NARRATIVE REVIEW / MEDICAL & HEALTH SCIENCES

Artificial Intelligence in Dermatology

Recent Scientific Advances, Clinical Applications, and Future Perspectives

Torello Lotti[1] · Milo Dolci[2]*

[1] Istituto Superiore di Medicina, Florence, Italy | professor@torellolotti.it

[2] Prisca Sapientia Institute, Zurich, Switzerland | director@priscasapientia.university

CORRESPONDING AUTHOR

Milo Dolci · director@priscasapientia.university

PUBLICATION

Volume 2 · Issue 1 · 2026

ABSTRACT

Background: Artificial intelligence (AI) is increasingly applied to dermatological image interpretation, diagnostic classification, clinical decision support, and digital healthcare. Objective: This narrative review examines major advances, clinical applications, methodological limitations, and future directions for AI in dermatology. Methods: Selected foundational and methodological literature is critically synthesized, emphasizing diagnostic accuracy, external validation, clinical utility, fairness, and regulatory considerations. Findings: Deep-learning systems have demonstrated strong performance in selected skin-lesion classification tasks, including melanoma recognition. However, retrospective image-based results cannot automatically be generalized to clinical populations, imaging devices, and diverse skin phototypes. Human-AI collaboration may improve decisions under appropriate conditions, while large language models and multimodal systems introduce opportunities and risks involving factual accuracy, privacy, and automation bias. Conclusions: Dermatological AI should be evaluated as a clinically supervised support technology. Responsible adoption requires representative datasets, prospective validation, calibrated predictions, ongoing monitoring, equitable performance, and sustainable integration into clinical workflows.

KEYWORDS

Dermatology; Artificial Intelligence; Machine Learning; Deep Learning; Dermoscopy; Melanoma; Skin Cancer; Clinical Decision Support; Explainable AI; Large Language Models; Algorithmic Bias; Digital Dermatology.

© 2026 World Health Academy Publishing House

1. INTRODUCTION Artificial intelligence (AI) has become a major area of medical innovation, introducing computational methods for pattern recognition, image analysis, clinical decision support, and the organization of health information. Dermatology is particularly suited to investigation using computer vision because many cutaneous disorders are assessed through visual morphology, pigmentation, lesion distribution, and dermoscopic structures.

Dermatological diagnosis, however, extends beyond the classification of an isolated photograph. Patient history, anatomical location, lesion evolution, symptoms, skin phototype, risk factors, and histopathological findings can all influence clinical interpretation. Consequently, performance achieved on curated image datasets must be distinguished from the reliability of a system used in routine clinical care.

Esteva and colleagues demonstrated dermatologist-comparable performance for selected skin cancer image-classification tasks using deep neural networks [1]. Haenssle and colleagues subsequently compared a convolutional neural network with dermatologists in melanoma recognition using dermoscopic images [2], while later investigations considered broader lesion spectra [3]. These findings established technical feasibility, not the clinical equivalence of autonomous AI and comprehensive specialist examination.

Applications now extend to inflammatory skin diseases, teledermatology, disease-severity estimation, clinical documentation, and multimodal systems integrating text with images. This review examines the clinical promise of these technologies alongside the evidence needed to establish their safety, fairness, and usefulness.

2. REVIEW METHODOLOGY This article is a critical narrative review of selected peer-reviewed investigations, methodological literature, and institutional guidance relevant to AI in dermatology. It considers foundational classification studies, human-AI collaboration, differential diagnosis, dataset bias, external validation, generative AI, and regulatory requirements.

The synthesis distinguishes retrospective discrimination from clinical effectiveness and examines sensitivity, specificity, calibration, generalizability, and patient-safety implications. Because this is not a systematic review, it does not claim exhaustive retrieval, formal risk-of-bias scoring, or pooled effect estimates. The conclusions are qualitative and intended to identify clinically meaningful research priorities.

3. COMPUTATIONAL FOUNDATIONS

3.1. Machine Learning and Feature-Based Classification

Machine learning refers to methods that infer predictive patterns from data. Conventional dermatological systems often rely on engineered image features such as lesion asymmetry, border irregularity, color variation, texture, and geometry. These features may be processed by support vector machines, random forests, or other classifiers. Their performance depends on feature relevance, labeling quality, and the representativeness of training examples.

3.2. Deep Learning and Convolutional Neural Networks

Deep learning uses multilayer networks to learn progressively complex representations. Convolutional neural networks (CNNs) can extract spatial patterns from clinical and dermoscopic images without requiring every discriminative feature to be specified in advance. Early layers may respond to edges and textures, whereas deeper representations combine these into lesion-related patterns. A high model confidence score, however, should not be confused with a calibrated probability of disease.

3.3. Transfer Learning and Multimodal Architectures

Transfer learning adapts models trained on large datasets to specialized dermatological tasks and can be useful when expertly annotated images are scarce. Vision transformers and multimodal architectures may integrate photographs, dermoscopy, patient histories, and clinical notes. Such integration better reflects real clinical reasoning but creates challenges involving missing data, privacy, spurious correlations, and model complexity.

4. ARTIFICIAL INTELLIGENCE IN SKIN CANCER DETECTION

4.1. Melanoma Recognition

Melanoma detection remains among the most extensively studied applications of dermatological AI. Dermoscopy reveals pigment networks, vascular structures, and other patterns that support image-based classification. Landmark studies reported promising performance in selected experimental settings [1-3]. Nevertheless, clinical melanoma diagnosis involves lesion evolution, risk factors, examination of multiple lesions, and decisions regarding biopsy or surveillance.

False-negative predictions may delay cancer diagnosis, whereas false-positive predictions may increase unnecessary procedures, patient anxiety, and healthcare expenditure. Clinical evaluation must therefore consider decision thresholds, prevalence, and downstream outcomes, rather than classification accuracy alone.

4.2. Non-Melanoma Skin Cancer

Basal cell carcinoma and cutaneous squamous cell carcinoma present heterogeneous morphological appearances and may overlap visually with benign or premalignant conditions. Models trained on restricted binary comparisons cannot be assumed to recognize the complete spectrum of cutaneous malignancy. Representative differential diagnoses and independent validation are essential.

4.3. Human-AI Collaboration

The most clinically plausible role for many current systems is to support the dermatologist. Tschandl and colleagues investigated human-computer collaboration for skin cancer recognition [4]. Such approaches must be evaluated at the level of the clinician-AI team: an algorithm that performs well independently may not improve clinical decisions if users overtrust incorrect predictions or disregard appropriate uncertainty.

5. INFLAMMATORY AND CHRONIC DERMATOLOGICAL DISEASES AI research increasingly addresses psoriasis, atopic dermatitis, acne, rosacea, vitiligo, and other chronic disorders. Potential applications include lesion segmentation, automated lesion counting, severity estimation, and longitudinal photographic comparison. Differential diagnosis across multiple skin diseases has also been investigated using deep learning [5].

For psoriasis, image analysis may assist estimation of erythema, scaling, and affected surface area. However, symptoms, quality of life, joint involvement, and treatment history remain important. In atopic dermatitis, pruritus, sleep disturbance, and patient-reported outcomes cannot be inferred reliably from photographs alone. Acne and rosacea assessment may benefit from lesion counting, but lighting, cosmetic coverage, image resolution, and skin pigmentation can alter computational outputs.

Accordingly, automated visual scores should complement rather than replace validated clinical instruments and patient-centered assessment.

6. TELEDERMATOLOGY AND DIGITAL HEALTHCARE Teledermatology offers remote specialist assessment and follow-up. AI may assist image-quality screening, triage, referral prioritization, and standardized documentation. However, remote photographs may omit relevant anatomical or tactile information, and poor lighting or focus can compromise interpretation.

Clinical pathways should specify when in-person examination, dermoscopy, biopsy, or urgent referral is required. The safety of AI-assisted teledermatology should be demonstrated in prospective care settings, including measurement of diagnostic delay, inappropriate reassurance, and avoidable referrals.

7. LARGE LANGUAGE MODELS AND GENERATIVE AI Large language models (LLMs) primarily process and generate text, while multimodal variants may also analyze images. Dermatological applications include drafting clinical notes, summarizing records, preparing patient education

materials, and supporting scientific literature organization. Their fluent outputs can nevertheless contain incorrect medical statements, fabricated references, or omitted clinical details.

Clinical documentation generated with AI must be checked against the actual encounter and must not imply that examinations or procedures occurred when they did not. Patient-facing information should be reviewed for accuracy and must not be presented as a substitute for individualized medical assessment. In scientific publishing, authors retain responsibility for all factual statements and bibliographic references.

8. ALGORITHMIC BIAS AND GENERALIZABILITY Dermatological AI can be affected by unequal representation of skin phototypes, age groups, disease categories, imaging devices, and healthcare settings. Daneshjou and colleagues documented disparities in model performance on a diverse curated clinical image dataset [6]. Systematic appraisal of performance across darker skin phototypes has also highlighted the importance of equitable evaluation [7].

Dataset bias may arise through case selection, diagnostic labeling, and the overrepresentation of images from specialist centers. External validation should include independent institutions, heterogeneous populations, different devices, and realistic disease prevalence. Models may also degrade when deployed data differ from development data, a phenomenon known as distribution shift.

Clinical governance should therefore include subgroup reporting, post-deployment monitoring, and mechanisms to detect performance deterioration.

9. EXPLAINABLE ARTIFICIAL INTELLIGENCE Explainable AI attempts to clarify the features influencing a prediction. Saliency maps and related attribution methods can help clinicians inspect whether a model appears to attend to lesion regions. Yet a visually plausible explanation is not proof that the underlying prediction is correct or that the explanation faithfully represents the model's computation.

Explainability should be evaluated alongside diagnostic validity, calibration, reproducibility, and clinical safety. Transparency about intended use, training limitations, and uncertainty is at least as important as the availability of a visual explanation.

10. DIAGNOSTIC PERFORMANCE AND CLINICAL UTILITY Accuracy can be misleading when the prevalence of disease is low or classes are imbalanced. Sensitivity and specificity describe different aspects of classification, while

positive and negative predictive values vary with prevalence. The area under the receiver operating characteristic curve summarizes discrimination across thresholds but does not by itself demonstrate improved care.

Calibration describes agreement between predicted risks and observed outcomes. Clinical utility requires additional assessment of referral decisions, biopsy rates, diagnostic delay, adverse events, clinician workload, and patient experience. A model may demonstrate strong retrospective discrimination while producing little benefit-or even harm-when introduced into an unsuitable workflow.

Table 1. Principal dimensions of AI evaluation in dermatology.

DIMENSION	CLINICAL SIGNIFICANCE
Discrimination	Ability to distinguish diagnostic categories
Sensitivity	Detection of true positive cases
Specificity	Correct identification of negative cases
Calibration	Agreement between predicted risk and observed frequency
External validity	Performance in independent populations and settings
Fairness	Performance across relevant demographic and clinical subgroups
Clinical utility	Effect on decisions, outcomes, and resource use
Safety	Consequences of incorrect predictions and system failures

11. REGULATORY AND ETHICAL CONSIDERATIONS Clinical AI systems must be evaluated according to their intended medical purpose, applicable regulatory classification, and demonstrated safety. The United States Food and Drug Administration maintains information about AI-enabled medical devices [10]. Regulatory authorization for a defined indication should not be interpreted as proof of suitability for all dermatological settings.

In Europe, relevant requirements may include medical-device legislation, data-protection rules, and provisions of the European Artificial Intelligence Act. Applicable obligations depend on the system's intended purpose and regulatory status. Dermatological images may contain identifiable information, including facial features, scars, and tattoos; appropriate consent, lawful processing, access controls, and cybersecurity safeguards are therefore essential.

Human oversight must be substantive. Clinicians should understand intended use and known limitations and retain the ability to challenge or disregard algorithmic recommendations. Healthcare organizations and manufacturers also bear responsibilities for validation, maintenance, and safe deployment. International guidance emphasizes ethics and governance as central requirements for trustworthy health AI [9].

12. ARTIFICIAL INTELLIGENCE IN COSMETIC DERMATOLOGY Applications in aesthetic dermatology include facial image analysis, pigmentation mapping, wrinkle assessment, erythema quantification, and longitudinal monitoring of cosmetic outcomes. Standardized imaging may support more reproducible documentation, but apparent changes can be influenced by lighting, camera angle, facial expression, and skin hydration.

Commercial claims about AI-derived skin age, attractiveness, or treatment efficacy should not be equated with validated clinical measurements. Meaningful assessment requires defined endpoints, reproducible acquisition protocols, and independent validation.

13. ECONOMIC AND ORGANIZATIONAL IMPLICATIONS AI may support triage, documentation, image management, and specialist resource allocation. However, implementation entails acquisition, integration, training, cybersecurity, maintenance, regulatory compliance, and monitoring costs. High algorithmic accuracy does not necessarily establish cost-effectiveness.

An overly sensitive triage model could increase unnecessary referrals, while improved prioritization might shorten waiting times for high-risk lesions. Economic studies should consider complete diagnostic pathways, cost per appropriate referral or correctly identified malignancy, clinician time, and patient outcomes. Workflow compatibility and professional training are important determinants of actual value.

14. FUTURE PERSPECTIVES Multimodal systems may combine clinical photographs, dermoscopy, histopathology, patient histories, and structured records to support more contextualized decisions. Longitudinal models may eventually assist disease monitoring and prediction of treatment response, but require high-quality prospective datasets.

Federated learning may facilitate collaborative model development across institutions without centralizing all patient data, although it does not eliminate privacy or security risks. The next research priority should be pragmatic, multicenter prospective evaluation of AI-assisted care compared with appropriate standard-care pathways, with explicit measurement of patient outcomes, fairness, and operational consequences.

15. DISCUSSION The principal contribution of dermatological AI research has been to demonstrate that computational systems can recognize diagnostically informative visual patterns [1-5]. Nevertheless, reliable clinical implementation requires evidence beyond isolated image classification. Generalizability across patient populations, skin phototypes, and care settings remains a central challenge [6,7].

Four interdependent dimensions can organize evaluation: diagnostic validity, clinical utility, equity and safety, and organizational sustainability. Diagnostic validity includes discrimination, calibration, and external reproducibility. Clinical utility concerns improvements in real decisions and outcomes. Equity and safety require attention to underrepresented groups and the consequences of errors. Organizational sustainability concerns the integration of technology into professional workflows at acceptable cost.

These dimensions should be considered jointly. An accurate model that is poorly calibrated, inequitable, or difficult to integrate may not improve patient care. Conversely, carefully designed clinician-AI collaboration may be valuable when its benefits are prospectively demonstrated.

16. LIMITATIONS This narrative review is selective rather than exhaustive and does not present a systematic search strategy, formal risk-of-bias assessment, or meta-analysis. Included research spans heterogeneous image sources, diagnostic categories, reference standards, and study designs. Comparisons across studies should therefore be interpreted cautiously. Many available evaluations are retrospective and cannot establish the effects of AI on patient outcomes.

17. CONCLUSIONS Artificial intelligence offers important opportunities for dermatological image interpretation, diagnostic support, disease monitoring, and healthcare organization. Its promise is supported by foundational research on skin-lesion classification and human-computer collaboration, but experimental performance must not be confused with clinical effectiveness.

Responsible adoption requires representative training and evaluation data, independent validation, calibrated predictions, prospective clinical studies, regulatory compliance, privacy safeguards, and continuous monitoring. AI should augment dermatological expertise and support patient-centered decisions, not replace professional judgment. The long-term value of these technologies will depend on their demonstrated capacity to improve safety, equity, clinical outcomes, and sustainable healthcare delivery.

REFERENCES

1. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, Thrun S. Dermatologist-level classification of skin cancer with deep neural networks. Nature. 2017;542:115-118. doi:10.1038/nature21056.

2. Haenssle HA, Fink C, Schneiderbauer R, et al. Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. Ann Oncol. 2018;29(8):1836-1842. doi:10.1093/annonc/mdy166.

3. Haenssle HA, Fink C, Toberer F, et al. Man against machine reloaded: performance of a market-approved convolutional neural network in classifying a broad spectrum of skin lesions in comparison with 96 dermatologists working under less artificial conditions. Ann Oncol. 2020;31(1):137-143. doi:10.1016/j.annonc.2019.10.013.

4. Tschandl P, Rinner C, Apalla Z, et al. Human-computer collaboration for skin cancer recognition. Nat Med. 2020;26:1229-1234. doi:10.1038/s41591-020-0942-0.

5. Liu Y, Jain A, Eng C, et al. A deep learning system for differential diagnosis of skin diseases. Nat Med. 2020;26:900-908. doi:10.1038/s41591-020-0842-3.

6. Daneshjou R, Vodrahalli K, Novoa RA, et al. Disparities in dermatology AI performance on a diverse, curated clinical image set. Sci Adv. 2022;8(32):eabq6147. doi:10.1126/sciadv.abq6147.

7. Aggarwal P. Performance of artificial intelligence imaging models in detecting dermatological manifestations in higher Fitzpatrick skin color classifications: a systematic review. J Am Acad Dermatol. 2022;87(4):819-823. doi:10.1016/j.jaad.2021.05.078.

8. Han SS, Park GH, Lim W, et al. Deep neural networks show an equivalent and often superior performance to dermatologists in onychomycosis diagnosis: automatic construction of onychomycosis datasets by region-based convolutional deep neural network. PLoS One. 2018;13(1):e0191493. doi:10.1371/journal.pone.0191493.

9. World Health Organization. Ethics and Governance of Artificial Intelligence for Health: WHO Guidance. Geneva: WHO; 2021. https://www.who.int/publications/i/item/9789240029200.

10. US Food and Drug Administration. Artificial Intelligence-Enabled Medical Devices. https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-enabled-medical-devices.

DECLARATIONS Author contributions: Torello Lotti: clinical and dermatological conceptualization, medical interpretation, and scientific review. Milo Dolci: interdisciplinary conceptualization, technological innovation, healthcare organization, economic implications, and manuscript development. Contributions are subject to author confirmation.

Funding and conflicts of interest: Disclosed in separate author forms supplied to the journal.

Ethics approval: Not applicable; this narrative review reports no original research involving human participants or animals.

Data availability: No original patient-level dataset was generated.