but substantial, as an independent contributor to delay (Pagani et al., 2023). Future triage programs should thus monitor the patient along the entire journey and measure not just the rate of in-triage accuracy, but progress rate through the teledermatology consultation process and, most of all, through in-person referral success, especially as it relates to the groups identified previously (Section 9.0) where delay is disproportionately experienced.

11.2 Adopting an Innovative Approach to Policy nd Implementation Considerations

When moving this framework to the practice, there are multiple points to be considered, based on the evidence found. Skin-tone-disaggregated performance reporting is recommended for skin cancer devices that aim to be authorized for primary care, as done in the DDI benchmark methodology (Daneshjou et al., 2022), through regulatory evidence. The key equity considerations about technology access raised in the rural teledermatology literature (Parga et al. 2025) should be factored into the design of teledermatology infrastructure, so as to not exclude communities that already have the best connectivity. By offloading these costs and services as assistance or support (such as language-concordant communication, arranging travel and/or scheduling assistance), referral navigation becomes a core program component, not an unfunded add-on, and is directly addressing the pharmaceutical expenditure non-completion problem that Pagani et al. (2023) reported that is non-randomly distributed.

XII. DISCUSSION

12.1 What is the challenge facing innovators to achieve equity

Many of the bks cited reflect a conflict between the sophistication of today's dermatological AI and its slower development in terms of achieving equitable performance in the broader diversity of patients who are being treated. The story of Esteva et al. (2017) proving that dermatologists can perform at the level of their patients, followed by the evidence from Daneshjou et al. (2021, 2022) elucidating significant disparities on darker skin tones, presents an example reminiscent of the stark inequities that were only clearly visible through targeted and deliberate disaggregated study uncovered by Obermeyer et al. (2019) outside dermatology as well. The key point of this review is that innovation and equity need to be one and the same, prioritizing both innovation and equity, and both should start in the initial stages of development.

12.2 Tackling a drill with some pressure doesn't make it harder

An imbalance in research attention is a second pattern: between upstream (detection technology) and downstream (referral completion). Both the experience with machine learning and the experience with bias are quite vast and growing (see Section 5.0 and 6.0) and the experience with referral completion is quite small, but nonetheless important because found to have a very high 11% non-completion rate, and showed very strong language and racial disparities (Pagani et al., 2023) which may prove to be as consequential for outcomes for underserved communities as the combined work so far on machine learning and bias. Innovation in early detection must not replace, and should not divert attention from, provision of the specialist care follow-up to which the flagged patients are referred.

12.3 Research Gaps

Several gaps remain. The evidence base here for the fully integrated proposed triage pathway in Section 11.0 is not evidence-based as a single pathway, but rather separate evidence-based literatures on the accuracy of the pathway, pathway bias and teledermatology, and pathway completion through referrals. Data linked to AI-assisted triage and eventual completion of the referral and diagnosis remains limited; an extension of Daneshjou et al.'s (2022) approach into the domain of referral completion and diagnosis confirmation (Pagani et al., 2023). Further in-depth research into referral navigation interventions themselves, and not just the patterns of not completing them, is comparatively under-developed, and is a clear direction for future intervention focused research.

12.4 Limitations of This Review

There are a number of limitations to this review. It does not formally register a systematic review with dual independent screening or quantitative meta analytic pooling, but rather is a narrative synthesis of twenty carefully selected sources of information; readers who are interested in pooled estimates on any specific question — such as the exact size of the racial difference in melanoma survival, or the exact quality of a specific diagnostic classifier — are advised to consult specific systematic reviews on this limited question. The 20 sources had different methodological approaches and included registry analyses, single-institution cohorts, algorithm validation studies, scoping reviews and cross-sectional surveys, with no attempt made or appropriate for the analyses to compare the relative merits of the methodological approaches. Lastly, the synthesis of the integrated framework proposed in Figure 1 and Section 11.0 is a conceptual framework based on evidence reviewed in the literature as opposed to a previously validated protocol, and its actual effectiveness in practice has not been directly tested in the published literature.

XIII. CONCLUSION

Despite the generally improving prognosis for Americans developing skin cancer, there are significant and persistent disparities: although patients from minority racial/ethnic groups are diagnosed at a somewhat later stage of skin cancer than are non-Hispanic White patients, in independent analyses of both skin or other cancer registry data, those same minority groups had worse survival than non-Hispanic White skin cancer patients, and, a dermatology health care workforce growing more and more concentrated away from rural areas and the underserved populations that benefit from its palliative role most. The problem can be tackled in this way and is helped by advanced machine learning models that have been shown to deliver dermatologist-class accuracy and that are beginning to receive regulatory approval for primary-care use of AI devices – for the first time, as seen in their translation into real-world applications. This same study has revealed a major problem: dermatology AI created with historically unrepresentative datasets is significantly less effective for dark-colored skin and less common conditions exactly the populations where effect of current healthcare treatment has been well documented as being suboptimal. Added to this, the majority of patients referred to specialist redeployment for successful cross-cultural specific triage often fail to follow up for those in-person consultations, especially those who are not English-speaking and minority communities. Transformation of skin cancer triage in underserved communities in the U.S. can't happen with only technology; it must be purposeful, effective, and integrated into a skin cancer care pathway that is equitably evaluated and continuously monitored while the promise of advanced machine learning is substantial, the reality of being a patient living in a community that benefits most will look very different.

REFERENCES

[1] Daneshjou, R., Smith, M. P., Sun, M. D., Rotemberg, V., & Zou, J. (2021). Lack of transparency and potential bias in artificial intelligence data sets and algorithms: A scoping review. JAMA Dermatology, 157(11), 1362–1369.https://doi.org/10.1001/jamadermatol.2021.312

[2] Daneshjou, R., Vodrahalli, K., Novoa, R. A., Jenkins, M., Liang, W., Rotemberg, V., Ko, J., Swetter, S. M., Bailey, E. E., Gevaert, O., Mukherjee, P., Phung, M., Yekrang, K., Fong, B., Sahasrabudhe, R., Allerup, J. A. C., Okata-Karigane, U., Zou, J., & Chiou, A. S. (2022). Disparities in dermatology AI performance on a diverse, curated clinical image set. Science Advances, 8(32), Article eabq6147. https://doi.org/10.1126/sciadv.abq6147

[3] Dawes, S. M., Tsai, S., Gittleman, H., Barnholtz-Sloan, J. S., & Bordeaux, J. S. (2016). Racial disparities in melanoma survival. Journal of the American Academy of Dermatology, 75(5), 983–991. https://doi.org/10.1016/j.jaad.2016.06.006

[4] Duniphin, D. D. (2023). Limited access to dermatology specialty care: Barriers and teledermatology. Dermatology Practical & Conceptual, 13(1), Article e2023031. https://doi.org/10.5826/dpc.1301a31

[5] Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542(7639), 115–118. https://doi.org/10.1038/nature21056

[6] Fahmy, L. M., Schreidah, C. M., & Geskin, L. J. (2023). Racial and ethnic disparities in the perception of respect from physicians among skin cancer patients in the United States. JAAD

International, 11, 78–82. https://doi.org/10.1016/j.jdin.2023.01.009

[7] Feng, H., Berk-Krauss, J., Feng, P. W., & Stein, J. A. (2018). Comparison of dermatologist density between urban and rural counties in the United States. JAMA Dermatology, 154(11), 1265–1271. https://doi.org/10.1001/jamadermatol.2018.3022

[8] Fliorent, R., Fardman, B., Podwojniak, A., Javaid, K., Tan, I. J., Ghani, H., Truong, T. M., Rao, B., & Heath, C. (2024). Artificial intelligence in dermatology: Advancements and challenges in skin of color. International Journal of Dermatology, 63(4), 455–461. https://doi.org/10.1111/ijd.17076

[9] Freeman, K., Dinnes, J., Chuchu, N., Takwoingi, Y., Bayliss, S. E., Matin, R. N., Jain, A., Walter, F. M., Williams, H. C., & Deeks, J. J. (2020). Algorithm based smartphone apps to assess risk of skin cancer in adults: Systematic review of diagnostic accuracy studies. BMJ, 368, Article m127. https://doi.org/10.1136/bmj.m127

[10] Lam, M., Zhu, J. W., Hu, A., & Beecker, J. (2022). Racial differences in the prognosis and survival of cutaneous melanoma from 1990 to 2020 in North America: A systematic review and meta-analysis. Journal of Cutaneous Medicine and Surgery, 26(2), 181–188. https://doi.org/10.1177/12034754211052866

[11] Liu, Y., Jain, A., Eng, C., Way, D. H., Lee, K., Bui, P., Kanada, K., de Oliveira Marinho, G., Gallegos, J., Gabriele, S., Gupta, V., Singh, N., Natarajan, V., Hofmann-Wellenhof, R., Corrado, G. S., Peng, L. H., Webster, D. R., Ai, D., Huang, S., … Coz, D. (2020). A deep learning system for differential diagnosis of skin diseases. Nature Medicine, 26(6), 900–908. https://doi.org/10.1038/s41591-020-0842-3

[12] Mahendraraj, K., Sidhu, K., Lau, C. S. M., McRoy, G. J., Chamberlain, R. S., & Smith, F. O. (2017). 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(15), Article e6258. https://doi.org/10.1097/MD.0000000000006258

[13] Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464), 447–453. https://doi.org/10.1126/science.aax2342

[14] Pagani, K., Lukac, D., Olbricht, S. M., Aronson, M. D., Benneyan, J. C., Fernandez, L., Salant, T., Schiff, G. D., Shafiq, U., Sternberg, S. B., & McGee, J. S. (2023). Urgent referrals from primary care to dermatology for lesions suspicious for skin cancer: Patterns, outcomes, and need for systems improvement. Archives of Dermatological Research, 315(5), 1397–1400. https://doi.org/10.1007/s00403-022-02456-7

[15] Parga, A. D., Peng, D. S., Vu, T. N., Kizy, S. M., & Khan, A. (2025). Diagnostic accuracy, implementation barriers, and equity implications of teledermatology in rural skin cancer: Scoping review. JMIR. https://doi.org/10.2196/77443

[16] Qian, Y., Johannet, P., Sawyers, A., Yu, J., Osman, I., & Zhong, J. (2021). The ongoing racial disparities in melanoma: An analysis of the Surveillance, Epidemiology, and End Results database (1975–2016). Journal of the American Academy of Dermatology, 84(6), 1585–1593. https://doi.org/10.1016/j.jaad.2020.08.097

[17] Rezk, E., Eltorki, M., & El-Dakhakhni, W. (2022). Leveraging artificial intelligence to improve the diversity of dermatological skin color pathology: Protocol for an algorithm development and validation study. JMIR Research Protocols, 11(3), Article e34896. https://doi.org/10.2196/34896

[18] Tschandl, P., Rosendahl, C., & Kittler, H. (2018). The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Scientific Data, 5, Article 180161. https://doi.org/10.1038/sdata.2018.161

[19] Venkatesh, K. P., Kadakia, K. T., & Gilbert, S. (2024). Learnings from the first AI-enabled skin cancer device for primary care authorized by FDA. npj Digital Medicine, 7(1), Article 156. https://doi.org/10.1038/s41746-024-01161-1

[20] Zavala, V. A., Bracci, P. M., Carethers, J. M., Carvajal-Carmona, L., Coggins, N. B., Cruz-Correa, M. R., Davis, M., de Smith, A. J., Dutil,

J., Figueiredo, J. C., Fox, R., Graves, K. D., Gomez, S. L., Llera, A., Neuhausen, S. L., Newman, L., Nguyen, T., Palmer, J. R., Palmer, N. R., Pérez-Stable, E. J., Piawah, S., Rodriquez, E. J., Sanabria-Salas, M. C., Schmit, S. L., Serrano-Gomez, S. J., Stern, M. C., Weitzel, J., Yang, J. J., Zabaleta, J., Ziv, E., & Fejerman, L. (2021). Cancer health disparities in racial/ethnic minorities in the United States. British Journal of Cancer, 124(2), 315–332. https://doi.org/10.1038/s41416-020-01038-6