Global Health & Medicine. 2026; 8(3):193-204.
MINI-REVIEW
DOI: 10.35772/ghm.2026.01058

Artificial intelligence (AI)-assisted diagnosis of skin diseases: From image classification to dermatology-specific multimodal clinical reasoning

Yuhan Cheng[1,2,§], Chu Zhou[3,§], Ping Wang[3], Huanran Liu[4], Yue Han[3,*]

[1] School of Medicine, Tongji University, Shanghai, China;
[2] Department of Nursing, Shanghai Tenth People's Hospital, School of Medicine, Tongji University, Shanghai, China;
[3] Department of Dermatology, The Union Hospital, Fujian Medical University, Fuzhou, Fujian, China;
[4] Graduate School of Engineering, Tamagawa University, Machida, Tokyo, Japan.

Abstract: Artificial intelligence (AI) in dermatology has moved beyond the early paradigm of single-image classification. Dermatological diagnosis is achieved based on morphology, distribution, symptoms, tactile findings, temporal evolution, patient history, histopathology, and treatment response. Clinically important differentials, such as eczema versus psoriasis, cutaneous T-cell lymphoma versus chronic dermatitis, drug eruption versus viral exanthem, lupus erythematosus versus dermatomyositis, and melanoma versus atypical nevus, are rarely resolved with one photograph alone. This review therefore frames AI-assisted dermatology around a central argument: the field must progress from lesion recognition to dermatology-specific multimodal clinical reasoning. We summarize major advances in convolutional neural networks, dermoscopic benchmarks, clinical-image datasets, large language models, vision-language systems, and dermatology foundation models. We also analyze challenges that are particularly relevant to dermatology, including morphologic overlap, skin-tone bias, reduced erythema visibility on darker skin, dataset imbalance, variable smartphone imaging, imperfect reference standards, and the gap between benchmark performance and clinical deployment. Special attention is given to fairness, regulatory oversight, software as a medical device, human-AI collaboration, prognosis prediction, biologic-response modeling, longitudinal monitoring, and treatment optimization. Finally, we discuss future directions, including skin-tone-aware foundation models, lesion-level and body-site grounding, pathology-genomics integration, dermatology copilots, post-marketing surveillance, and prospective clinical trials. By prioritizing dermatological reasoning rather than generic AI architecture, this review outlines a clinically grounded pathway for building safe, interpretable, equitable, and useful AI systems for skin disease management.

***Keywords*:** dermatology, artificial intelligence, multimodal reasoning, foundation models

1. Introduction

Skin diseases are among the most common human disorders and encompass inflammatory, infectious, neoplastic, autoimmune, genetic, and drug-related conditions. Dermatology is also an unusually visible specialty: patients, primary care clinicians, dermatologists, pathologists, and digital platforms can all observe the same organ, although at different levels of resolution and with different amounts of context. This visibility has made dermatology an early adopter of computer vision, but it has also encouraged the misleading view that diagnosis is primarily an image classification task. In clinical practice, dermatologists rarely make decisions based on an isolated image. They integrate lesion morphology with body site, distribution, skin tone, symptoms, palpation, tempo of evolution, medication exposure, comorbidities, occupational and environmental triggers, dermoscopy, histopathology, microbiology, serology, and the response to treatment (*1,2*).

The first generation of dermatology artificial intelligence (AI) relied largely on convolutional neural networks (CNNs) trained on clinical or dermoscopic images. Landmark work reported dermatologist-level or near-dermatologist-level performance in selected skin cancer tasks, and public datasets such as HAM10000 and ISIC accelerated algorithm benchmarking (*3-9*). These studies were important because they showed that visual pattern recognition could be scaled, standardized, and compared across groups. At the same time, they revealed

a substantial dataset-to-clinic gap: algorithms usually performed best on curated pigmented-lesion images and often decreased in reliability when confronted with different devices, image quality, disease spectra, clinical settings, and patient populations. This gap is particularly vast in dermatology, where real-world photographs vary in lighting, distance, compression, hair, scale, ulceration, cosmetics, anatomical site, and background pigmentation.

A second generation of systems is now emerging. Large language models (LLMs) and multimodal vision-language models can process patient narratives, electronic health records, pathology reports, and clinical instructions along with images. General medical foundation models, pathology copilots, and dermatology-specific multimodal models offer a route to align morphology with clinical text, histology, and longitudinal data (*10-15*). The goal is not simply to improve top-1 accuracy. Clinically useful systems should generate a ranked differential diagnosis, explain which visual and historical features substantiate each option, indicate when biopsy or referral is needed, express uncertainty, and maintain performance across skin tones and care settings.

This review departs from a generic 'AI in medicine' framework by asking a dermatology-specific question: what must an AI system understand to reason in a clinically dermatological manner? We propose four requirements. First, the system should encode morphology and distribution rather than disease labels alone. Second, it must account for skin color and optical visibility because erythema, scale, pigment network, and vascular structures are not equally visible across skin tones. Third, it should model time, symptoms, and treatment response since many inflammatory and lymphoproliferative diseases declare themselves longitudinally. Fourth, it must be embedded in a supervised workflow with regulatory, ethical, and fairness safeguards.

Dermatology is fundamentally a multimodal specialty, making image-only AI intrinsically incomplete for many inflammatory, lymphoproliferative, drug-related, and longitudinal disorders. The most clinically useful question is therefore not whether more modalities always improve accuracy but which diagnostic decisions are impossible, unsafe, or poorly calibrated without history, distribution, time, pathology, and treatment response.

2. Why dermatology is uniquely amenable to and uniquely difficult for AI

2.1. Morphologic overlap is the central clinical problem

The core challenge for dermatology AI is not merely whether a CNN can detect a border or a pigment network but whether a model can separate diseases that share the same visible vocabulary. Erythematous scaly plaques may represent psoriasis, chronic eczema, tinea corporis, cutaneous lupus erythematosus, pityriasis rubra pilaris, or early mycosis fungoides. Vesicles may indicate allergic contact dermatitis, herpesvirus infection, dyshidrotic eczema, bullous pemphigoid, or a drug eruption depending on distribution, age, symptoms, mucosal involvement, and chronology. A cropped trunk lesion can be photographed and classified, but dermatologists often make a diagnosis by comparing multiple lesions and asking whether they are symmetrical, grouped, photo-distributed, acral, flexural, follicular, dermatomal, annular, targetoid, retiform, or livedoid.

This morphologic overlap helps explain why dermatology AI may involve more difficulty than many standardized imaging tasks. Radiologic images usually follow consistent acquisition protocols and utilize consistent anatomical planes, whereas dermatology photographs may be taken by patients, nurses, primary care physicians, or specialists using different devices, angles, distances, and lighting conditions. The relevant signal may include background skin, hair-bearing status, nail or mucosal involvement, scale texture, or even the absence of a finding. In addition, the diagnostic reference standard is often mixed: some diseases are diagnosed clinically, some require clinicopathologic correlation, and some become evident only after follow-up or treatment failure. AI systems therefore need to learn a diagnostic hierarchy that includes lesion type, color, scale, border, configuration, distribution, temporal behavior, and clinicopathologic consistency. Table 1 summarizes key dermatology-specific problems and the multimodal capabilities required to address them.

2.2. Skin tone is not a secondary fairness issue; it changes the visual signal

Skin-tone bias is one of the defining challenges in dermatology AI. Many training datasets overrepresent lighter skin, while darker skin types, acral lesions, mucosal disease, and conditions common in underserved populations remain comparatively scarce. This imbalance is not only statistical; it changes the visual basis of diagnosis. Erythema may appear bright red on lightly pigmented skin but violaceous, gray, brown, or subtle on deeply pigmented skin. Post-inflammatory hyperpigmentation may mask active inflammation, and pigment networks or vascular patterns may differ in contrast. Psoriasis, atopic dermatitis, lupus erythematosus, dermatomyositis, and drug eruptions can therefore present with skin-tone-dependent cues that a narrow model may not have learned.

Fairness studies have shown that dermatology AI can perform worse on datasets enriched for darker skin tones and uncommon diseases and that post-hoc reweighting cannot fully compensate for models trained on narrow data (*16,17*). Fitzpatrick skin type labels are useful but incomplete because they describe ultraviolet response rather than the full range of skin color, hue, undertone,

Table 1. Dermatology-specific clinical problems that multimodal AI should address

Clinical problem	Typical examples	Why image-only AI is insufficient	Desired multimodal capability
Morphologic overlap	Eczema *vs.* psoriasis, CTCL *vs.* chronic dermatitis, lupus *vs.* dermatomyositis, drug eruption *vs.* viral exanthem	Single images may show the same erythematous plaques, scale, vesicles, or annular patterns without enough context.	Modeling primary and secondary lesion morphology, symptoms, distribution, chronicity, medications, pathology, and treatment response.
Distribution and body-site logic	Extensor psoriasis, flexural eczema, photo-distributed lupus, dermatomal zoster, acral melanoma, nail disease	Cropped lesion photographs remove the body map and may hide symmetry, clustering, or site-specific clues.	Integrating close-up images, distant distribution photos, body maps, anatomical metadata, and prior visits.
Temporal evolution	Changing nevus, flare-remitting eczema, treatment-resistant CTCL, delayed drug eruption	A one-time image cannot show growth, recurrence, latency, dechallenge, or response to therapy.	Using longitudinal images, symptom diaries, medication timelines, and disease trajectory models.
Tactile and bedside signs	Induration, warmth, tenderness, blanching, Nikolsky's sign, scale texture, edema	The model cannot palpate or perform bedside procedures based on pixels alone.	Prompting clinicians or patients for structured signs and flag tests needed before decision-making.
Skin tone and optical visibility	Subtle erythema on darker skin, post-inflammatory hyperpigmentation, low-contrast vascular features	Color cues learned from light-skin datasets may not transfer and may amplify racial disparities.	Reporting skin-tone-stratified performance, calibrating color, enriching darker skin datasets, and using fairness audits.
Weak or mixed gold standards	Clinical diagnosis without biopsy, clinicopathologic discordance, evolving diagnoses	A single disease label can be noisy or misleading.	Using consensus labels, uncertainty-aware training, pathology linkage, and follow-up outcome labels.

Abbreviations: AI, artificial intelligence; CTCL, cutaneous T-cell lymphoma.

and imaging contrast. Future dermatology AI should report performance by skin tone, race and ethnicity when ethically collected, age, sex, anatomical site, image source, and disease prevalence. It should also test erythema-dependent conditions explicitly. Equity must be designed into data collection, annotation, model training, external validation, and post-deployment monitoring rather than added after model development.

2.3. Dermatological diagnosis is longitudinal and often therapeutic

Time is diagnostically important in dermatology. Melanoma screening depends on changes in size, shape, color, and structure. Psoriasis and atopic dermatitis fluctuate with infection, stress, season, adherence, and treatment. Cutaneous T-cell lymphoma (CTCL) may mimic eczema for years before the clinicopathologic pattern reaches a diagnostic threshold. Drug eruptions require attention to medication chronology, latency, dechallenge, rechallenge, and systemic features. Models that rely on a single image at a given time therefore discard information that dermatologists routinely consider essential.

The clinical horizon of dermatology AI is also expanding beyond diagnosis. In chronic inflammatory diseases, the central question may be whether a patient will respond to a biologic, have a flare-up, experience treatment toxicity, or need to be escalated. In melanoma, AI may help integrate histology, molecular markers, and clinical staging for risk assessment. In wounds and ulcers, the task may be to quantify healing trajectory and detect infection or ischemia. These use cases require multimodal and longitudinal modeling rather than isolated image classification (*18-20*).

3. From image classification to multimodal clinical reasoning

3.1. What image-based AI has achieved and where it remains limited

CNNs, EfficientNet, ResNet, Vision Transformers, and ensemble methods have performed well in selected dermoscopic and clinical-image tasks. Early studies demonstrated that deep networks could classify skin cancers at a level comparable to dermatologists under experimental conditions, and ISIC-style challenges created shared benchmarks for lesion segmentation, attribute detection, and classification (*3-9,21-23*). These studies remain foundational because they made dermatology one of the most visible specialties in medical computer vision.

Nevertheless, high area under the curve (AUC) values in curated datasets should not be mistaken for clinical readiness. A pigmented-lesion classifier trained on dermoscopy cannot necessarily diagnose inflammatory rashes, infections, vasculitis, connective tissue disease, or genodermatoses. A model validated on biopsy-confirmed lesions may perform poorly on low-quality smartphone images from primary care. Likewise, a melanoma-versus-nevus classifier may still fail to recommend biopsy when the lesion is amelanotic, the image is incomplete, or the history is concerning. Image AI should therefore be

treated as one component of clinical reasoning and not as the reasoning process itself.

3.2. LLMs may be especially important in dermatology because diagnosis is narrative

LLMs are often described as documentation tools, but in dermatology their deeper value may lie in narrative reasoning. Dermatologists transform a history into diagnostic constraints: acute or chronic, localized or generalized, itchy or painful, febrile or afebrile, drug-associated or spontaneous, recurrent episodes or an initial episode, and treatment-responsive or refractory. LLMs can extract these constraints from consultation notes, referral letters, pathology reports, patient-submitted descriptions, and teledermatology intake forms. They can also translate patient language into dermatological terminology; for example, 'spreading red itchy patches after antibiotics' can be mapped to a differential that includes morbilliform drug eruption, viral exanthem, urticaria, and an early severe cutaneous adverse reaction.

LLMs also create safety risks. They may generate unsupported diagnoses, overstate certainty, suggest an unsafe treatment, or miss emergencies such as Stevens-Johnson syndrome/toxic epidermal necrolysis, meningococcemia, necrotizing infection, or rapidly progressive melanoma. Dermatology-specific deployment therefore requires retrieval from curated guidelines, awareness of local formularies, calibrated uncertainty, escalation rules, and clinician review. A dermatology copilot should not merely provide an answer; it should show which visual signs, historical features, and evidence sources substantiate its differential diagnosis (*12,19*).

3.3. Vision-language and foundation models: Opportunities for dermatology-specific grounding

Vision-language models align images and text in a shared representation space and may enable open-vocabulary recognition, image-text retrieval, and explanations in dermatological terms. General medical foundation models and multimodal systems show how images, text, structured variables, and long clinical records can be interpreted through a unified interface (*10-15*). This architecture is attractive in dermatology because the disease label alone is rarely sufficient. A system should understand that an 'annular scaly plaque with central clearing on the trunk' implies a different differential from a "well-demarcated extensor plaque with silvery scale," even if both appear as erythematous plaques.

Segmentation and multimodal foundation models may aid in lesion boundary detection, body-site mapping, ulcer area monitoring, total-body photography, and clinicopathologic alignment. Dermatology-specific models trained on large collections of clinical photographs, dermoscopy, histopathology, and weakly labeled reports are especially promising (*15*). The next step is not only larger pretraining but clinically grounded pretraining. Models should learn relationships between morphological terms and image regions, lesion distribution and body maps, histopathologic patterns and clinical phenotypes, and longitudinal changes and treatment response.

Foundation models may change dermatology more profoundly than classic CNNs because they are better aligned with the three characteristics of the specialty: label ambiguity, an open vocabulary, and morphology-language coupling. A CNN can learn that a lesion resembles training examples labeled psoriasis, but a vision-language model can represent the phrase "well-demarcated erythematous plaques with silvery scale on extensor surfaces" and relate it to competing diagnoses, body-site logic, and clinical history. This matters because dermatological labels are often provisional, overlapping, or refined after biopsy and follow-up rather than fixed at the moment of imaging (*10,14,15*).

Dermatology may therefore be one of the medical specialties most naturally amenable to multimodal foundation models. Its diagnostic vocabulary is visual but not purely visual: morphology is described in language, distribution is mapped across the body, histopathology provides tissue-level confirmation, and chronic diseases evolve across visits. The field should judge foundation models not only by top-1 disease classification but by whether they can retrieve visually similar cases, ground morphological terms in image regions, recognize missing clinical context, and express uncertainty when a single photograph cannot lead to a safe diagnosis.

3.4. Multimodal fusion should follow the dermatologist's workflow

Early fusion, late fusion, cross-modal attention, contrastive pretraining, and retrieval-augmented generation are useful engineering strategies, but their value depends on whether they reproduce dermatological reasoning. A practical multimodal pipeline should assess image quality and skin tone, segment relevant lesions, identify morphology and distribution, extract a structured history from text, retrieve guidelines or prior images when appropriate, generate a ranked differential diagnosis, calibrate uncertainty, and recommend triage or next steps only when the available information is sufficient.

This workflow differs from generic medical AI because the dermatological examination is partly visual, partly tactile, and partly historical. A model cannot palpate an induration or gauge warmth, tenderness, or blanching, but it can ask the clinician or patient to provide those findings. It cannot directly perform bedside procedures, but it can flag which tests are needed before making a decision. Multimodal dermatology AI should therefore be interactive: it should improve

data acquisition and clinical questioning rather than simply classify what has already been uploaded. The dermatological multimodal reasoning architecture for AI-assisted skin disease diagnosis is shown in Figure 1.

3.5. Which dermatological tasks truly require multimodal AI?

A critical weakness of early dermatology AI was the implicit assumption that every skin task should be solved as an image classification problem. The opposite error would be to assume that every task requires maximal multimodalities. Multimodal AI is valuable when additional context changes the clinical decision, reduces unsafe uncertainty, or distinguishes diseases with overlapping morphology. It is less necessary when the target is a narrowly defined visual triage problem under standardized acquisition conditions.

Melanoma screening illustrates this distinction. For many dermoscopic images of pigmented lesions, visual structure represents a large fraction of the diagnostic signal, and image-based models can provide useful risk stratification. And yet the same model is incomplete for amelanotic lesions, lesions selected from high-risk total-body photography, changing nevi, immunosuppressed patients, or cases in which the relevant clinical decision is biopsy versus short-interval monitoring. In these settings, prior images, lesion evolution, age, site, personal and family history, and a clinician's suspicions become part of the diagnostic signal (*3-9,21-23*).

Inflammatory, drug-related, infectious, and lymphoproliferative dermatoses are different. A photograph of an erythematous scaly plaque may be consistent with eczema, psoriasis, tinea, lupus, dermatomyositis, or early CTCL. The decisive information may be itching, pain, a fever, the timing of medication, mucosal involvement,

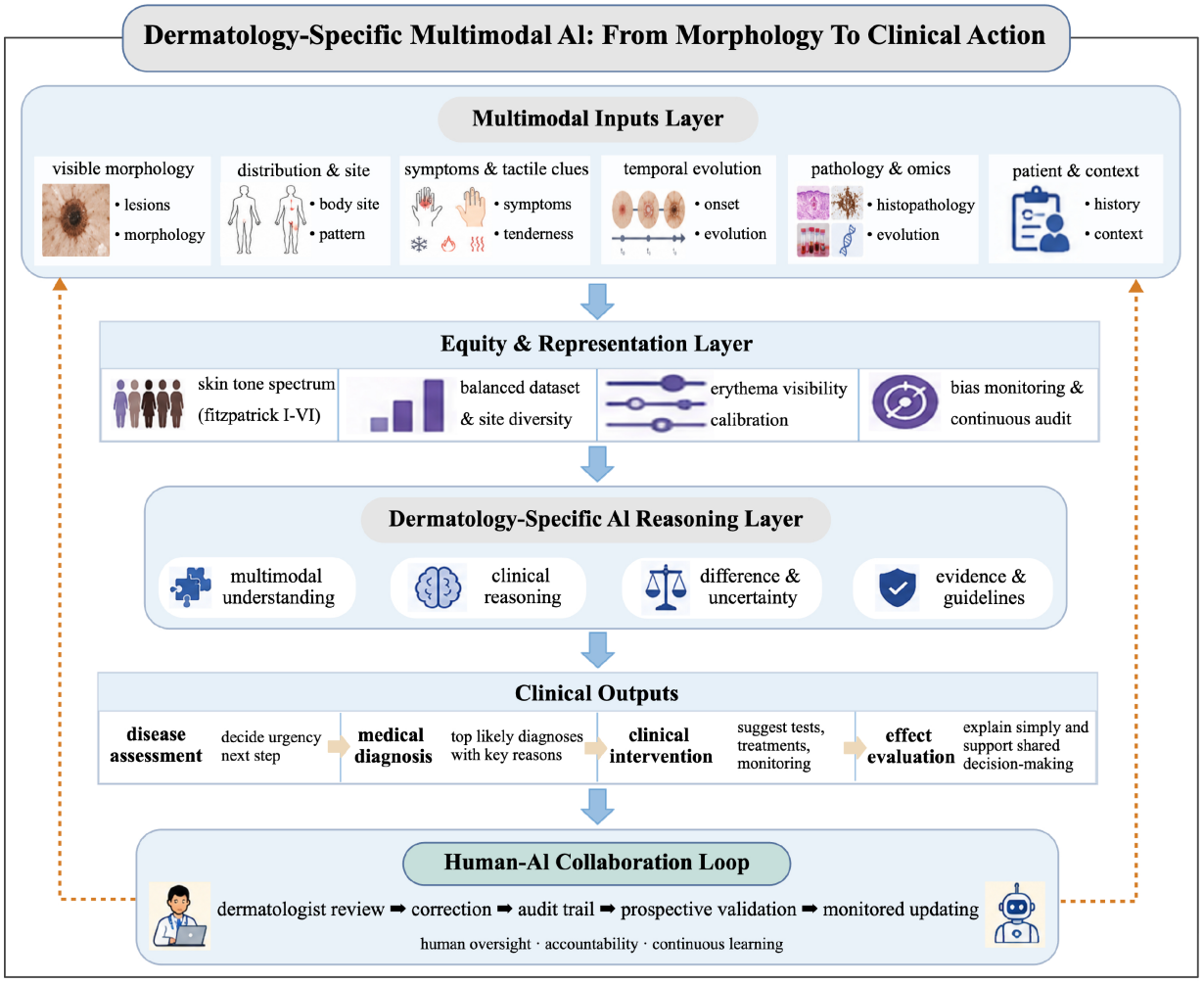

Figure 1. Dermatological multimodal reasoning architecture for AI-assisted skin disease diagnosis. The schematic illustrates a clinically grounded pipeline that integrates multiple data modalities to aid in dermatological decision-making. Inputs include clinical images, dermoscopy, patient history, symptoms, medication timelines, prior images, and histopathology reports. The core reasoning engine performs lesion segmentation, morphology and distribution analysis, detection of temporal changes, and clinicopathologic alignment using vision-language and foundation models. Outputs consist of a ranked differential diagnosis with uncertainty calibration, visual and textual explanations linked to dermatological signs, triage recommendations, and suggestions for missing information. The architecture emphasizes interactive data acquisition, clinician oversight, and fairness across skin tones.

distribution, chronicity, histology, immunophenotype, laboratory findings, or treatment failure. For these tasks, image-only AI is not merely incomplete; it can yield methodologically unsound results because the required clinical variables are absent from the input.

4. Dermatology-specific bottlenecks and challenges

4.1. Data quality and annotation: Labels must depict morphology and not just diagnosis

Many dermatology AI datasets provide disease labels, but dermatologists reason with intermediate descriptors. A label such as 'psoriasis' is less informative than a structured description such as 'well-demarcated erythematous plaque, silvery scale, extensor surface, chronic recurrent course, family history, and no fungal hyphae.' Similarly, a melanoma dataset becomes more clinically useful when it includes body site, lesion size, evolution, dermoscopic structures, histopathology, and whether the lesion was identified during high-risk surveillance or routine care. Disease-only labels may encourage models to learn shortcuts, including device type, background skin, ruler marks, biopsy ink, or center-specific artifacts.

Annotation should therefore move from single labels towards structured dermatological ontologies that include primary lesion type, secondary changes, arrangement, distribution, body site, symptom profile, skin tone, image quality, and diagnostic certainty. Consensus annotation, adjudication of difficult cases, and uncertainty-aware labels are essential because dermatology reference standards are often imperfect. For rare diseases, sharing of registry-based data, active learning, and expert-in-the-loop annotation may be more useful than indiscriminate data scaling.

4.2. Learning of hidden shortcuts by dermatology AI

Dermatology AI can fail even when retrospective accuracy appears high because models may learn features that correlate with labels but do not represent disease biology. Common hidden shortcuts include ruler marks placed beside suspicious lesions, skin-marker ink, biopsy-site framing, surgical drapes, dermoscopy device signatures, clinic-specific backgrounds, compression artifacts, hair removal patterns, and anatomical-site or referral-center cues. A model can therefore appear to detect melanoma while partly detecting the clinical behavior that preceded biopsy or photography.

These failures are especially dangerous because dermatology images are often collected after a clinician has already identified a lesion as concerning. The dataset may encode selection bias, workup bias, and spectrum bias: benign lesions photographed casually, malignant lesions photographed with rulers and dermoscopy, and inflammatory rashes photographed only after referral. If these cues are not controlled, external validation may reveal an abrupt drop in performance in primary care, teledermatology, darker skin tones, or limited-resource settings (*16,17,19,23-25*).

Robust development should therefore include shortcut stress tests. Investigators should perform metadata ablation, background masking, lesion-only versus context-aware comparisons, device-stratified validation, counterfactual removal of rulers and markers, compression robustness testing, and prospective evaluation in the intended workflow. Explanations should be audited for whether they highlight true dermatological signs rather than artifacts. This failure analysis is as important as reporting the aggregate AUC.

4.3. Real-world deployment

Many AI systems perform well on images acquired under standardized conditions but face a different reality in teledermatology and primary care. Patient-submitted photographs may be blurred, underexposed, overcompressed, too close, too far, or taken after the application of emollients, makeup, antiseptics, topical corticosteroids, or dressings. Hair, tattoos, scale, crust, ulceration, nail polish, and anatomical curvature can obscure lesions. Smartphone color calibration and automatic white balance can alter erythema and pigmentation. These factors are not merely technical noise; they directly change the diagnostic signal available to both a clinician and a model (*24,25*).

Deployment-ready systems should provide image-quality feedback, standardized acquisition instructions, color calibration or reference cards when feasible, validation across devices, and the ability to safely reject inadequate images. They should also be tested in common teledermatology scenarios, including multiple-lesion uploads, rashes involving several body sites, darker skin tones, pediatric images, nail and scalp disease, genital or mucosal lesions requiring careful consent, and follow-up photographs taken at different distances or under different lighting conditions.

4.4. Bias, global inequity, and dataset colonialism

Global dermatology AI should not be created based solely on images collected at wealthy, urban, specialist centers and then exported to settings with different diseases, skin tones, pathogens, occupational exposure, climate, access to care, and treatment pathways. Such a pattern, sometimes discussed as dataset colonialism, risks turning underrepresented populations into data sources or target markets without giving them governance authority, clinical benefit, or locally valid tools. It may also miss conditions common in limited-resource settings, including neglected tropical diseases, pigmentary disorders, leprosy, scabies, deep fungal infections, Kaposi sarcoma, and HIV-associated dermatoses.

Equitable dermatology AI requires locally meaningful datasets, partnerships between communities and clinicians, transparent consent, shared governance, subgroup performance reporting, and mechanisms for sharing benefits. Fairness should be assessed not only with an equal AUC but also based on clinically meaningful outcomes: missed malignancies, delayed diagnosis, unnecessary biopsies, referral burden, treatment access, patient trust, and whether the system improves care for groups historically underserved by dermatology (*16,17*).

4.5. Interpretability must be dermatology-specific

Generic heatmaps are insufficient if they simply highlight a lesion without explaining the relevant signs. Dermatologists need explanations that map to clinical language, such as an atypical pigment network, blue-white veil, arborizing vessels, Wickham striae, collarette scale, follicular plugging, Gottron papules, a targetoid morphology, retiform purpura, or palmoplantar involvement. A useful explanation should connect a visible region to a morphology term, the term to a differential diagnosis, and the differential diagnosis to a concrete next step.

Dual-modality explanations are therefore preferable: visual evidence should be paired with a textual rationale and calibrated uncertainty. A system might rank psoriasis as the leading diagnosis because it identifies sharply demarcated scaly plaques on extensor surfaces and a chronic recurrent history while also retaining eczema as a possibility because of pruritus and flexural involvement. Such explanations allow clinicians to challenge the model, identify missing information, and reduce automation bias.

4.6. Regulation and medico-legal responsibility

Clinical dermatology AI may qualify as software as a medical device when it provides diagnostic or triage recommendations. Regulatory evaluation should consider the intended use, target population, user type, autonomy level, risk of harm, and whether the model is locked or adaptive. Earlier devices such as MelaFind demonstrated the feasibility of computerized lesion assessment but also underscored problems related to specificity, workflow fit, and clinical utility; newer primary care systems have renewed interest in regulated AI-assisted skin cancer triage (*26,27*).

Modern multimodal and adaptive systems require representative validation, skin-tone subgroup analysis, human-factor testing, cybersecurity review, data governance, change-control planning, and post-marketing performance monitoring (*28-30*). Responsibility must also be explicit. When an AI output conflicts with a clinician's judgment, the medical record should show who reviewed the output, what evidence was considered, and why the final decision was made. High-risk recommendations should require clinician confirmation and should be linked to visible evidence and current guidance.

5. Future breakthroughs and technical pathways

5.1. Dermatology foundation models should be morphology-aware, skin-tone-aware, and longitudinal

The next generation of dermatology foundation models should not be evaluated only in terms of disease classification accuracy. They should be tested on morphology recognition, body-site grounding, lesion segmentation, image-quality assessment, skin-tone fairness, rare-disease retrieval, detection of temporal changes, and clinicopathologic correlation. Self-supervised and weakly supervised learning can capitalize on large image archives, but clinically meaningful pretraining should pair images with morphology-rich reports, dermoscopic structures, pathology captions, distribution maps, and longitudinal treatment data. Table 2 summarizes landmark evidence and emerging model families.

Longitudinal modeling is especially important. Total-body photography, sequential dermoscopy, patient-reported flare-up diaries, laboratory trends, and treatment changes can all be represented as trajectories. Such models could identify changing melanocytic lesions, quantify psoriasis severity over time, detect atopic dermatitis flare-ups, monitor wound healing, and recognize chronic dermatitis cases that warrant a biopsy for cutaneous lymphoma. They should also report uncertainty and clearly distinguish diagnostic prediction, prognosis, and estimation of the treatment response.

5.2. Clinicopathologic foundation models

The most consequential frontier for dermatology foundation models may be clinicopathologic integration. Dermatologists rarely treat clinical photographs and pathology slides as independent evidence streams; instead, they ask whether the clinical morphology, dermoscopic structures, histopathologic pattern, immunophenotype, molecular alterations, and disease course support the same diagnosis. AI systems should be designed to model this alignment rather than simply concatenate modalities.

Melanoma, CTCL, autoimmune connective-tissue disease, vasculitis, blistering disorders, and complex drug reactions are particularly amenable to this approach. In melanoma, clinical and dermoscopic images can be connected with histologic subtype, Breslow thickness, ulceration, mitotic rate, genomics, and outcomes. In CTCL, persistent patches or plaques, serial photographs, repeated biopsies, T-cell receptor clonality, and treatment response form a longitudinal clinicopathologic pattern. In autoimmune disease, morphology and distribution must

Table 2. Landmark evidence and emerging model families in dermatology AI

Area or model family	Representative evidence or dataset	Dermatology-specific contribution	Remaining gap
Classic skin cancer CNNs	Esteva *et al*. (*3*), Haenssle *et al.* (*4*), Tschandl *et al.* (*5*): human-computer studies.	Established dermatologist-level performance in selected dermoscopic or clinical image tasks.	Often have a narrow disease spectrum, curated images, and limited real-world workflow evaluation.
Public dermoscopy benchmarks	ISIC challenges, HAM10000, BCN20000.	Enabled reproducible lesion segmentation and pigmented-lesion classification.	Underrepresentation of darker skin, non-pigmented lesions, rare disorders, and primary care images.
Clinical image + history systems	Liu *et al.* (*6*): deep learning system for 26 common skin diseases; PAD-UFES-20 smartphone dataset.	Demonstrated that history and metadata improve differential diagnosis beyond lesion pixels.	Require prospective triage evaluation and better handling of incomplete patient contexts.
Broad dermatology classifiers	Han *et al.* (*7*): 134 skin disorders, mobile and few-shot platforms.	Expanded AI from melanoma to inflammatory, infectious, and benign conditions.	May still rely on disease labels rather than morphology and distribution annotations.
Fairness datasets and audits	DDI, Fitzpatrick17k, skin-tone subgroup analysis.	Exhibited clinically significant performance degradation on darker skin tones and uncommon diseases.	Require global, prospective, skin-tone-aware validation and local governance.
Vision-language models	CLIP, BiomedCLIP, GPT-4V dermatology prompting.	Enabled open-vocabulary matching, image-text retrieval, and morphology-grounded explanations.	Prone to prompt sensitivity, hallucinations, and weak spatial grounding without clinical guardrails.
Medical generalist models	Med-PaLM M, Med-Gemini, multimodal pathology copilots.	Supported flexible reasoning over text, images, records, pathology, and long contexts.	Dermatology-specific performance, safety, and fairness remain insufficiently validated.
Dermatology foundation models	PanDerm-like multimodal dermatology pretraining.	Can learn transferable representations across clinical, dermoscopic, and histopathologic modalities.	Need prospective external validation, transparent failure analysis, and equitable deployment.
Regulated or near-regulated systems	MelaFind, DermaSensor, Skin Analytics-type triage tools.	Moved AI from benchmark datasets towards clinical pathways and device regulation.	Need evidence for specificity, workflow impact, liability, and post-marketing monitoring.
Clinicopathologic foundation models	Multimodal pathology copilots, dermatology foundation models, and paired clinicopathologic datasets.	Connected clinical morphology, dermoscopy, histopathology, immunophenotype, genomics, transcriptomics, and outcomes into one reasoning space.	Require lesion-to-biopsy linkage, tissue-level grounding, temporal alignment, external validation, and explicit handling of discordant evidence.

Abbreviations: AI, artificial intelligence; CNNs, convolutional neural networks; DDI, Diverse Dermatology Images; ISIC, International Skin Imaging Collaboration.

be aligned with serology, histology, immunofluorescence, and transcriptomic signatures (*14,15,18*).

Clinicopathologic foundation models would require paired and temporally linked datasets, tissue-level grounding, lesion-level mapping between the clinical image and biopsy site, and outcome labels that distinguish diagnosis, prognosis, and the therapy response. Their value should be judged by whether they detect discordance, recommend a repeat biopsy when morphology and histology conflict, and support cautious reasoning in instances of diseases where a single gold standard is unrealistic.

5.3. Agentic dermatology AI: A copilot, not an autonomous dermatologist

Agentic AI refers to systems that can plan actions, call tools, retrieve information, ask follow-up questions, and coordinate workflow steps. In dermatology, a safe agent might check image quality, request close-up and distant distribution photographs, ask about pain, itching, a fever, medication exposure, mucosal involvement, pregnancy, immunosuppression, and duration, retrieve prior images, generate a differential diagnosis, recommend whether urgent referral is needed, and draft a structured note for clinician review.

Such systems should be designed as copilots rather than autonomous dermatologists. They require escalation rules for red-flag conditions, refusal behavior when data are inadequate, transparent uncertainty, and audit trails. Patient-facing agents should avoid a definitive diagnosis when the risk is high and should direct patients to urgent care when systemic symptoms, a rapidly spreading rash, mucosal involvement, necrosis, purpura, or suspected

melanoma is present. Clinician-facing agents can reduce documentation and the cognitive burden while preserving professional accountability.

5.4. Fairness-by-design and global data collaboration

Federated learning, secure aggregation, differential privacy, and trusted research environments can facilitate collaboration across institutions without centralizing sensitive images or records (*31*). However, privacy-preserving learning does not automatically ensure fairness. A federated network dominated by similar populations can still reproduce bias. Collaboration should therefore be deliberately structured to include diverse skin tones, geographic regions, age groups, anatomical sites, disease categories, and care settings. Synthetic images may help balance rare categories, but they must be reviewed by experts and tested for leakage, artifact amplification, and skin-tone distortion.

A practical global reporting standard should require every dermatology AI study to describe its dataset composition, skin-tone distribution, image source, reference standard, subgroup performance, failure modes, calibration, and external validation. Journals and regulators should also require explicit reporting on darker skin, inflammatory disorders, rare diseases, and real-world smartphone images.

5.5. Prospective clinical trials and post-deployment learning

Most dermatology AI evidence remains retrospective. Prospective studies should determine whether AI improves outcomes that matter clinically: time to melanoma diagnosis, appropriateness of referral, biopsy yield, missed cancer rate, flare-up control, treatment adherence, clinician workload, patient satisfaction, and equity. The SPIRIT-AI, CONSORT-AI, and DECIDE-AI reporting frameworks provide useful principles for AI clinical trials and early-stage evaluations, but dermatology-specific endpoints are still needed (*30,32,33*).

Post-deployment monitoring is equally important because imaging devices, clinical practice, disease prevalence, and treatment options change over time. Drift detection, periodic recalibration, safety audits, and clinician feedback mechanisms should be required. Adaptive updates should remain separate from real-time decision-making support unless a regulated change-control plan is in place. Table 3 provides a practical deployment checklist for dermatology AI.

6. AI's potential applications and future prospects in dermatology

6.1. Teledermatology and primary care triage

Teledermatology is the most immediate setting for multimodal AI because it already depends on asynchronous images and text. AI can assist by checking image quality, collecting a structured history, identifying high-risk lesions, suggesting differentials for common rashes, and prioritizing referrals. In primary care, AI may help clinicians decide whether a lesion requires an urgent dermatology assessment, whether a rash can be treated empirically, or whether biopsy, culture, serology, or emergency evaluation is needed (*24,25*).

The purpose of these systems should not be framed as replacing dermatologists. Rather, AI can help direct the right patient to the right level of care. Low-risk benign lesions may be managed with reassurance and follow-up, whereas suspected melanoma, a rapidly progressing infection, vasculitis, or a severe drug eruption should be escalated. Successful triage depends on a high level of sensitivity to dangerous diseases, transparent uncertainty, skin-tone equity, and workflows that reduce rather than increase the clinician's burden.

6.2. Rare and difficult diseases

Rare and diagnostically difficult diseases are natural targets for multimodal dermatology AI because individual clinicians may encounter only a small number of cases, whereas aggregated data can reveal recognizable patterns. CTCL, autoimmune blistering diseases, genodermatoses, vasculitis, connective tissue diseases, and rare infections often require integration of clinical morphology, distribution, laboratory data, pathology, immunofluorescence, molecular testing, and follow-up. AI systems may aid in these cases by retrieving similar examples, ranking differential diagnoses, and detecting discordance between clinical and histopathologic findings.

Few-shot learning, retrieval-augmented reasoning, and knowledge graphs may be particularly useful. Even when a model cannot confidently classify a rare disease, it may provide safety-oriented reasoning, such as noting that a chronic treatment-resistant patch or plaque dermatitis with poikiloderma and atypical lymphocytes warrants concern about mycosis fungoides and calls for a repeat biopsy. This type of cautious, evidence-linked output may be more clinically valuable than forcing a high-confidence label.

6.3. Treatment decision-making support and efficacy prediction

Dermatology AI is moving from diagnosis towards decision-making support. In psoriasis and atopic dermatitis, multimodal models may combine images, severity scores, itching and sleep measures, comorbidities, laboratory data, prior therapies, adherence, and pharmacogenomic or transcriptomic signals to predict the response to biologics, Janus kinase inhibitors, phototherapy, or systemic immunomodulators. In acne,

Table 3. Dermatology AI deployment checklist: From validation to clinical use

Risk domain	Dermatology-specific manifestation	Required evaluation	Practical mitigation
Skin-tone fairness	Erythema and vascular cues vary by pigmentation; darker skin is underrepresented in many datasets.	Reporting performance by skin tone, disease type, anatomical site, and image source; auditing false negatives and false positives.	Diverse recruitment, skin-tone-aware sampling, color calibration, fairness dashboards, and local validation.
Real-world image quality	Blur, lighting, compression, cosmetics, hair, scale, ulceration, and variable smartphone cameras.	Stress-testing teledermatology and primary care images; measuring the reject rate and clinician override rate.	Image-quality prompts, acquisition guidance, reference cards, and safe rejection when images are inadequate.
Clinical safety	Missed melanoma, severe drug eruption, vasculitis, necrotizing infection, or immunosuppression-related disease.	Evaluate high-risk triage sensitivity, calibration, uncertainty, and red-flag recognition.	Escalation rules, clinician confirmation, conservative thresholds, and emergency warnings.
Workflow integration	Standalone tools add clicks and may not fit EHR or teledermatology pathways.	Ascertaining time, referral quality, biopsy yield, user trust, and documentation burden.	EHR integration, structured output, audit trails, and human-centered interface design.
Regulation and liability	Adaptive multimodal models may change over time and influence biopsy, referral, or therapy decisions.	Defining intended use, autonomy level, locked *vs.* adaptive status, cybersecurity, and change-control plans.	SaMD pathway, predetermined change-control plans, post-marketing surveillance, and clear responsibility.
LLM hallucination	Confident but unsupported diagnosis, unsafe treatment advice, or missed red flags.	Benchmarking factuality, guideline consistency, uncertainty expression, and failure cases.	Retrieval-augmented generation, curated knowledge bases, guardrails, and clinician review.
Longitudinal drift	Disease prevalence, cameras, guidelines, therapies, and patient populations evolve.	Monitoring calibration and subgroup performance over time; comparing with respect to prior model versions.	Drift detection, scheduled recalibration, rollback mechanisms, and prospective auditing.

Abbreviations: AI, artificial intelligence; EHR, electronic health record; LLM, large language model; SaMD, software as a medical device.

hidradenitis suppurativa, and rosacea, longitudinal images and patient-reported outcomes may help monitor severity and treatment response. In melanoma, AI may integrate histology, clinicopathologic variables, and molecular markers for risk stratification and trial matching (*18-20*).

These applications require higher safety standards than classification because they affect treatment exposure, cost, and patient expectations. Recommendations should be linked to guidelines, they should factor in local formulary constraints, and they should be patient-centered and explainable. The model should distinguish diagnostic confidence from therapeutic evidence, and clinicians should remain responsible for final treatment decisions.

6.4. Education and research innovation

Multimodal AI can enhance dermatology education by generating case-based learning materials that include images across skin tones, morphology labels, distribution maps, histopathology, and differential reasoning. It can also give trainees immediate feedback, not only on whether an answer is correct but also on which morphology clues were missed. In research, AI can mine image-text-pathology repositories to identify phenotypes, disease subtypes, biomarker associations, and treatment response patterns. Foundation models may also accelerate clinical trial screening by matching patient phenotypes to eligibility criteria.

Educational and research applications must still ensure privacy, consent, and fairness. Synthetic cases should be clearly labeled, and generated educational material should be reviewed by experts. For trainee education, AI should emphasize uncertainty and differential diagnosis rather than encouraging single-label pattern recognition.

6.5. From diagnosis-centric AI to continuous skin intelligence

The future paradigm of dermatology AI should not be limited to "taking a photograph and receiving a diagnosis". A more ambitious and clinically realistic direction is continuous skin intelligence: systems that track lesions, inflammation, wounds, symptoms, exposure, and treatment response over time. This paradigm fits dermatology because many important decisions depend on trajectory rather than a single visual snapshot.

Such systems could combine patient-submitted images, total-body photography, wearable or smartphone sensing, itching and pain diaries, environmental triggers, medication timelines, sleep and activity data, and clinician-verified outcomes. They could monitor melanoma risk through detection of changes,

quantification of psoriasis and atopic dermatitis flare-ups, follow hidradenitis suppurativa activity, gauge wound healing, and identify patients whose condition needs to be managed before a severe flare-up occurs.

However, continuous monitoring also creates new risks: overdiagnosis, anxiety, privacy leakage, inequitable access, false alarms, and unclear responsibility for unattended alerts. The goal should be clinician-supervised longitudinal decision-making support and not surveillance for its own sake. Safe systems must define what is monitored, who receives alerts, how uncertainty is displayed, and when the model should remain silent.

7. Conclusion

Dermatology AI has progressed from experimental CNN classifiers to multimodal systems that can integrate images, text, structured patient information, and histopathology. The key conceptual advance is that dermatology AI should be built around clinical reasoning rather than generic computer vision. The field must learn morphology, distribution, skin tone, symptoms, temporal evolution, clinicopathologic correlation, and therapeutic context.

The major barriers are also dermatology-specific. Image-only models are vulnerable to morphologic ambiguity, acquisition variability, and missing context. Skin-tone bias can alter the visual signal itself, and particularly in erythema-dependent diseases. Benchmark accuracy may not translate to teledermatology, primary care, rare-disease diagnosis, or treatment decision-making support. LLMs and agentic systems add important capabilities but also risks, including hallucinations, automation bias, privacy leakage, and unclear medico-legal responsibility.

Future progress will depend on foundation models that are morphology-aware, skin-tone-aware, longitudinal, and clinically grounded; datasets that are diverse and responsibly governed; explanations that connect visual evidence to dermatological language; regulatory pathways that address adaptive multimodal software; and prospective studies that measure outcomes valued by patients and clinicians. With these foundations, AI can become a trusted component of digital dermatology - not a replacement for dermatologists, but a tool that expands access, improves triage, improves diagnostic reasoning, and facilitates more personalized care. The central methodological challenge is to decide when AI should recognize a lesion, when it should reason across modalities, and when it should defer because the available inputs do not support a safe dermatological determination.

Funding: This work was supported by the National Natural Science Foundation of China (82203935), the Fujian Province Natural Science Foundation (2026J001590), Joint Funds for the innovation of Science and Technology, Fujian Province (2025Y9299), and Fujian Medical University Union Hospital's Project to Foster Excellent Young Scholars (2022XH027).

Conflict of Interest: The authors have no conflicts of interest to disclose.

References

1. Hay RJ, Johns NE, Williams HC, Bolliger IW, Dellavalle RP, Margolis DJ, Marks R, Naldi L, Weinstock MA, Wulf SK, Michaud C, J L Murray C, Naghavi M. The global burden of skin disease in 2010: An analysis of the prevalence and impact of skin conditions. J Invest Dermatol. 2014; 134:1527-1534.
2. Karimkhani C, Dellavalle RP, Coffeng LE, Flohr C, Hay RJ, Langan SM, Nsoesie EO, Ferrari AJ, Erskine HE, Silverberg JI, Vos T, Naghavi M. Global skin disease morbidity and mortality: An update from the Global Burden of Disease Study 2013. JAMA Dermatol. 2017; 153:406-412.
3. 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.
4. 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; 29:1836-1842.
5. Tschandl P, Rinner C, Apalla Z, *et al*. Human-computer collaboration for skin cancer recognition. Nat Med. 2020; 26:1229-1234.
6. Liu Y, Jain A, Eng C, *et al*. A deep learning system for differential diagnosis of skin diseases. Nat Med. 2020; 26:900-908.
7. Han SS, Park I, Eun Chang S, Lim W, Kim MS, Park GH, Chae JB, Huh CH, Na JI. Augmented intelligence dermatology: Deep neural networks empower medical professionals in diagnosing skin cancer and predicting treatment options for 134 skin disorders. J Invest Dermatol. 2020; 140:1753-1761.
8. Tschandl P, Rosendahl C, Kittler H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Sci Data. 2018; 5:180161.
9. Marchetti MA, Codella NCF, Dusza SW, *et al*. Results of the 2016 International Skin Imaging Collaboration International Symposium on Biomedical Imaging challenge: Comparison of the accuracy of computer algorithms to dermatologists for the diagnosis of melanoma from dermoscopic images. J Am Acad Dermatol. 2018; 78:270-277.e1.
10. Moor M, Banerjee O, Abad ZSH, *et al*. Foundation models for generalist medical artificial intelligence. Nature. 2023; 616:259-265.
11. Haug CJ, Drazen JM. Artificial intelligence and machine learning in clinical medicine, 2023. N Engl J Med. 2023; 388:1201-1208.
12. Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. Nat Med. 2022; 28:31-38.
13. Topol EJ. High-performance medicine: The convergence of human and artificial intelligence. Nat Med. 2019; 25:44-

56.
14. Lu MY, Chen B, Williamson DFK, *et al*. A multimodal generative AI copilot for human pathology. Nature. 2024; 634:466-473.
15. Yan S, Yu Z, Primiero C, *et al*. A multimodal vision foundation model for clinical dermatology. Nat Med. 2025; 31:2691-2702.
16. Daneshjou R, Vodrahalli K, Novoa RA, *et al*. Disparities in dermatology AI performance on a diverse, curated clinical image set. Sci Adv. 2022; 8:eabq6147.
17. Adamson AS, Smith A. Machine learning and health care disparities in dermatology. JAMA Dermatol. 2018; 154:1247-1248.
18. Choy SP, Kim BJ, Paolino A, Tan WR, Lim SML, Seo J, Tan SP, Francis L, Tsakok T, Simpson M, Barker JNWN, Lynch MD, Corbett MS, Smith CH, Mahil SK. Systematic review of deep learning image analyses for the diagnosis and monitoring of skin disease. NPJ Digit Med. 2023; 6:180.
19. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. BMC Med. 2019; 17:195.
20. Widiawaty A, Indriatmi W, Jatmiko W, Novianto E, Kekalih A, Gunawan H, Palar PS, Rachmadi MF, Dermawan S, Malahayati TL, Ramadhan AW. Multimodal machine learning approach for diagnosing atopic dermatitis. F1000Res. 2025; 14:952.
21. Tschandl P, Rosendahl C, Akay BN, *et al*. Expert-level diagnosis of nonpigmented skin cancer by combined convolutional neural networks. JAMA Dermatol. 2019; 155:58-65.
22. Brinker TJ, Hekler A, Enk AH, Klode J, Hauschild A, Berking C, Schilling B, Haferkamp S, Schadendorf D, Holland-Letz T, Utikal JS, von Kalle C; Collaborators. Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task. Eur J Cancer. 2019; 113:47-54.
23. Goyal M, Knackstedt T, Yan S, Hassanpour S. Artificial intelligence-based image classification methods for diagnosis of skin cancer: Challenges and opportunities. Comput Biol Med. 2020; 127:104065.
24. Pacheco AGC, Lima GR, Salomão AS, *et al*. PAD-UFES-20: A skin lesion dataset composed of patient data and clinical images collected from smartphones. Data Brief. 2020; 32:106221.
25. Freeman K, Dinnes J, Chuchu N, Takwoingi Y, Bayliss SE, Matin RN, Jain A, Walter FM, Williams HC, Deeks JJ. Algorithm based smartphone apps to assess risk of skin cancer in adults: Systematic review of diagnostic accuracy studies. BMJ. 2020; 368:m127.
26. Monheit G, Cognetta AB, Ferris L, Rabinovitz H, Gross K, Martini M, Grichnik JM, Mihm M, Prieto VG, Googe P, King R, Toledano A, Kabelev N, Wojton M, Gutkowicz-Krusin D. The performance of MelaFind: A prospective multicenter study. Arch Dermatol. 2011; 147:188-194.
27. Venkatesh KP, Kvedar JC. Learnings from the first AI-enabled skin cancer device for primary care authorized by FDA. NPJ Digit Med. 2024; 7:167.
28. Muehlematter UJ, Daniore P, Vokinger KN. Approval of artificial intelligence and machine learning-based medical devices in the USA and Europe (2015-20): A comparative analysis. Lancet Digit Health. 2021; 3:e195-e203.
29. Benjamens S, Dhunnoo P, Mesko B. The state of artificial intelligence-based FDA-approved medical devices and algorithms: An online database. NPJ Digit Med. 2020; 3:118.
30. Vasey B, Nagendran M, Campbell B, *et al*. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. Nat Med. 2022; 28:924-933.
31. Hossen MN, Panneerselvam V, Koundal D, Ahmed K, Bui FM, Ibrahim SM. Federated machine learning for detection of skin diseases and enhancement of IInternet of Medical Things (IoMT) Security. IEEE J Biomed Health Inform. 2023; 27:835-841.
32. Cruz Rivera S, Liu X, Chan AW, Denniston AK, Calvert MJ; SPIRIT-AI and CONSORT-AI Working Group; SPIRIT-AI and CONSORT-AI Steering Group; SPIRIT-AI and CONSORT-AI Consensus Group. Guidelines for clinical trial protocols for interventions involving artificial intelligence: The SPIRIT-AI extension. Nat Med. 2020; 26:1351-1363.
33. Liu X, Rivera SC, Moher D, Calvert MJ, Denniston AK; SPIRIT-AI and CONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: The CONSORT-AI extension. Nat Med. 2020; 26:1364-1374.

Received April 28, 2026; Revised May 21, 2026; Accepted June 9, 2026.

Released online in J-STAGE as advance publication June 13, 2026.

[§]*These authors contributed equally to this work.*
**Address correspondence to*:
Yue Han, Department of Dermatology, The Union Hospital, Fujian Medical University, No. 29 Xinquan Road, Fuzhou 350001, China.
E-mail: dr_hanyue@126.com