PATHFINDER PUBLISHER

PATHFINDER OF RESEARCH

PATHFINDER OF RESEARCH
THE INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

Review Research

Artificial Intelligence for Skin Cancer Imaging: Performance, Translational Gaps, and a Three-Tier Clinical Readiness Framework

Sumana Das [1], Sheikh Farhan Ibn Rahman [2], Md. Rokibul Hasan [3*], M A S Ansari [4]

[1] Khulna Medical College, Khulna, Bangladesh

[2] Faculty of Biotechnology and Genetic Engineering, Sylhet Agricultural University, Sylhet, Bangladesh

[3] Sylhet M A G Osmani Medical College, Sylhet, Bangladesh

[4] Faculty of Medicine, Manipal University College Malaysia, Melaka, 75150, Bukit Baru, Malaysia

Abstract: Artificial Intelligence **(**AI)-driven image analysis has made significant strides in dermatologic cancer care, showing dermatologist-level accuracy in melanoma detection through various publications. However, its integration into routine clinical practice remains limited, resulting in minimal impact in real-world scenarios. This review assesses AI applications across three main imaging methods in skin cancer management: clinical photography, dermoscopy, and histopathological slides. Key factors affecting diagnostic efficacy include image quality, annotation methods, dataset composition, and deployment context. Convolutional neural networks, particularly when provided with high-resolution images and validated biopsy labels, demonstrate effectiveness comparable to skilled dermatologists. However, performance diminishes significantly when models are tested on diverse devices, skin types, or lesion types. Clinical photograph systems, while less accurate, offer advantages in scalable triage and teledermatology. Histopathological models produce robust confirmatory results; however, they face obstacles including inadequate supervision, variability in staining, and substantial computational demands. Common issues across these modalities include dataset bias, underrepresentation of darker skin tones and benign lesions, noisy ground truth, and poor integration into clinical workflows. The authors put forward a three-tier framework to get the medical field ready for AI. They suggest using it in primary care to screen for lesions, as a decision-support tool in dermoscopy, and as a computational biomarker in digital pathology. Future developments will hinge on creating representative datasets, advancing training techniques, ensuring interpretability, long term monitoring of lesions, and achieving regulatory validation. Ultimately, mature AI systems are anticipated to act as diagnostic collaborators rather than replacements for clinicians, focusing on reducing unnecessary procedures, standardizing biopsy criteria, and improving equity in skin cancer management.

Keywords: Artificial intelligence; Skin cancer; Dermoscopy; Deep learning; Clinical photography; Histopathology; Diagnostic imaging

Check for Updates

Article History:
Received: 11 October 2025
Accepted: 13 December 2025
Online: 15 December 2025

Corresponding author:
rokibhasan54somc@gmail.com

Citations: Das, S., Rahman, S. F. I., Hasan, M. R. & Ansari, M. A. S. (2025). Artificial Intelligence for Skin Cancer Imaging: Performance, Translational Gaps, and a Three-Tier Clinical Readiness Framework. *Pathfinder of Research*, *3*(3), 47-63

1. Introduction

Skin cancer is the most frequent type of cancer found around the world, and in many places, the number of cases is continually rising. Melanoma is not extremely common, but it kills many people with skin cancer because it spreads to other parts of the body rapidly. Basal cell carcinoma (BCC) and squamous cell carcinoma (SCC) are more common; however, they aren't usually as fatal when treated right away. Recent global estimates suggest that non-melanoma skin cancers affect several million individuals annually, while tens of thousands of new melanoma cases are reported each year just in high-income countries **(American Cancer Society, 2024; WHO, 2024).** The survival rate for early-stage melanoma exceeds 95%; however, the prognosis markedly declines with a delayed diagnosis **(Salinas et al., 2024).** The primary clinical challenge is not the lack of suitable treatment but the failure to consistently and swiftly detect high-risk lesions.

There are several reasons why diagnostic delays develop, and they depend on the situation. In low-resource settings, limited access to dermatologists, lengthy referral processes, and insufficient biopsy facilities lead to many patients initially presenting with severe disease. In more affluent health systems, the issue is distinct: there exists a prevalence of benign lesions, considerable discord among observers regarding their visual assessment, and significant ambiguity when lesions reside in diagnostic gray zones, such as early melanoma versus atypical or dysplastic nevi **(Liu et al., 2023; Akhter et al., 2025).** Dermoscopy has made it easier to find subsurface features and unique pigment patterns, but even experienced doctors still misclassify many lesions, especially when they are busy or have many patients to see.

Some people believe that AI, especially deep learning, could be valuable in these situations. Convolutional neural networks (CNNs) can learn hierarchical representations directly from large collections of images. This means they don't have to do manual feature engineering and can employ subtle, high-dimensional cues that are difficult for individuals to put into words. Leading benchmark studies demonstrate that CNNs can attain or exceed dermatologist-level performance in melanoma classification when trained on rigorously curated dermoscopic datasets with biopsy-validated labels **(Esteva et al., 2017; Brinker et al., 2019).** These results have generated considerable research interest in AI-assisted pipelines that include lesion detection, segmentation, classification, and risk stratification.

But people become so enthused about headline accuracy figures that they forget that most models are built in scenarios that are far different from how things really work. Algorithms typically perform less effectively when presented with images of new institutions, other technologies, or a broader spectrum of individuals. Changes in lighting, camera settings, resolution, lesion framing, and skin phototype, as well as changes in the types of benign and malignant lesions, all make it difficult to generalize and cause domain shift **(Goyal et al., 2020; Wu et al., 2022).** Additionally, many datasets are created from patients who have been pre-selected for expert evaluation, leading to an overrepresentation of diagnostically complex or high-risk lesions and an underrepresentation of common benign conditions. Such variation makes both training and testing less accurate, which makes it seem like performance in the real world is better than it really is. AI research in skin cancer has also grown beyond dermoscopy. There are three main imaging contexts that are common in the literature right now. First, clinical photos are the first thing patients see when they come into contact with the health system.

These images are usually captured with mobile phones or normal digital cameras. These photos are diverse from one another, noisy, and usually feature more than one lesion or a background structure that is challenging to understand. But they are highly useful for scalable triage and teledermatology. Second, dermoscopic images provide ordered, high-contrast representations of lesion anatomy and have become the primary basis for AI studies claiming dermatologist-level performance **(Tschandl et al., 2019b; Shetty et al., 2022).** Third, histopathological whole-slide images are the best way to tell if a tumor is cancerous, but they are difficult to work with because they are so big (gigapixels), the labels are not obvious, and different labs use different methods for staining and scanning **(Liopyris et al., 2022; Melarkode et al., 2023).** Because of this, each modality is at a different point in the diagnostic process and needs distinct techniques to model, validate, and combine.

Even while the literature is developing swiftly, there is still not much agreement on how to look at the information as a whole in terms of being ready for real clinical situations. Many assessments summarize model architectures and performance measures but inadequately address dataset bias, ground-truth integrity, external validation, or practical workflow constraints **(Mukherjee Behara et al., 2024; Martinez-Vargas et al., 2025)**. Some people only care about dermoscopy and don't consider the wider picture of digital pathology and clinical photography. For clinicians and policymakers, the pertinent questions are not "Can AI compete with experts on a test set?" but rather "In what circumstances does AI enhance patient care?", "At which stages of the care pathway is AI most beneficial?", and "How can we ensure these tools do not exacerbate existing disparities, particularly for marginalized skin types and health systems?"

This review tackles these shortcomings by scrutinizing various AI types and their applications in the context of skin cancer diagnosis. To start, we look at AI systems that use clinical photographs, dermoscopic pictures, and histological slides. We examine how the data's quality, labeling, and task setup affect the system's ability to identify issues. Second, we generated a list of cross-cutting issues, such as dataset bias, class imbalance, not enough skin tone variability, noisy labeling, domain shift, limited explainability, privacy issues, and difficulty with integrating workflows. Third, we discuss new technical and organizational ideas that could help the field advance beyond proof-of-concept models to tools that can be utilized in real life. Some of them are federated learning, domain-generalization strategies, self-supervised pretraining, and interpretability frameworks **(Team, Scientific Data Curation, 2021; Li et al., 2022; Gohil & Desai, 2024).**

To ground this discussion in clinical reality, we propose a three-tier clinical readiness framework that delineates the specific roles AI can assume in skin cancer care: (i) autonomous suspicion screening in community and primary care utilizing clinical photographs, (ii) high-specificity decision support during dermoscopic evaluation prior to biopsy, and (iii) computational biomarker extraction from histopathology for diagnostic validation and prognostic enhancement. Putting current evidence into this layered structure makes it easier to honestly assess what AI can and can't do right now, where it is most likely to have an effect, and which research areas are most important if AI is to make a real difference in skin cancer by finding it earlier, reducing overtreatment, and making outcomes fairer.

2. AI-Based Imaging and Diagnostic Methods for Skin Cancer

Artificial intelligence systems developed for skin cancer diagnosis have diverse behaviors contingent upon the

imaging modalities employed during their training. Therapeutic photographs, dermoscopic images, and histopathological whole-slide scans create distinct visual conditions that influence model performance, alter annotation reliability, and determine therapeutic utility **(Tschandl et al., 2019b; Shetty et al., 2022).** People have had great hopes for these systems because they have often been regarded as equivalent. A realistic evaluation must, therefore, assess modality-specific performance and its alignment with genuine therapeutic pathways. This phase puts together the evidence while retaining the key framework that a high-impact review need.

2.1 AI on Dermoscopic Images: Great Performance When There Are Limits

Dermoscopy offers you pictures that are bigger, polarized, and arranged in a way that assists convolutional structures a lot. These pictures have diagnostic traits such as pigment networks, streaks, regression zones, and vascular patterns. This feature enables deep neural networks to extract hierarchical representations that are closely associated with melanoma risk **(Winkler et al., 2023)**. Initial seminal studies indicated that deep learning could perform comparably to or beyond dermatologists on standardized dermoscopic datasets **(Haenssle et al., 2018; Brinker et al., 2019).** AI is frequently quite sensitive in these controlled circumstances and often does better than doctors at being specific, which means fewer unnecessary biopsies.

However, this strength primarily stems from the quality of the dermoscopic datasets, not their actual strength. Public datasets such as ISIC and HAM10000 contain lesions that are more diagnostically intriguing than common benign ones. Most of the patients they see have lighter skin types, and they take pictures in the most prestigious locations **(Goyal et al., 2020; Santoro, 2023)**. Models trained on these datasets perform markedly worse when exposed to unfamiliar devices, institutions, demographics, or lighting conditions. Figure 1 illustrates the variety of dermoscopic lesion categories typically present in study datasets.

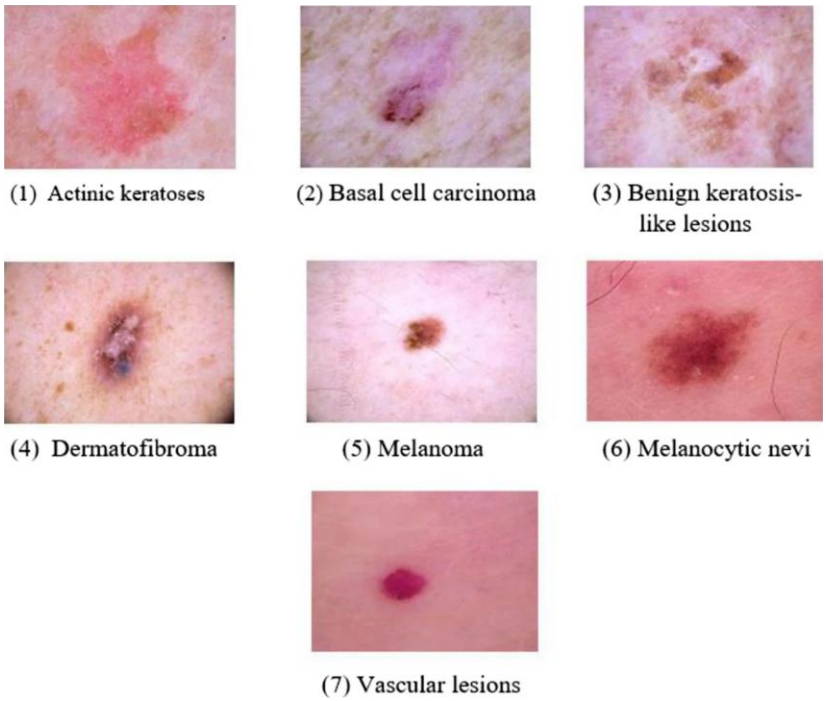

Figure 1: Different categories of dermoscopic skin lesions (Adopted from Shetty et al., 2022).

The structured detail evident in figure 1 explains why dermoscopy-based AI performs strongly in controlled environments yet becomes fragile under domain shift. Similarly, figure 02, adopted from **Goyal et al. (2020)**, shows an ensemble CNN interpretation for lesion categorization, highlighting how model predictions rely on high-quality, centrally framed dermoscopic fields.

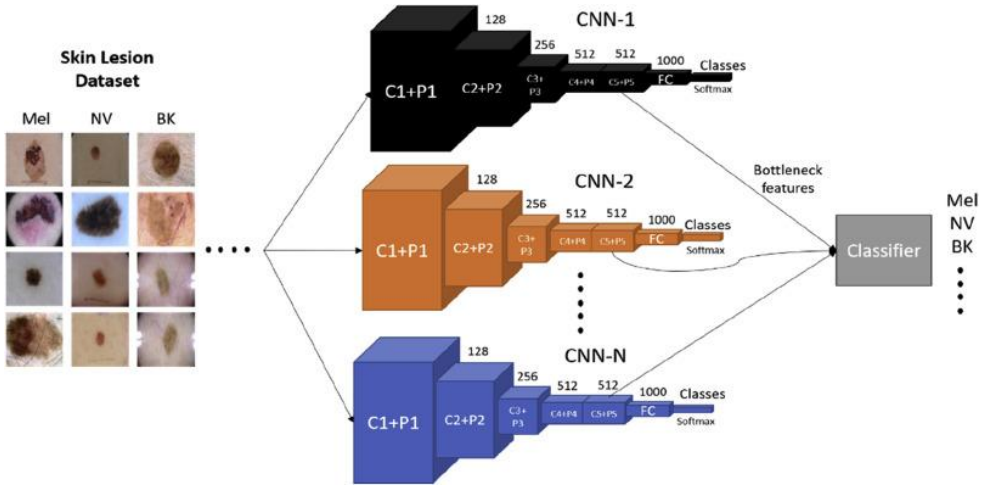

Figure 2: Skin lesion categorization using an ensemble CNN technique (Adopted from Goyal et al., 2020).

Although dermoscopy provides the richest visual substrate for training, its clinical readiness is limited by dataset bias and poor generalization. This places dermoscopic AI within the category of diagnostic refinement tools, useful once a lesion has already reached specialist evaluation but not reliable enough for autonomous diagnosis.

Table 1: Major Publicly Used Image Datasets in AI-Based Dermatology Research.

Dataset name	Approximate image count	Primary label source	Predominant skin types represented	Key characteristics and notes
ISIC Archive	> 33,000 dermoscopic images	Expert review plus biopsy for subsets	Mostly Fitzpatrick I-III	Largest open dermoscopic repository; benchmark for many CNN studies; substantially biased toward suspicious lesions **(Goyal et al., 2020)**.
HAM10000	~10,015 dermoscopic images	Histology, confocal imaging and follow-up	Mostly lighter phototypes, small non-Caucasian fraction	Frequently used for multiclass lesion classification; better class balance than older sets but still limited diversity in skin tone **(Mendonça et al., 2013)**.
PH2 Dataset	200 dermoscopic images	Biopsy-verified	Limited range of skin types	Small, high-quality dataset with very reliable labels; mainly used for

Dataset name	Approximate image count	Primary label source	Predominant skin types represented	Key characteristics and notes
				segmentation and calibration work **(Mendonça et al., 2013)**.
Dermofit	~1,300 clinical and dermoscopic images	Dermatologist visual diagnosis	Predominantly light phototypes	Includes multiple benign and malignant classes; useful for model calibration but relatively small and geographically narrow **(DERMOFIT, n.d.)**.
Asan Dataset	~17,000+ dermoscopic images	> 12,000 biopsy-confirmed lesions	Largely Asian population	Institution-specific dataset with strong histologic confirmation; valuable for assessing generalization to non-European cohorts **(Han et al., 2018; Goyal et al., 2020)**.

2.2 AI on Clinical Photographs: Broader Reach, Lower Predictive Precision

Clinical photographs show how patients first see the healthcare system. These photographs were taken with smartphones or regular cameras in places where the lighting isn't always adequate. The quality and backdrops are different, and they show hair, freckles, inflammation, and other things that take away from the picture. Lesions often only cover a small part of the frame. In these cases, AI must first identify the lesion before putting it into a category. When full-frame pictures are used without segmentation, models often mistake background textures for diagnostic signs, which leads to inconsistent results. The use of segmentation or adaptive cropping greatly improves how well categorization works **(Gessert et al., 2020).**

Clinical-image databases demonstrate markedly greater diversity compared to dermoscopic archives, including a wide range of benign and non-neoplastic dermatological conditions. This unpredictability increases the false-positive rate for melanoma classification, especially when models are mostly trained on dermoscopic datasets that focus on melanoma. Despite their shortcomings, clinical-image AI algorithms are essential for real-world triage. Future primary-care assessments show that AI-assisted screening cuts down on unnecessary referrals and accelerates the process of finding high-risk cases **(Giavina-Bianchi et al., 2021).**

Table 2: Summary of AI Performance Across Imaging Modalities

Imaging modality	Typical input source	Data resolution	Main strengths	Main limitations	Most appropriate clinical role
Clinical	Smartphone	or Low	to Very	accessible; High	noise; Early suspicion

Imaging modality	Typical input source	Data resolution	Main strengths	Main limitations	Most appropriate clinical role
photographs	consumer camera images	medium	supports teledermatology and large-scale triage; cheap to collect	background clutter; strong device and lighting variability; broad mix of benign conditions	screening and triage in community and primary-care settings
Dermoscopic images	Polarized, magnified dermatoscope images	Medium to high	Highly discriminative morphology; best benchmark performance; strong feature richness	Dataset bias; over-representation of suspicious lesions; poor representation of darker skin; domain shift across centers	Diagnostic refinement and biopsy decision support in specialist dermatology clinics
Histopathology whole slides	Digitized histopathology (WSI) from scanners	Very high (gigapixel)	Definitive ground truth; fine-grained cellular and architectural detail; potential for quantitative biomarkers	Weak supervision at slide level; severe stain/device variability; high computational burden; reproducibility issues	

Clinical photographs therefore provide accessibility rather than high diagnostic precision. Their most realistic clinical role aligns with early-stage triage, allowing clinicians to determine which lesions warrant dermoscopic evaluation or urgent referral. This shift from diagnostic ambition to triage utility is consistent with current evidence **(Team, 2020; Mukherjee et al., 2025)**.

2.3 Histopathology-Based AI: Definitive Ground Truth with Gigapixel Complexity

Histopathology is still the greatest way to detect if someone has melanoma, but it's also the toughest thing for AI to perform. Whole-slide images (WSIs) are too enormous to look at all at once; therefore, they need to be broken up into smaller pieces. Weak monitoring makes matters considerably difficult because slide-level labels are used even when only a small part of the tissue is malignant. Contemporary models employ multiple-instance learning and attention mechanisms to ascertain the presence of diagnostic signals within patches **(Melarkode et al., 2023).**

When employed in a single lab, AI based on histopathology shows a lot of potential for discovering tumor nests,

mitotic activity, and architectural problems. But even small changes to the chemical used for staining, the thickness of the slides, or the technology used for scanning can cause a lot of domain drift. Models trained in one place don't always work well when used outside of that place, unless strong stain-normalization methods or domain-adaptation methods are used **(Li et al., 2022).**

Histopathology AI should be viewed as an additional analytical layer that validates existing knowledge among pathologists, rather than a substitute for their clinical expertise. The clinical pathway should position it at the confirming end of the diagnostic spectrum, where adherence to regulations, the ability to repeat tests, and the comprehension of results are all crucial aspects.

2.4 Integrating Modality-Specific AI into a Unified Diagnostic Pathway

The three types of imagery work together to create a layered AI landscape. Clinical photographs are easy to access to, but they don't always provide accurate predictions. Dermoscopy presents organized morphological data that makes it easier to classify things in controlled settings. Histopathology presents clear ground truth, but it can be problematic for AI systems because of staining variability and gigapixel complexity. Your charts and graphs make it obvious that there is a big difference between the two methods of imaging.

3. Major Challenges and Limitations in AI-Based Skin Cancer Diagnosis

Even though they are becoming better quickly and often say they are as competent as dermatologists, AI systems for diagnosing skin cancer still have difficulties with their structure, methods, and use that make them less dependable in real life. Many of the best results come from datasets and evaluation methods that are too limited and don't represent how varied and difficult real clinical settings are. As a result, there is still a big difference between benchmark performance and actual clinical readiness. Understanding these limitations is essential for evaluating present results and guiding future research that authentically meets clinical needs rather than merely technical convenience **(Tschandl et al., 2019a).**

3.1 Dataset Bias, Class Imbalance, and Insufficient Skin Tone Diversity

A primary challenge in dermatologic AI is the intrinsic bias found in publicly available datasets. ISIC and HAM10000 contain too many lesions that look strange or suspicious since most of the samples come from tertiary dermatological centers, where patients travel after an early stage of clinical triage. Such bias makes the case spectrum look weird because there are a lot more melanomas and other high-risk lesions than there are in typical practice. Models trained on such data inevitably learn decision limits that are overly tuned for differentiating between high-risk lesions rather than properly detecting the substantially broader spectrum of benign diseases encountered in typical settings **(Gajjar et al., 2025).**

People are increasingly acknowledging another drawback: the scarcity of diverse skin phototypes. Fitzpatrick skin types I-III make up the majority of datasets. This means that models don't work well on skin tones that are darker. Lesions on darker skin show different contrast properties, pigment distribution, erythema visibility, and backdrop texture, all of which make it harder to categorize them. Numerous studies have documented reductions in accuracy and sensitivity exceeding ten to fifteen percentage points when models are evaluated using images from populations not included in the training set **(Tiva et al., 2025a).** This inconsistency not only

undermines clinical reliability but also threatens to worsen existing imbalances, particularly for areas already facing delayed diagnoses and institutional barriers to dermatologic care.

3.2 Noisy Labels, Weak Ground Truth, and Intrinsic Diagnostic Ambiguity

The quality of the labels used to train an AI system is what makes it trustworthy. In dermatology, many image databases rely on professional visual interpretation instead of biopsy-confirmed diagnoses. Even among experienced dermatologists, differences between observers are common, especially for borderline lesions like early melanoma, dysplastic nevi, and collision tumors. Research indicates that disagreement rates for certain cases are between 15% and 25%, which means that AI models trained on these labels learn how people make mistakes instead of a solid diagnostic truth **(Liopyris et al., 2022).**

Histopathology is much worse off when there is insufficient oversight. Whole-slide images are given one diagnostic label, yet the tumor areas may only represent a minor part of the tissue. When each tile obtains the slide-level label, the model has to determine which areas fit the pathological signal. This situation causes the forecasts at the patch level to be unstable. This can result in models that appear proficient in slide-level categorization yet fail to comprehend essential clinical factors such as tumor edges, architectural disarray, or mitotic activity **(Melarkode et al., 2023).** The absence of uniform pixel-level annotation limits the interpretability and overall clinical reliability of these systems.

3.3 Failure to Generalize and Weak External Validation

One of the main issues with AI research in dermatology is that models always fare worse when they are evaluated on data sets that are not their own. A lot of research looks at models by splitting one dataset into training and testing groups. This keeps the imaging conditions, lesion spectrum, and device properties the same across these groups without meaning to. This results in performance metrics that appear satisfactory but deteriorate upon the inclusion of additional center images. Many external validation studies show that using model predictions in real clinical contexts leads to big drops in sensitivity, specificity, and area-under-curve values **(Wu et al., 2022).**

Generalization failure occurs due to various factors, including differences in camera sensors, dermatoscope characteristics, lighting conditions, color calibration, anatomical locations, and patient demographics. In actual practice, lesions are often only partially visible, not well defined, or surrounded by other tissues like hair or inflammation. In these uncontrolled conditions, dermatologists generally do better than the models, especially when it comes to lesions that are challenging to diagnose. Regulatory bodies are hesitant to approve fully autonomous diagnostic systems, and one of the key reasons is that AI is still vulnerable to domain shift **(Salinas et al., 2024).**

3.4 Limited Explainability and Weak Clinical Trust

Usually, deep learning models produce probability scores without saying why. Saliency maps and gradient-based representations attempt to delineate key regions; nevertheless, their outputs frequently lack consistency, exhibit noise, or fail to correlate with clinically established dermoscopic patterns such as regression structures, blue-whitish veils, or pigment networks. Doctors, especially dermatologists and pathologists, need adequate

justifications for making judgments that could lead to a biopsy, surgery, or reassurance. AI systems that can't explain things clearly make it challenging for doctors to do their jobs and the output of algorithms.

Researchers found that doctors typically don't pay attention to AI ideas when the reasons aren't obvious or when the results don't match what they think **(Giavina-Bianchi et al., 2021).** In a field where medical and legal accountability is quite high, skepticism of black-box forecasts is understandable. AI is still not used in healthcare decision-making because its explanations are inadequate.

3.5 Privacy Issues, High Annotation Costs, and Data Fragmentation in Institutions

You need enormous, diversified, and well-annotated datasets to make strong AI models. But it takes a lot of time and money to develop these kinds of databases. Clinical photographs need to be carefully chosen, dermoscopic images need to be carefully labeled, and histopathology slides need to be carefully prepared and checked. Privacy restrictions like HIPAA and GDPR make it extra tough for organizations to share data. This leads to separate silos instead of unified databases that show the complete population.

Federated learning has emerged as a feasible approach by enabling model training across several locations without the requirement of sharing raw data; nonetheless, its effectiveness is dependent on stringent standardization of imaging methods, labeling conventions, and preprocessing workflows. Institutions rarely share similar workflows, hence limiting the practical applicability of federated approaches **(Melarkode et al., 2023).** AI systems are still limited to becoming instruments that only work in certain institutions, not diagnostic companions that everyone can utilize without a lot of standardized data.

3.6 Workflow Mismatch and Lack of Integration into Real Clinical Practice

Clinical contexts are complex decision-making processes that take into account a patient's medical history, the course of lesions, other health problems, risk factors, imaging, and the doctor's gut feeling. Most AI systems just look at one picture at a time and don't consider other things that could change the diagnosis. AI outputs frequently function as additional second opinions rather than contributing significantly to the diagnostic process. Dermatologists frequently disregard recommendations when the outcomes are illogical or when their incorporation into electronic health records becomes challenging **(Liopyris et al., 2022).** People probably won't use AI technologies that make it difficult to think, get in the way of work, or add steps that aren't necessary. AI should simplify clinical reasoning and documentation, not complicate them, to enable their effective integration.

3.7 Computational Burden and Hardware Limitations

Lastly, adopting this technology remains challenging due to various technological issues. It is extremely difficult to apply histopathology models in regions with few resources since they demand powerful GPUs and a lot of storage space to process gigapixel slides. Clinical-photo models, on the other hand, need to function with light technologies like cellphones. This means that the models can't be too intricate and may not be as precise. Different camera qualities, compression methods, and image resolutions make things even less stable **(Li et al., 2022).** In many low- and middle-income countries, the challenges get worse when imaging tools and network connections aren't always available. This makes the gap between experimental results and real-world viability much worse. As long as the quality of the hardware is closely related to how well the AI system can make

diagnoses, it will be impossible for AI systems to work equitably in different healthcare settings.

4. Future Opportunities for Artificial Intelligence in Skin Cancer Diagnosis

The growing study on artificial intelligence for skin cancer diagnosis shows that while there are some great technical advances, there are also big problems in making them work in the real world. To go from high-performance prototypes to clinically reliable systems, we need to change the way we train, test, integrate, and monitor models. Even while existing algorithms can reach expert-level accuracy in controlled settings, AI systems will only be truly useful in clinical settings when they can handle variability, explain their rationale clearly, treat all groups fairly, and be integrated into diagnostic workflows. To comprehend the future prospects for AI in dermatologic oncology, it is essential to view them not merely as incremental technical enhancements but as integral to a comprehensive shift towards clinically informed design.

In the future, it's vital to create large datasets that represent the entire population. The absence of variation in current datasets, including skin tone, lesion kind, imaging technology, or clinical situation, perpetuates the challenge of generalization. Collaborations across multiple institutions, along with privacy-preserving methods like federated learning, offer a viable way to create unified datasets without violating patient privacy. Federated learning enables institutions to provide model updates instead of raw data; nonetheless, its efficacy is contingent upon shared labeling techniques, uniform preprocessing pipelines, and defined imaging circumstances **(Urbi et al., 2021).** When done right, these kinds of distributed data ecosystems could help eradicate biases that have been around for a long time and make models that work well with a wide range of patients and imaging settings.

Another big chance is to design algorithms that focus on outward validity instead of optimizing for internal benchmark results. Traditional supervised learning frequently falters in the face of domain transition, necessitating innovative methodological strategies. Domain-generalization techniques, adversarial variance reduction, style-transfer augmentation, and self-supervised representation learning offer intriguing avenues for developing models that exhibit reduced sensitivity to lighting conditions, device variability, and background noise **(Wu et al., 2022; Tiva al., 2025a).** Transformer-based architectures, which use global receptive fields instead of just local ones, might make stability even better when lesions show up in places that aren't in the center or when there is a lot of background noise. These methods focus on resilience instead of pure accuracy, which shows a move toward models that can handle the volatility of real-world clinical imaging.

Explainability will also be a key factor in whether AI becomes a trusted diagnostic partner. Current saliency-based explanations frequently lack the precision or stability required to meet therapeutic standards. Future systems must produce outputs that are both interpretable and therapeutically significant, such as emphasizing specific dermoscopic structures like pigment networks, streaks, or regression zones, or distinguishing histopathologic traits such as mitotic figures and architectural anomalies. In prospective experiments, hybrid human-AI interaction models that incorporate uncertainty estimations or confidence overlays with explanations have been shown to enhance clinician trust and facilitate improved decision-making **(Gajjar et al., 2022).** For AI to be accepted in dermatology and pathology, it needs to be able to reason in a way that is consistent, clear, and anatomically correct with clinical heuristics. Adding time-based information to static photographs presents a transformative opportunity. Clinicians regularly evaluate the progression of lesions throughout time, noting

alterations in color, border geometry, and overall symmetry. AI systems that can look at long-term image sequences could identify subtle changes that people miss, making it easier to detect malignant transformations earlier. Adding patient metadata such as age, family history, immunologic status, or previous lesion history should improve risk stratification and make predictions more in line with the fact that skin cancer risk is caused by many things **(Li et al., 2022, Tiva et al., 2025b).** These multimodal systems might render more clinically consistent evaluations than models that solely use images.

Histopathology is also a wonderful place for new ideas to grow. Improvements in multiple-instance learning, attention-based aggregation, and graph neural networks are helping models identify both small cellular problems and bigger patterns in architecture. These systems could develop into instruments that facilitate tumor staging, margin evaluation, or the determination of prognostic indicators, including Breslow thickness and mitotic density **(Gessert et al., 2020).** Histopathology AI may eventually go beyond just making diagnoses to help with prognosis, treatment planning, and making individualized decisions about treatment. To succeed in this field, staining techniques must be standardized and validated by multiple institutions, and models that work in all lab settings must be created.

Simultaneously, deployment-focused AI presents prospects for influence across diverse healthcare environments. Smartphone-based triage solutions may be most useful in places where there aren't many dermatological specialists and resources are limited. Initial research indicates that AI-enhanced teledermatology systems can expedite referrals for high-risk lesions and diminish superfluous clinic visits **(Urbi et al., 2025).** In urban hospitals, AI might be an internal quality-control layer that lowers the difference between observers or finds lesions that were missed during busy clinical operations. In digital pathology, algorithms can assist pathologists in finding areas of interest more quickly, which can help them stay awake and make more accurate diagnoses. In the end, the best chance for progress is not to make self-sufficient diagnostic systems but to create long-lasting partnerships between humans and AI in the clinic. Dermatology and pathology are interpretive disciplines that necessitate the integration of pattern recognition with clinical judgment, patient history, and contextual knowledge. AI systems that enhance human expertise by standardizing thresholds, quantifying risk, emphasizing pertinent structures, or delivering consistent second readings are significantly more likely to be embraced than those intended to supplant physicians. As technical and organizational advancements increase, AI is set to evolve into a structured decision-support collaborator rather than a standalone adjudicator of malignancy **(Sazzad et al., 2025).** To make this vision a reality, we need to keep working on having a variety of training data, strict cross-institutional validation, clinically useful interpretability, open reporting of biases, and integration methodologies that take into account how clinical work actually works. It is also important to ensure that AI is used fairly so that it improves results for everyone instead of making existing differences worse. If these chances are taken advantage of in a planned way, AI can go from being an intriguing experimental tool to a dependable and morally sound part of global skin cancer therapy.

5. Conclusion

Artificial intelligence has made great strides in dermatologic oncology, showing that deep learning models can match or even beat expert performance in very controlled, limited settings. These accomplishments, especially

in dermoscopic imaging, validate that AI can discern diagnostically significant patterns from cutaneous lesions with a granularity that enhances human sensibility. The evidence from different imaging methods shows that outstanding accuracy on curated datasets does not mean that the system is ready for clinical use in general. Variations in equipment, skin tones, acquisition settings, and lesion epidemiology consistently cause models to deteriorate when tested in real-world scenarios. This means that lab and clinic result still differ greatly. Dermoscopy-based models are still the best at technical tasks, but they can't diagnose on their own because they use biased datasets and are sensitive to changes in the domain. Clinical-photograph AI algorithms, albeit less precise, provide considerable public health benefits by facilitating scaled triage and the prompt identification of high-risk cases in environments with little dermatological expertise. AI that focuses on histopathology and is backed by whole-slide imaging and more advanced architectures shows outstanding internal performance, but it is still limited by insufficient supervision, staining variability, and unmet needs for cross-institutional reproducibility. In all modalities, biased representation, lack of skin tone variety, noisy ground truth, limits on how simple it is to understand, and workflow misalignment still make it difficult to use these tools in everyday clinical practice.

Even with these problems, current research indicates that there is a way to meaningful integration in the future. Improvements in federated learning, domain generalization, and self-supervised representation learning could lead to models that stay stable across different groups of people and imaging settings. Parallel attempts to improve explainability, combine data from different sources, record how lesions change over time, and create stain-invariant histopathology models are breaking down long-standing barriers to clinician trust and regulatory acceptability. As these improvements develop, AI is becoming less of a replacement for dermatologists or pathologists and more of a structured decision-support tool that improves consistency, cuts down on unnecessary tests, and makes it easier for more people to undergo early evaluations. In its clinically advanced state, AI is expected to function across three tiers of the diagnostic continuum: scalable suspicion screening utilizing clinical photographs at the community interface, diagnostic enhancement via dermoscopic assessment in specialized environments, and confirmatory and prognostic enhancement within digital pathology processes. This kind of stratified model makes sure that AI helps rather than hinders in the way of clinical thinking and that its benefits correspond with the strengths and weaknesses of each imaging method. To make this vision a reality, we will need to keep working on making datasets more diverse, open, and validated by outside sources. Additionally, we must ensure their equitable use and alignment with authentic clinical workflows. If these prerequisites are met, AI-enabled diagnostic systems could significantly increase the early identification of melanoma, lessen worldwide inequalities in dermatologic treatment, and improve the accuracy and consistency of skin cancer diagnostic decision-making.

Funding

This work had no outside funding.

Author Contribution

The authors were involved in the creation of the study design, data analysis, and execution stages. Every writer

gave their consent after seeing the final work.

Acknowledgments

We would like to thank the authors of the reviewed articles

A statement of conflicting interests

The authors declare that none of the work reported in this study could have been impacted by any known competing financial interests or personal relationships.

5. References

Akhter, S., Ansari, M. A. S., Tiva, M. G., & Bhuyian, M. S. (2025). Improving Treatments for Oral Diseases, Head and Neck Cancers, as well as Developing New Technologies. Pathfinder of Research, 3(1), 1-25.

Brancaccio, G., Balato, A., Malvehy, J., Puig, S., Argenziano, G., & Kittler, H. (2023). Artificial intelligence in skin cancer diagnosis: A reality check. *Journal of Investigative Dermatology, 144*(3), 492–499. https://doi.org/10.1016/j.jid.2023.10.004

Brinker, T. J., Hekler, A., Enk, A. H., Klode, J., Hauschild, A., Berking, C., Schilling, B., Haferkamp, S., Schadendorf, D., Holland-Letz, T., Utikal, J. S., von Kalle, C., Ludwig-Peitsch, W., … Alamri, A. S. M. (2019). Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task. *European Journal of Cancer, 113*, 47–54. https://doi.org/10.1016/j.ejca.2019.04.001

Codella, N. C., Nguyen, Q. B., Pankanti, S., Gutman, D. A., Helba, B., Halpern, A. C., & Smith, J. R. (2017). Deep learning ensembles for melanoma recognition in dermoscopy images. *IBM Journal of Research and Development, 61*(4/5), 5-1. https://doi.org/10.1147/JRD.2017.2706689

DERMOFIT. (n.d.). University of Edinburgh Database. https://homepages.inf.ed.ac.uk/rbf/DERMOFIT/

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

Gajjar, S., Rathod, O., Joshi, D., Joshi, H., & Barot, V. (2025, July). A Hybrid ResNet-ViT Architecture for Skin Cancer Classification. In *2025 IEEE 4th World Conference on Applied Intelligence and Computing (AIC)* (pp. 31-36). IEEE. https://.doi.org/10.1109/AIC66080.2025.11212073

Gessert, N., Nielsen, M., Shaikh, M., Werner, R., & Schlaefer, A. (2020). Skin lesion classification using ensembles of multi-resolution EfficientNets with meta data. *MethodsX*, *7*, 100864. https://doi.org/10.1016/j.mex.2020.100864

Giavina-Bianchi, M., De Sousa, R. M., De Almeida Paciello, V. Z., Vitor, W. G., Okita, A. L., Prôa, R., … Machado, B. S. (2021). Implementation of artificial intelligence algorithms for melanoma screening in a primary care setting. *PLoS ONE, 16*(9), e0257006. https://doi.org/10.1371/journal.pone.0257006

Gohil, Z. M., & Desai, M. B. (2024). Revolutionizing dermatology: a comprehensive survey of AI-enhanced early

skin cancer diagnosis. *Archives of Computational Methods in Engineering*, *31*(8), 4521-4531. https://doi.org/10.1007/s11831-024-10121-7

Goyal, M., Knackstedt, T., Yan, S., & Hassanpour, S. (2020). Artificial intelligence-based image classification methods for diagnosis of skin cancer: Challenges and opportunities. *Computers in Biology and Medicine, 127*, 104065. https://doi.org/10.1016/j.compbiomed.2020.104065

Haenssle, H. A., Fink, C., Schneiderbauer, R., Toberer, F., Buhl, T., Blum, A., … Zalaudek, I. (2018). Man against machine: Diagnostic performance of a deep learning CNN for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Annals of Oncology, 29*(8), 1836–1842. https://doi.org/10.1093/annonc/mdy166

Han, S. S., Kim, M. S., Lim, W., Park, G. H., Park, I., & Chang, S. E. (2018). Classification of clinical images for benign and malignant cutaneous tumors using a deep learning algorithm. *Journal of Investigative Dermatology, 138*(7), 1529–1538. https://doi.org/10.1016/j.jid.2018.01.028

Li, Z., Koban, K. C., Schenck, T. L., Giunta, R. E., Li, Q., & Sun, Y. (2022). Artificial intelligence in dermatology image analysis: current developments and future trends. *Journal of clinical medicine*, *11*(22), 6826. https://doi.org/10.3390/jcm11226826

Liopyris, K., Gregoriou, S., Dias, J., & Stratigos, A. J. (2022). Artificial intelligence in dermatology: Challenges and perspectives. *Dermatology and Therapy, 12*(12), 2637–2651. https://doi.org/10.1007/s13555-022-00833-8

Liu, Q., Zhang, J., & Bai, Y. (2023). Mapping the landscape of artificial intelligence in skin cancer research: a bibliometric analysis. *Frontiers in oncology*, *13*, 1222426. https://doi: 10.3389/fonc.2023.1222426

Maron, R. C., Weichenthal, M., Utikal, J. S., Hekler, A., Berking, C., Hauschild, A., … Glutsch, V. (2019). Systematic outperformance of 112 dermatologists in multiclass skin cancer image classification by convolutional neural networks. *European Journal of Cancer, 119*, 57–65. https://doi.org/10.1016/j.ejca.2019.06.013

Martínez-Vargas, E., Mora-Jiménez, J., Arguedas-Chacón, S., Hernández-López, J., & Zavaleta-Monestel, E. (2025). The emerging role of artificial intelligence in dermatology: A systematic review of its clinical applications. *Dermato, 5*(2), 9. https://doi.org/10.3390/dermato5020009

Melarkode, N., Srinivasan, K., Qaisar, S. M., & Plawiak, P. (2023). AI-powered diagnosis of skin cancer: a contemporary review, open challenges and future research directions. *Cancers*, *15*(4), 1183. https://doi.org/10.3390/cancers15041183

Mendonça, T., Ferreira, P. M., Marques, J. S., Marcal, A. R. S., & Rozeira, J. (2013). PH2 – A dermoscopic image database for research and benchmarking. *EMBC Conference Proceedings*, 5437-5440. https://doi.org/10.1109/EMBC.2013.6610779

Mukherjee Behara, K., Bhero, E., & Agee, J. T. (2024). AI in dermatology: a comprehensive review into skin cancer detection. *PeerJ Computer Science*, *10*, e2530. https://doi.org/10.7717/peerj-cs.2530

Mukherjee, S., Rao, S. R., & Poddar, A. (2025). Artificial intelligence powered mobile health apps for skin cancer detection: current challenges and a systems thinking approach for improved public health outcomes in low-and middle-income countries. Melanoma Research, 10-1097. https://doi: 10.1097/CMR.0000000000001074

Salaheldin, A. M., Wahed, M. A., Talaat, M., & Saleh, N. (2024). An evaluation of AI-based methods for papilledema detection in retinal fundus images. *Biomedical Signal Processing and Control, 92*, 106120. https://doi.org/10.1016/j.bspc.2024.106120

Salinas, M. P., Sepúlveda, J., Hidalgo, L., Peirano, D., Morel, M., Uribe, P., … Navarrete-Dechent, C. (2024). AI versus clinicians for skin cancer diagnosis: A systematic review and meta-analysis. *NPJ Digital Medicine, 7*(125). https://doi.org/10.1038/s41746-024-01103-x

Santoro, C. (2023). Review examines pros and cons in AI detection of skin cancer. *AJMC – Managed Care News*. https://www.ajmc.com/view/review-examines-pros-and-cons-in-ai-detection-of-skin-cancer

Sazzad, S. A., Chowdhury, R., Hasan, M. R., Tiva, M. G., Rahman, K., Ansari, M. A. S., & Sunny, A. R. (2025). Public Health, Risk Perception, and Governance Challenges in the 2025 Los Angeles Wildfires: Evidence from a Community-Based Survey. *Pathfinder of Research*, *3*(1), 26-41.

Shetty, B., Fernandes, R., Rodrigues, A. P., Chengoden, R., Bhattacharya, S., & Lakshmanna, K. (2022). Skin lesion classification of dermoscopic images using machine learning and convolutional neural network. *Scientific Reports*, *12*(1), 18134. https://doi.org/10.1038/s41598-022-22644-9

Team, Scientific Data Curation (2021). Metadata record for: A patient centric dataset of images and metadata for identifying melanomas using clinical context. figshare. Dataset. https://doi.org/10.6084/m9.figshare.13070345.v1

Tiva, M. G., Tarin, N. N., Hasan, M. R. K., Urbi, S. R. C., & Sazzad, S. A. (2025a). Post-COVID-19 Work force Management in US Healthcare: Burnout, Retention, and Strategies for Enhancing Cultural Competency. Pathfinder of Research, 3(1), 98-119.

Tiva, M. G., Urbi, S. R. C. & Hasan, M. R. K. (2025b). Preventable Readmissions and Financial Management in Healthcare: A Comparative Study of Cost Containment in Non-Profit and For-Profit Hospitals. Pathfinder of Research,*3*(2), 1-21

Tschandl, P., Codella, N., Akay, B. N., Argenziano, G., Braun, R. P., … Kittler, H. (2019). Expert-level diagnosis of nonpigmented skin cancer by combined CNNs. *JAMA Dermatology, 155*(1), 58–65. https://doi.org/10.1001/jamadermatol.2018.4378

Tschandl, P., Rosendahl, C., Akay, B. N., Argenziano, G., Blum, A., Braun, R. P., Cabo, H., Gourhant, J. Y., Kreusch, J., Lallas, A., Lapins, J., Marghoob, A., Menzies, S., Neuber, N. M., Paoli, J., Rabinovitz, H. S., Rinner, C., Scope, A., Soyer, H. P., Sinz, C., … Kittler, H. (2019). Expert-Level Diagnosis of Nonpigmented Skin Cancer by Combined Convolutional Neural Networks. *JAMA dermatology*, *155*(1), 58–65. https://doi.org/10.1001/jamadermatol.2018.4378

Urbi, S. R. C., & Tiva, M. G. (2025). Technology and Innovation in Healthcare: Adoption of AI and Predictive Analytics in Hospital Management. *Pathfinder of Research*, *3*(2), 22-45.

Winkler, J. K., Blum, A., Kommoss, K., Enk, A., Toberer, F., Rosenberger, A., & Haenssle, H. A. (2023). Assessment of diagnostic performance of dermatologists cooperating with a convolutional neural network in a prospective clinical study. *JAMA Dermatology*, *159*(6), 621. https://doi.org/10.1001/jamadermatol.2023.0905

Wu, Y., Chen, B., Zeng, A., Pan, D., Wang, R., & Zhao, S. (2022). Skin cancer classification with deep learning: A systematic review. *Frontiers in Oncology, 12*, 893972. https://doi.org/10.3389/fonc.2022.893972

Zhang, L., Yang, G., & Ye, X. (2019). Automatic skin lesion segmentation by combining deep convolutional networks and shallow networks. *Journal of Medical Imaging, 6*(2), 024001. https://doi.org/10.1117/1.jmi.6.2.024001