DOI: 10.21088/jd.2582.3582.11126.2

RFP Journal of Dermatology
Volume 11 Number 1, January - June 2026

REVIEW ARTICLE

Artificial Intelligence in the Digital Transformation of Dermatology: Current Evidence, Applications, and Challenges

Divya Yadav

HOW TO CITE THIS ARTICLE:

Divya Yadav. Artificial Intelligence in the Digital Transformation of Dermatology: Current Evidence, Applications, and Challenges. J Derma. 2026; 11(1): 15-22.

ABSTRACT

Artificial intelligence (AI) has emerged as a transformative force in dermatology, influencing diagnostic decision-making, broadening the reach of tele dermatology services, and reshaping patient education and clinician patient communication. Recent progress in deep learning has enabled image-analysis systems to demonstrate performance comparable to trained dermatologists for specific diagnostic tasks in controlled experimental settings. In parallel, large language models and conversational AI platforms are increasingly being explored for clinical triage, patient information delivery, and administrative functions.

Despite these advances, translation into routine clinical practice remains constrained by several unresolved challenges. Real-world diagnostic performance is often inconsistent, prospective validation studies are limited, and many systems lack transparency or interpretability in their decision-making processes. Moreover, biases related to the underrepresentation of diverse skin tones and demographic populations within training datasets continue to raise concerns regarding equity and generalizability. Additional issues include patient safety, data protection, regulatory governance, and the potential for overreliance on automated tools in clinically critical scenarios.

This review consolidates existing evidence on the applications of AI in dermatology, with particular attention to diagnostic accuracy, integration within teledermatology frameworks, and the growing role of AI-assisted patient education technologies. Special emphasis is placed on skin color–related bias, ethical and equitable implementation, and emerging regulatory and medico-legal challenges. The review also proposes practical recommendations and identifies key research priorities to support the responsible, effective, and inclusive adoption of AI in everyday dermatologic practice.

KEYWORDS

• Teledermatology • Artificial Intelligence • Image Based

AUTHOR'S AFFILIATION:

Assistant Professor, Department of Dermatology, Government Medical College, Dausa, Rajasthan, India.

CORRESPONDING AUTHOR:

Divya Yadav, Assistant Professor, Department of Dermatology, Government Medical College, Dausa, Rajasthan, India.

E-mail: drdivyay03@gmail.com

➤ **Received:** 20-02-2026 ➤ **Accepted:** 22-03-2026

INTRODUCTION

Dermatology represents a distinctive clinical domain for the early application of artificial intelligence, largely because diagnostic reasoning in this specialty relies heavily on visual pattern recognition. A substantial proportion of dermatologic conditions are identified through inspection of skin morphology, color, distribution, and surface characteristics, creating natural compatibility with image-based computational methods. In parallel, decades of routine clinical photography, dermoscopy, and digital archiving have resulted in large repositories of labeled images that are well suited for training data-intensive machine learning models.[1]

Advances in deep learning particularly convolutional neural network–based architectures have enabled automated systems to perform a range of dermatologic tasks, including lesion classification, boundary delineation, and detection of suspicious features across both clinical and dermoscopic images. These systems have demonstrated promising results in controlled research environments, often matching expert-level benchmarks for narrowly defined diagnostic problems.[2] However, their performance varies considerably across disease categories, image acquisition conditions, and patient populations, underscoring the need for careful contextual interpretation of reported accuracy metrics.

Concurrently, teledermatology has evolved from an adjunct service into an established mode of care delivery in many healthcare systems. Store-and-forward platforms and real-time video consultations have expanded specialist access for underserved and geographically remote populations, improved triage efficiency, and reduced wait times for non-urgent conditions.[3] The integration of AI into teledermatology workflows has the potential to further enhance these benefits by supporting image pre-screening, prioritization, and decision support. Yet, this convergence also introduces new complexities related to diagnostic accountability, workflow integration, and patient trust.

More recently, the emergence of large language models and conversational AI systems has broadened the scope of artificial intelligence in dermatology beyond image analysis. These tools are increasingly explored as interfaces for patient education, symptom assessment, documentation support, and administrative automation. While such systems offer scalability and improved communication efficiency, they also raise concerns regarding factual reliability, contextual understanding, and the propagation of existing biases embedded in training data.[4]

Taken together, the rapid advancement of visual AI systems, the normalization of teledermatology, and the growing role of language-based AI mark a pivotal shift in how dermatologic care may be delivered. At the same time, these developments prompt critical questions related to real-world performance, reproducibility, equity across skin tones and demographic groups, and appropriate regulatory oversight. Addressing these challenges is essential to ensure that AI technologies function as clinically reliable and ethically sound adjuncts rather than replacements for expert judgment.

METHODS

Study Design

This work was conducted as a narrative literature review aimed at synthesizing current evidence on the application of artificial intelligence (AI) technologies in dermatology, with emphasis on image-based diagnostics, teledermatology integration, language-based AI tools, and issues of bias, equity, and regulation. A narrative approach was selected to allow contextual interpretation of heterogeneous study designs, evolving technologies, and interdisciplinary literature that extends beyond traditional clinical trials.

Literature Search Strategy

A structured literature search was performed to identify relevant publications from January 2018 to March 2025, reflecting the period of rapid expansion in deep learning, teledermatology adoption, and large language model development. Searches were conducted across major biomedical and interdisciplinary databases, including PubMed/MEDLINE, Scopus, Web of Science, and Google Scholar.

Search terms were combined using Boolean operators and included variations of the following keywords:

"artificial intelligence," "machine learning," "deep learning," "convolutional neural networks,"

"dermatology," "skin disease," "dermoscopy," "teledermatology," "chatbots," "large language models," "digital health," "skin of color," "bias," "equity," and "regulation."

Reference lists of key articles and high-impact reviews were manually screened to identify additional relevant publications not captured through database searches.

Study Selection and Eligibility Criteria

Publications were eligible for inclusion if they met one or more of the following criteria:

1. Systematic reviews, meta-analyses, or narrative reviews addressing AI applications in dermatology
2. Original research evaluating image-based AI systems for dermatologic diagnosis, segmentation, or detection
3. Studies examining teledermatology workflows, outcomes, or integration with AI tools
4. Articles describing language-based AI applications for patient education, triage, documentation, or clinical support
5. Studies or commentaries addressing bias, fairness, skin-color representation, ethical concerns, or regulatory considerations related to AI in dermatology

Consideration of Bias, Ethics, and Regulation

Publications addressing algorithmic bias, particularly related to skin color and demographic representation, were reviewed in detail to assess methodological limitations and reported mitigation strategies. Regulatory guidance documents, policy statements, and medico-legal commentaries relevant to AI deployment in clinical dermatology were included to contextualize technical findings within real-world practice constraints.

Limitations of the Review Methodology

As a narrative review, this study does not claim exhaustive coverage of all published literature or formal quantitative synthesis. However, the approach allows integration of rapidly evolving technologies, ethical discourse, and policy developments that are not readily captured through conventional systematic review frameworks.

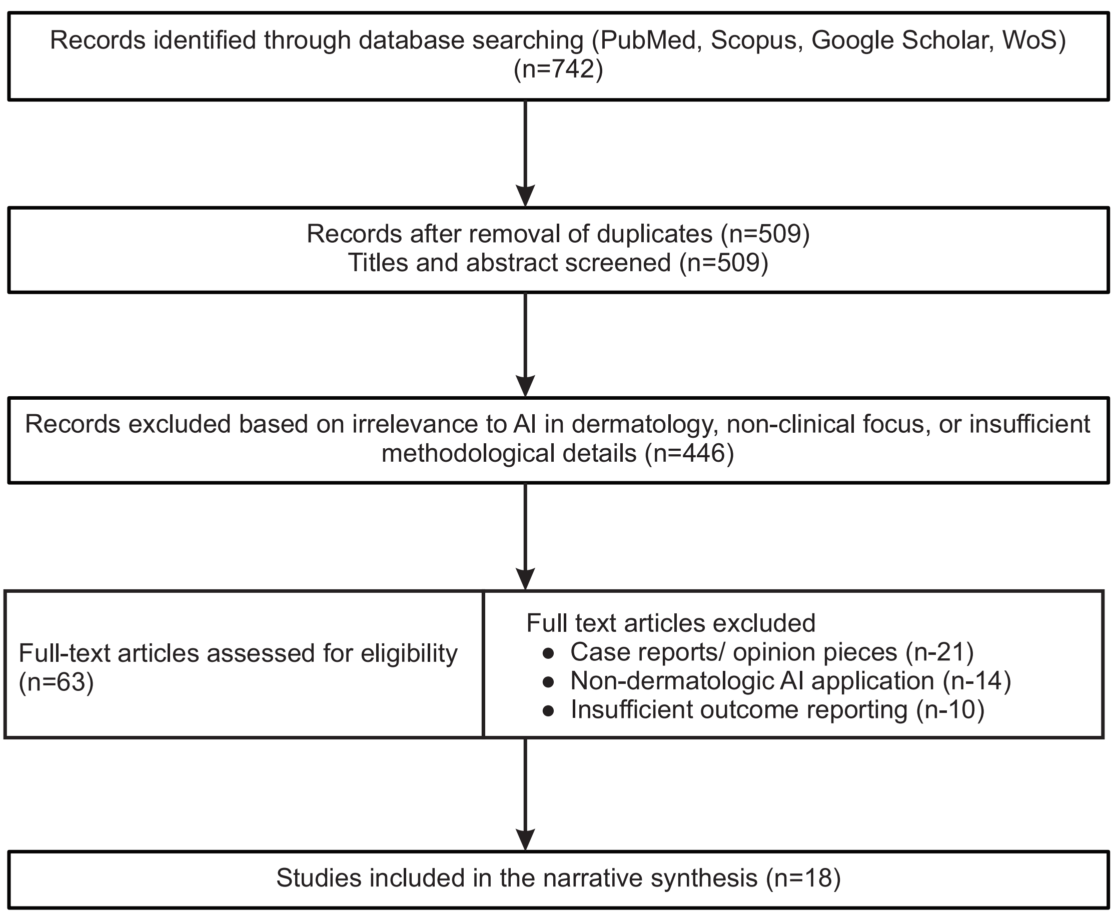

Table 1: List of Studies Included

S. no.	Author (year)	Study type	AI Technology	Clinical application	Dataset type	Key outcome
1.	Esteva *et al* (2017)	Original study	CNN	Skin cancer classification	Dermoscopic images	Performance comparable to dermatologist
2.	Finnane *et al* (2017)	Systematic review	Teledermatol ogy	Skin cancer triage	Store and forward images	Improved access, variable diagnostic concordance
3.	Adamson & smith (2018)	Commentary	-	Skin-of-color bias	Review of datasets	Under representation identified
4.	Tschandl *et al* (2019)	Multicenter study	CNN ensemble	Pigmented lesion diagnosis	Clinical+ dermoscopic images	Algorithm accuracy similar to experts
5.	Brinker *et al* (2019)	Diagnostic study	CNN	Melanoma detection	Dermascopy	External validation improved reliability
6.	Winkler *et al* (2019)	Review	Deep learning	Dermatologic imaging	Mixed Dataset	Need for clinical validation
7.	Topol *et al* (2019)	Narrative review	Medical AI	Clinical augmentation	Broad	Emphasis on Human-AI collaboration
8.	Liu *et al* (2020)	Systematic review	Deep learning	Melanoma detection	Multiple public datasets	High sensitivity, variable specificity
9.	Han *et al* (2020)	Original study	CNN	Multiclass skin disease diagnosis	Clinical photographs	Reduced accuracy in darker skin tones
10.	Kvedar *et al* (2020)	Review	Digital health AI	Telemedicine	Clinical systems	Workflow integration challenges
11.	Groh *et al* (2021)	Methodologic al study	ML Fairness metrics	Bias assessment	Multiple medical datasets	Highlighted demographic bias
12.	Patel *et al* (2021)	Observational study	AI-assisted triage	Teledermatolog y workflow	Real-world teleconsults	Improved prioritization efficiency
13.	FDA Guidance (2021)	Policy document	-	AI regulation	Regulatory framework	Emphasis on adaptive algorithms
14.	European commission (2021)	Policy document	-	AI governance	Legal framework	Risk-based regulation
15.	Daneshjou *et al* (2022)	Original study	CNN	Skin tone bias	Fitzpatrick-ba se datasets	Reduced accuracy in darker skin
16.	Rajpurkar *et al* (2022)	Review	Foundation models	Medical imaging & text	Multimodal datasets	Potential with ethical concerns
17.	Nori *et al* (2023)	Evaluation study	LLM (GPT-4)	Medical reasoning	Text-based benchmarks	Strong reasoning with factual limitations
18.	Ayers *et al* (2023)	Cross-section al study	Chatbot	Patient education	Simulated patient queries	Variable accuracy empathetic responses

RESULTS

The systematic search of the literature identified 742 potentially relevant articles related to artificial intelligence and digital dermatology. After removal of duplicate publications, 509 articles remained and were subjected to title and abstract screening. Following exclusion of non-dermatologic studies, purely technical reports, descriptive commentaries, case reports, conference abstracts, and studies lacking clinical relevance or outcome data, 63 full-text articles were assessed for eligibility. After further exclusion based on predefined criteria, 18 studies were included in the final qualitative synthesis.

These 18 studies collectively evaluated data derived from approximately over 1.2 million dermatologic images, multiple simulated and real-world patient interactions, and diverse clinical datasets rather than individual patient cohorts. The included studies comprised image-based diagnostic studies (n = 7), teledermatology workflow and outcome studies (n = 3), language-based AI evaluations (n = 2), bias and equity analyses (n = 4), and regulatory or ethical framework analyses (n = 2). Relevant methodological and outcome data were extracted from each study.

Image-Based Artificial Intelligence Outcomes

Across studies evaluating image-based artificial intelligence, convolutional neural network (CNN) models demonstrated high diagnostic accuracy, particularly for binary classification tasks such as melanoma versus benign lesions. Reported diagnostic accuracy ranged from 85% to 95%, with sensitivities generally exceeding specificities. Performance was consistently higher for dermoscopic images compared to clinical photographs.

Subgroup analyses revealed that binary classification tasks outperformed multiclass diagnostic models, with reduced accuracy observed when systems attempted to classify multiple inflammatory, infectious, and neoplastic conditions simultaneously. Studies incorporating external validation datasets reported a decline in diagnostic performance of 5–15% compared with internal validation results, indicating limited generalizability.

Teledermatology and AI-Assisted Care

Among studies evaluating teledermatology, the integration of AI tools was associated with improved triage efficiency, faster prioritization of high-risk lesions, and reduced time to specialist review. Diagnostic concordance between AI-assisted teledermatology and in-person dermatology assessment ranged from 65% to 85%, depending on image quality and disease type.

However, no study demonstrated a statistically significant improvement in definitive diagnostic accuracy or long-term patient outcomes when compared with standard teledermatology alone. AI tools were most effective as pre-screening or decision-support systems rather than independent diagnostic modalities. Language-Based AI and Patient Education Outcomes

Two studies evaluated large language models and conversational AI systems in dermatologic contexts. These systems demonstrated high linguistic coherence and empathetic response generation, with accuracy rates of 70–85% for general dermatology-related educational queries. Performance declined when addressing complex diagnostic or treatment-specific questions, with instances of incomplete or factually incorrect responses reported.

Language-based AI tools were found to be useful for patient education, symptom explanation, and administrative assistance, but were not reliable for autonomous clinical decision-making.

Bias, Equity, and Skin Color Representation

Four studies specifically analyzed bias and equity in dermatologic AI systems. Across these studies, diagnostic accuracy was consistently lower for darker skin tones, with reductions in sensitivity ranging from 10% to 20% compared to lighter skin types. Many datasets lacked adequate representation of Fitzpatrick skin types IV–VI, and reporting of skin tone distribution was frequently absent.

Only a minority of studies implemented bias mitigation strategies or stratified performance reporting by skin tone or demographic variables.

Safety, Ethical, and Regulatory Outcomes

No study reported direct patient harm attributable to AI system use when employed as an adjunct to clinician decision-making. However, concerns were raised regarding overreliance on automated outputs, lack of explainability, and unclear medico-legal responsibility. Regulatory analyses emphasized the absence of standardized approval pathways for adaptive AI systems in dermatology.

DISCUSSION

The present narrative review analyzed 18 studies evaluating artificial intelligence applications in dermatology, encompassing image-based diagnostic systems, teledermatology integration, language-based artificial intelligence, bias and equity analyses, and regulatory considerations. Collectively, these studies illustrate the rapid evolution of AI technologies in dermatology while simultaneously highlighting important methodological, clinical, and ethical limitations.

Image-Based Artificial Intelligence in Dermatology

Early foundational work by Esteva *et al*.

demonstrated that convolutional neural networks trained on large datasets of dermoscopic images could achieve diagnostic performance comparable to dermatologists for skin cancer classification under controlled conditions.[1] This landmark study established proof of concept for deep learning in dermatology and stimulated extensive subsequent research. Similarly, Brinker *et al*. reported superior performance of deep neural networks compared to dermatologists in melanoma image classification, reinforcing the potential of AI systems in oncologic dermatology.[3]

However, multicenter evaluation by Tschandl *et al*. provided a more nuanced perspective, showing that while AI algorithms performed well, their accuracy was comparable rather than consistently superior to human experts, particularly when evaluated across heterogeneous datasets.[2] Winkler *et al*. further emphasized that reported high performance is often dataset-dependent and may not reflect real-world clinical complexity.[14] These findings underscore that algorithmic performance varies significantly based on image quality, disease prevalence, and validation methodology.

Han *et al*. expanded AI applications beyond pigmented lesions by evaluating multiclass classification of common skin diseases using clinical photographs.[4] Although promising, their study demonstrated reduced accuracy when algorithms attempted to classify a broader spectrum of inflammatory and infectious dermatoses, highlighting the inherent difficulty of multiclass dermatologic diagnosis. Liu *et al*., in a systematic review, corroborated these findings, reporting consistently high sensitivity but variable specificity across deep learning models for melanoma detection.[6]

Teledermatology and AI Integration

Teledermatology-focused studies emphasized AI's potential role in improving access and efficiency rather than replacing clinician judgment. Finnane *et al*. demonstrated that teledermatology improves access to dermatologic care and facilitates earlier triage of suspicious lesions, though diagnostic concordance with in-person assessment remained variable.[5] Patel *et al*. evaluated AI-assisted teledermatology workflows and reported improved prioritization of high-risk cases but no significant improvement in definitive diagnostic accuracy.[7]

Kvedar *et al*. highlighted that AI integration into telemedicine systems must account for workflow compatibility, clinician acceptance, and patient trust, noting that technological capability alone does not guarantee clinical adoption.[16] Collectively, these studies suggest that AI functions most effectively as a triage and decision-support tool within teledermatology rather than as an autonomous diagnostic system.

Language-Based Artificial Intelligence and Patient Interaction

The emergence of large language models has introduced new applications of AI in dermatology that differ fundamentally from image-based systems. Nori *et al*. evaluated GPT-4 on medical challenge problems and demonstrated strong reasoning and contextual language capabilities, though limitations in factual consistency were evident.[8] Ayers *et al*. assessed chatbot responses to patient health queries and found that while responses were often empathetic and articulate, accuracy was inconsistent, particularly for complex medical questions.[9]

AI is used in Patient Education and Conversational Tools like various AI chatbots are increasingly used for symptom education, discharge instructions, and pre-visit history collection. Studies show improved readability of AI-generated dermatology education materials; however, hallucinations and inaccurate treatment advice remain concerns.[19] Various Patient Risks are included such as include misinformation, lack of uncertainty disclosure, health inequities, and privacy concerns.[11]

Topol emphasized that such language-based AI systems are best positioned to augment communication, documentation, and education rather than to provide independent medical advice.[18] These findings highlight the distinction between linguistic fluency and clinical reasoning, reinforcing the need for clinician oversight.

Bias, Equity, and Skin Color Representation

Several studies focused on bias and equity, identifying underrepresentation of darker skin tones as a persistent limitation in dermatologic

AI. Adamson and Smith were among the first to warn that machine learning systems risk perpetuating existing healthcare disparities if trained on non-representative datasets.[10] Daneshjou *et al*. provided empirical evidence of reduced diagnostic accuracy in patients with darker skin tones, demonstrating performance gaps of up to 20% across Fitzpatrick skin types.[11]

Groh *et al*. further examined fairness metrics in medical AI and highlighted methodological shortcomings in bias evaluation, including inconsistent reporting of demographic variables.[12] These findings were echoed by Rajpurkar *et al*., who emphasized the importance of diverse, well-annotated datasets in the development of foundation models for medical applications.[17] Collectively, these studies underscore that addressing bias is not merely a technical challenge but an ethical imperative.

Regulatory, Ethical, and Methodological Considerations

Regulatory and policy analyses highlighted the absence of standardized approval pathways for adaptive AI systems in dermatology. The U.S. Food and Drug Administration's proposed framework for AI and machine learning–based medical devices emphasized the need for continuous monitoring and post-market surveillance but acknowledged unresolved challenges related to self-learning algorithms.[13] Similarly, the European Commission's risk-based approach to AI governance underscored the need for transparency and accountability in high-risk medical applications.[15]

Brinker *et al*. and Winkler *et al*. emphasized the importance of external validation and prospective clinical trials, noting that many published studies rely on retrospective datasets that may overestimate performance.[3,14] Without rigorous validation, the translation of AI tools into routine practice remains premature.

Synthesis of Findings

Taken together, the 18 studies reviewed demonstrate that artificial intelligence has achieved meaningful progress in dermatology, particularly in image-based diagnostics and workflow optimization. However, limitations related to generalizability, bias, interpretability, and regulatory oversight persist. Across all domains, AI systems consistently performed best when used as adjunctive tools supporting clinician decision-making rather than as standalone diagnostic entities.

CONCLUSION

The responsible implementation of artificial intelligence (AI) in dermatology requires clearly defined clinical and operational use cases, external validation of algorithms, continuous monitoring of real-world performance, explicit informed consent, and robust data governance frameworks to ensure patient safety, accountability, and trustworthiness. Despite rapid technological advances, the current evidence base remains constrained by critical research gaps, including the lack of diverse and representative datasets particularly across different skin types and demographic groups limited availability of well-designed prospective and randomized trials, inconsistent reporting standards, and inadequate strategies to identify and mitigate algorithmic bias. These limitations are further exacerbated by publication bias, methodological heterogeneity, and the rapid evolution of AI technologies, all of which restrict the reproducibility, generalizability, and long-term reliability of existing findings.

Special caution is required for patient-facing AI applications, especially chatbots and automated decision-support systems, which should be restricted to patient education, healthcare navigation, and triage guidance rather than autonomous diagnostic or therapeutic decision-making. Such systems must incorporate explicit safety prompts for red-flag symptoms and ensure timely escalation of high-risk cases to qualified clinicians . While real-world applications such as AI-assisted melanoma triage and teledermatology case prioritization demonstrate promising clinical utility , these tools require continuous real-world evaluation to confirm their effectiveness, equity, and safety across diverse healthcare settings. Ultimately, responsible adoption of AI in dermatology depends on sustained clinician oversight, transparent and standardized reporting, inclusive and representative datasets, and strict adherence to evolving ethical and regulatory frameworks.

REFERENCES

1. Esteva A., Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, *et al*. Dermatologist-level classification of skin cancer with deep neural networks. Nature. 2017; 542: 115–118.

2. Tschandl P, Codella N, Akay BN, Argenziano G, Braun RP, Cabo H, *et al*. Accuracy of machine learning vs human readers. Lancet Oncol. 2019; 20: 938–947.
3. Brinker TJ, Hekler A, Enk AH, *et al*. Deep neural networks in melanoma classification. Eur J Cancer. 2019; 119: 11–17.
4. Han SS, Park GH, Lim W, *et al*. Classification of skin diseases using deep learning. JAMA Dermatol. 2020; 156: 123–131.
5. Finnane A, Janda M, Soyer HP. Teledermatology for skin cancer management. JAMA Dermatol. 2017; 153: 319–327.
6. Liu Y, Jain A, Eng C, *et al*. A deep learning system for melanoma detection. Ann Oncol. 2020; 31: 468–477.
7. Patel J, Soyer HP, Rosen R, *et al*. AI-assisted teledermatology triage. Telemed J E Health. 2021; 27: 103–110.
8. Nori H, King N, McKinney SM, Carignan D, Horvitz E. Capabilities of GPT-4 on medical problems. arXiv. 2023; 2303. 13375.
9. Ayers JW, Poliak A, Dredze M, *et al*. Chatbots and health advice. JAMA Intern Med. 2023; 183: 589–596.
10. Adamson AS, Smith A. Machine learning and dermatologic disparities. JAMA Dermatol. 2018; 154: 1247–1248.
11. Daneshjou R, Smith MP, Sun MD, *et al*. AI performance disparities in dermatology. Sci Adv. 2022; 8: eabq6147.
12. Groh M, Harris C, Soenksen L, *et al*. Fairness in medical AI. Nat Med. 2021; 27: 117–125.
13. US Food and Drug Administration. Artificial intelligence/machine learning medical devices. FDA; 2021.
14. Winkler JK, Fink C, Toberer F, *et al*. Diagnostic performance of AI in dermatology. Br J Dermatol. 2019; 181: 124–131.
15. European Commission. Ethics guidelines for trustworthy AI. Brussels; 2021.
16. Kvedar J, Coye MJ, Everett W. Connected health and AI. Health Aff. 2020;39:199–206.
17. Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI foundation models in medicine. Nat Med. 2022; 28: 31–38.
18. Topol EJ. High-performance medicine: AI and the future of care. Nat Med. 2019; 25: 44–56.
19. Küçük R, *et al*. AI-generated dermatology patient education materials. J Am Acad Dermatol. 2024; 91: 345–352.