REVIEW

Editorial Process: Submission:12/04/2025 Acceptance:07/18/2026 Published:07/27/2026

Artificial Intelligence in Skin Cancer Detection for Primary Care: A Review

Saeb Zakout[1], Ihab Zaqout[2,3]*

Abstract

Background: Skin cancer is the most common malignancy worldwide and can be lethal if not detected early, especially melanoma. Primary care is often the first point of contact, but limited dermoscopy expertise and increasing service pressure can delay diagnosis. Artificial intelligence (AI) offers a route to expand access to earlier detection by supporting assessment and triage decisions in primary care. **Methods:** We conducted a narrative review of peer-reviewed literature on AI systems for skin-lesion analysis, including traditional machine-learning pipelines, convolutional neural networks (CNNs), vision transformers, ensemble and hybrid models, and mobile/on-device tools. We prioritized studies that compared AI with clinicians, reported diagnostic performance and calibration, evaluated subgroup performance (e.g., skin tone and device), and examined implementation in primary care, teledermatology, or real-world outpatient workflows. **Results:** AI systems can achieve high sensitivity for melanoma and other malignant lesions on curated image sets, often approaching specialist-level discrimination. Reader studies and pragmatic evaluations suggest AI assistance can improve non-dermatologists' sensitivity and diagnostic confidence, but specificity, calibration, and robustness across devices and under-represented populations remain variable. Recent deployment-focused evidence highlights risks of automation bias, alert fatigue, and inequitable performance, underscoring the need for careful threshold selection, external validation, and post-deployment monitoring. **Conclusion:** AI-enabled skin cancer detection can support earlier diagnosis and more efficient triage in primary care, but clinical benefit depends on evaluation beyond headline accuracy. Diverse training data, robust external and prospective validation, explicit calibration and equity reporting, and governance frameworks with human oversight and continuous monitoring are essential for safe implementation.

Keywords: Skin cancer- artificial intelligence- primary care- deep learning- health equity

Asian Pac J Cancer Prev, **27 (7)**, 2437-2445

Introduction

Skin cancer is the most common malignancy worldwide and, for melanoma in particular, survival depends on early detection. Primary-care services are the first point of contact for most patients, yet growing demand, limited dermoscopy expertise and long waits can delay diagnosis. Artificial intelligence (AI) is now being evaluated in frontline settings to augment clinician judgement, streamline triage, and standardize assessment quality [1-4].

However, evidence is fragmented and unevenly reported. Questions persist about external validity across devices and sites, calibration of probability outputs, performance on under-represented skin tones and populations, and the real-world impact of AI assistance on clinician decisions and referral patterns [5-8]. Heterogeneous study designs and metrics make it difficult for primary-care decision-makers to judge what is ready for responsible deployment.

This review aims to orient clinicians and implementers quickly by providing concise synthesis. Our objectives are to: (i) map the main classes of AI systems for skin lesion analysis and where they sit in the clinical pathway, (ii) synthesize diagnostic accuracy from reader studies, meta-analyses and prospective evaluations with attention to calibration and equity, and (iii) translate the evidence into practical guidance for workflow integration and service design.

We focus on AI systems that analyze clinical or dermoscopic photographs to detect or triage pigmented skin lesions in outpatient and primary-care contexts. We adopt a human–AI collaboration perspective and organize the synthesis around a pragmatic evaluation framework balancing discrimination, calibration, external validation and subgroup equity, consistent with best-practice reporting guidance [3, 9, 10].

*[1]Zakout hudklinikk AS, Kongens gate 30, 4610 Kristiansand, Norway. [2]Department of Information Technology, Faculty of Engineering and Information Technology, Al-Azhar University, Gaza, Palestine. [3]School of Electronics and Computer Science, Faculty of Engineering and Physical Sciences, University of Southampton, SO17 1BJ, UK. *For Correspondence: i.zaqout@alazhar.edu.ps*

Materials and Methods

This narrative review synthesized evidence on AI-based skin-lesion assessment with an emphasis on primary-care workflows. We identified relevant peer-reviewed studies by targeted searches and citation-chaining across key reviews and high-impact evaluations. Evidence was summarized by model family (traditional ML, CNNs, transformers, ensemble/hybrid, and mobile/on-device) and by evaluation maturity (curated reader studies, external validation, prospective/pragmatic trials, and real-world deployment). Particular attention was given to clinically meaningful operating thresholds, calibration of probability outputs, subgroup performance (e.g., skin tone and device), and implementation risks (automation bias, referral burden, and monitoring requirements).

AI Model Architectures for Skin Cancer Detection

AI methods for skin lesion analysis range from traditional machine-learning (ML) pipelines using hand-crafted features to end-to-end deep learning. Figure 1 contrasts the classical pipeline with modern deep-learning pipelines. Advanced approaches include ensembles, hybrid CNN–Transformer models, multi-input networks that incorporate metadata, and pipelines that combine segmentation and classification [11-14]. Comprehensive reviews have catalogued these machine-learning tools and deep neural network approaches for skin cancer recognition and highlighted key challenges and future directions for the field [15].

Traditional machine-learning approaches

Strengths of these methods are interpretability of engineered cues and robustness on small datasets. However, they often underperform on large, diverse corpora, are sensitive to segmentation errors and illumination artefacts, and struggle with cross-device generalization. Performance typically plateaus without extensive feature engineering and careful threshold tuning. Public datasets like HAM10000 and ISIC accelerated benchmarking but also revealed domain shift and label noise issues that favor deep learning approaches [9, 16].

Traditional pipelines typically begin with image pre-processing (illumination correction, hair removal) and lesion segmentation, followed by hand-crafted descriptors. Common features include ABCD rule cues (asymmetry, border irregularity, color variegation, diameter), color histograms and ratios, geometric and fractal measures, and texture descriptors such as LBP and Gabor filters. Classifiers have included SVMs (often with RBF kernels), random forests and gradient boosting [11, 17].

Deep learning with convolutional neural networks (CNNs)

Important practicalities include class imbalance (addressed with focal loss, reweighting or resampling), calibration (temperature scaling, Platt/Isotonic), and external validation across devices and care settings. Ensembles of heterogeneous backbones or differently seeded replicas can improve recall and stability [13, 18]. Incorporating context either total-body photography crops or clinical metadata, often boosts generalization [19, 20]. Prospective studies and smartphone-based evaluations show feasibility in primary care and tele-dermatology workflows [21-23].

CNNs learn hierarchical representations directly from pixels and have become the dominant approach for lesion classification [24, 25]. Modern practice relies on transfer learning from ImageNet, aggressive augmentation (flips/rotations, color jitter, random crops, mixup/cutout), and regularization (dropout, label smoothing). Backbones such as ResNet, DenseNet, EfficientNet and MobileNet are common starting points.

Vision transformers and hybrid models

Explainability tools Grad-CAM, attention roll-out and counterfactuals can highlight model focus, but explanations may be unstable across seeds and augmentations. Care is needed to avoid over-interpreting saliency: quantitative trust should come from calibration, external validation and subgroup analysis rather than heat-maps alone [26, 27].

Vision transformers (ViTs) replace convolution with self-attention over patch embeddings, capturing long-range

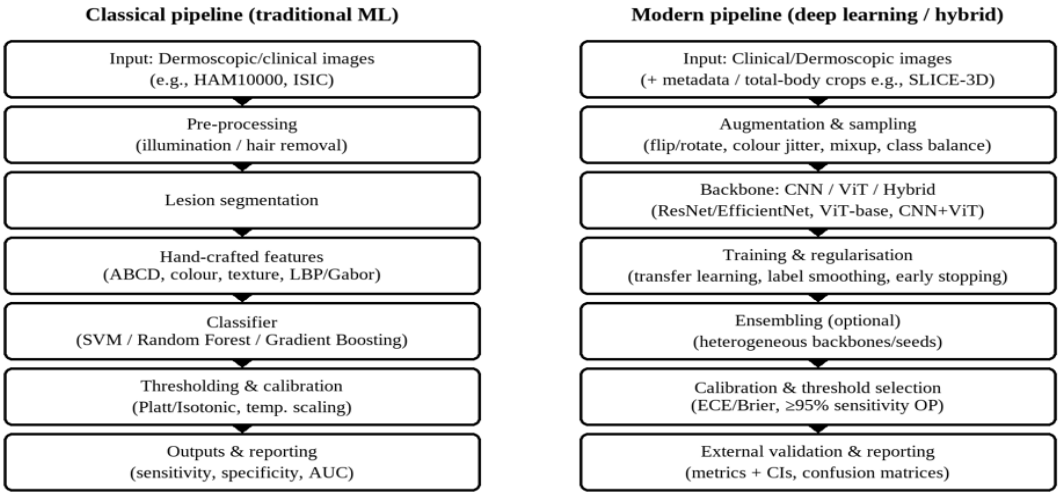

Figure 1. AI Skin Cancer Detection Pipelines (Classical vs Deep Learning). Abbreviations: ML, machine learning; CNN, convolutional neural network; ViT, vision transformer; ECE, expected calibration error; OP, operating point.

structure and color patterns relevant to dermoscopic criteria. When sufficient data and augmentation are available, ViTs can match or surpass CNNs, and hybrids fuse CNN feature pyramids (local texture/edges) with transformer blocks (global context) to leverage both inductive biases [14, 28, 29].

Lightweight models and mobile deployment

Operational considerations include camera guidance (focus, lighting, ruler/color card), device-to-device calibration transfer, and fail-safes for out-of-distribution inputs. User-centered interfaces can standardize acquisition and reduce retakes [30]. Real-world studies demonstrate viability of smartphone capture and AI-assisted triage in primary care and tele-dermatology [4, 23, 31, 32].

Primary-care deployment emphasizes speed, privacy and reliability. Lightweight architectures (MobileNet/EfficientNet-lite), pruning, quantization and knowledge distillation enable on-device inference with low latency. On-device models reduce dependence on network connectivity and mitigate privacy risk, whereas cloud inference can simplify updates and centralize monitoring, and hybrid designs are common. Table 1 summarizes representative methods and example studies.

Diagnostic Accuracy of AI vs. Clinicians

Comparative evidence comes from reader studies, meta-analyses and prospective clinical trials. Table 2

Table 1. Representative AI Approaches for Skin Lesion Analysis and Example Studies

Approach	Typical workflow/description	Example studies (as cited in this review)
Traditional ML (feature-based)	Pre-processing → segmentation → hand-crafted features (ABCD, color, texture) → classical classifier (SVM/RF/GB).	Bansal et al. (hand-crafted + classifier) [11]; Debelee et al. (survey) [17]
CNN deep learning	End-to-end convolutional networks with transfer learning and augmentation; may incorporate metadata; often calibrated post hoc.	Esteva et al. [24]; Wu et al. (review) [25]
Ensemble learning	Combine multiple backbones/seeds/losses for improved robustness; can tune for high-sensitivity triage.	Gessert et al. [18]; PROVE-AI (Marchetti et al.) [22]
Vision transformers	Self-attention over patch embeddings; can capture global structure; often needs strong augmentation/large data.	Dagnaw et al. [28]; Khan et al. (scoping review) [29]
Hybrid CNN–Transformer	Fuse CNN (local texture) with ViT (global context) using late or feature-level fusion.	Sarıateş & Özbay [14]
Context/metadata-aware models	Combine lesion image with patient metadata or total-body context to improve generalization.	Primiero et al. [20]; Kurtansky et al. (SLICE-3D) [19]
Segmentation-assisted pipelines	Segmentation used as preprocessing or multi-task regularizer to guide classification.	Kawahara et al. [12]
Explainable AI	Saliency/attention/counterfactual explanations to support audit and user trust (requires careful validation).	Han et al. [26]; Chanda et al. [27]
Mobile/on-device models	Lightweight backbones, pruning/quantization/distillation for privacy-preserving, low-latency inference on smartphones.	Gregoor et al. [21]; Khan et al. (distillation) [40]

Table 2. Core Performance Metrics Used for AI Skin-Lesion Evaluation

Metric	Definition (plain language)	Practical notes for primary care
Sensitivity (recall)	Proportion of true cancers correctly identified.	Often prioritize high sensitivity for triage; report the operating point used.
Specificity	Proportion of benign lesions correctly classified as benign.	Low specificity can overload referrals; report expected referral impact.
AUC (ROC)	Probability the model ranks a random cancer higher than a random benign case.	Summarizes discrimination but not calibration; report confidence intervals where possible.
PPV / NPV	Probability that a positive/negative result is truly malignant/benign.	Highly prevalence-dependent; useful for referral pathway planning.
Balanced accuracy	Average of sensitivity and specificity.	More stable under class imbalance than raw accuracy.
Brier score	Mean squared error of predicted probabilities.	Lower is better; captures both discrimination and calibration.
Expected calibration error (ECE)	Average difference between predicted probability and observed frequency across bins.	Report alongside a reliability plot; calibration slope/intercept can add clarity.
F1 score	Harmonic mean of precision and recall.	Useful for imbalanced classes, but less directly interpretable for triage than sens/spec.
Net benefit (decision-curve analysis)	Clinical utility across thresholds compared with treat-all/treat-none strategies.	Helps choose thresholds aligned with service capacity and harms of false positives/negatives.

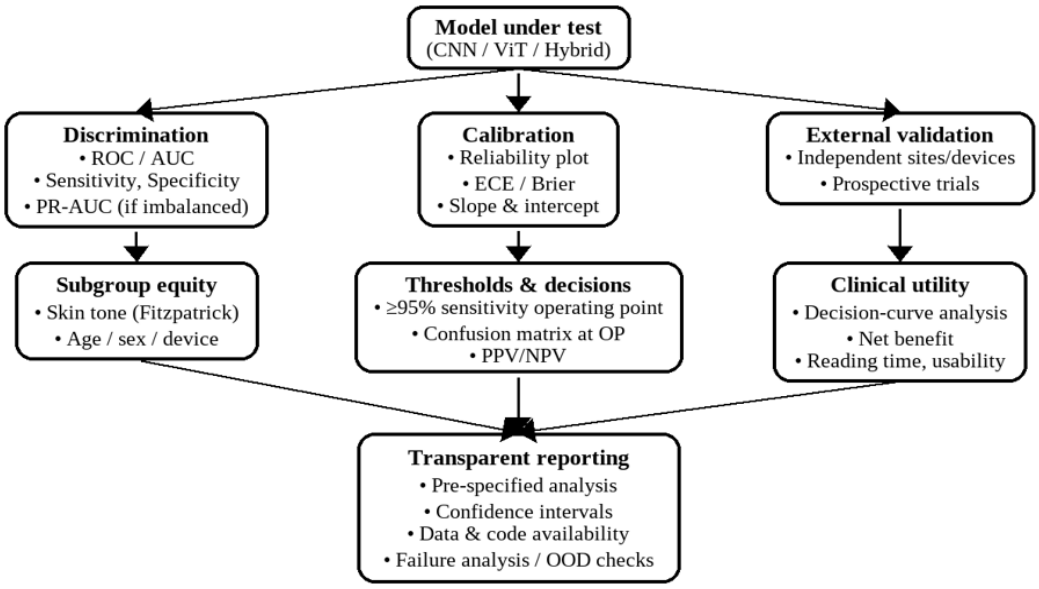

Figure 2. Evaluation Framework for AI Skin Lesion Models. Footnote: Recommended minimum reporting includes discrimination, calibration, external validation, subgroup equity, and clinical utility at a pre-specified operating point.

defines the core performance metrics used throughout this review, and Figure 2 outlines a practical evaluation framework for fair model assessment. Key metrics include sensitivity, specificity, area under the ROC curve (AUC) and calibration (e.g., Brier score, Expected Calibration Error [ECE]).

At first mention, sensitivity (true-positive rate) reflects the proportion of cancers detected, while specificity reflects the proportion of benign lesions correctly ruled out. AUC summarizes discrimination across thresholds but does not indicate whether predicted probabilities are well calibrated. Calibration describes how closely predicted risks match observed outcomes and can be summarized using reliability plots and metrics such as the Brier score and expected calibration error (ECE).

Head-to-head comparisons with dermatologists

Prospective assistance trials, closer to real practice, suggest AI can improve sensitivity and/or confidence while maintaining acceptable specificity, but raise issues of automation bias, alert fatigue and workflow fit [8, 10]. Alongside these assistance studies, Han et al. [33] reported the clinical performance of a novel artificial intelligence-based system for melanoma detection in a real-world setting. Best practice is to pre-specify decision thresholds, measure reading times and inter-rater agreement, and report calibration alongside discrimination to avoid over-referral at high-sensitivity operating points.

Reader studies typically compare unaided clinicians to an AI model on curated image sets under controlled conditions, and/or measure how clinicians change decisions with AI assistance. Seminal work showed CNNs matching or exceeding dermatologists on dermoscopic images [34, 35]. More recent experiments focus on specific triage decisions (e.g., melanocytic vs non-melanocytic) and report very high AUCs [36]. Complementing these reader studies, Yazdanparast et al. [37] compared face-to-face clinical examination, tele-dermoscopy and an AI assistant for melanoma diagnosis, finding that teledermoscopy achieved the highest specificity and agreement with histopathology while the AI tool had the lowest diagnostic accuracy.

Systematic reviews and meta-analyses

Robust reporting should include external validation, lesion-level metrics at clinically relevant thresholds, calibration measures (ECE/Brier), confusion matrices and decision-curve analysis to quantify net benefit [9]. Stratified analyses by skin tone and device are essential to surface disparities and guide mitigation [7]. Open-source validations such as PROVE-AI and trend analyses provide additional transparency into maturity and adoption [22].

Meta-analyses pool evidence across heterogeneous datasets and designs. Pooled sensitivity for AI systems typically lies in the high-80% range with specificity in the mid-70s, similar to expert readers, though heterogeneity is substantial [6]. Sources of heterogeneity include image modality (clinical vs dermoscopic), case mix, reference standard (histopathology vs follow-up), and thresholding strategies.

Table 3 summarizes the AI methods referenced in this study, grouped by single-, multi-, fusion/ensemble-, and hybrid-model categories. Additionally, includes traditional ML, segmentation and classification pipelines, multi-input/context-aware models, and lightweight/on-device architectures described in Section 2. Metrics are reported as in the cited works, but due to dataset and threshold differences, values should not be directly compared across studies.

Primary Care Implementation and Global Trends

Section 5 synthesizes practical considerations for adoption. Subsection 5.1 examines clinical deployment pathways, including regulatory status, intended-use statements, training, and post-market monitoring, while 5.2 reviews global research and adoption trends with an emphasis on equity. Figure 3 maps a typical primary-care

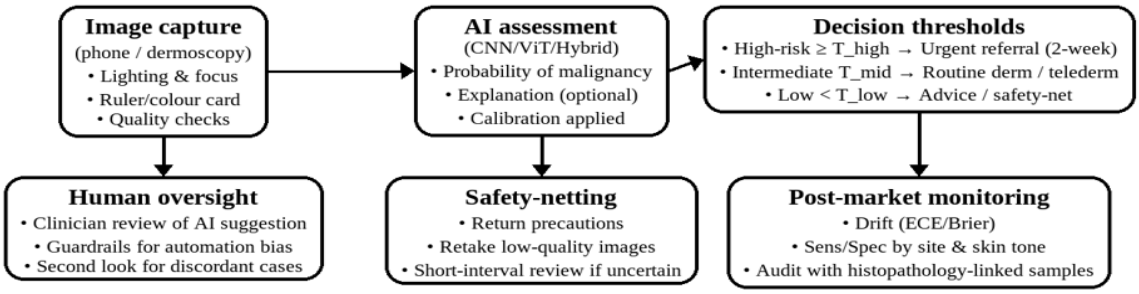

Figure 3. Primary Care Workflow Integrating AI Triage. Footnote: Example workflow showing how image quality, AI assessment, thresholds, human oversight, and post-market monitoring interact in primary care.

workflow showing where AI can assist triage, referral, and safety-netting.

From trials to practice

Recent regulatory and service-level developments demonstrate clinical feasibility of AI in primary care pathways. In 2024, the U.S. Food and Drug

Table 3. Selected Evidence on AI Performance and Clinical Applicability (Examples Cited in This Review)

Study / AI type	Target lesion (s) / task	Setting & image modality	Comparator / design	Key results (as reported)	Primary-care applicability
Haenssle et al. (CNN) [34]	Melanoma recognition	Curated dermoscopy reader study (100 images)	CNN vs 58 dermatologists	Sensitivity 88.9%; specificity 82.5%; AUC 0.86	Demonstrates specialist-level discrimination on curated images; needs calibration and external validation for primary care.
Esteva et al. (deep CNN) [24]	Skin cancer classification (multiple benign/ malignant)	Curated dermoscopy + clinical images	Model vs dermatologists (reader study)	Dermatologist-level performance reported	Seminal proof-of-concept; generalization to diverse devices/ populations remains key.
Marchetti et al. PROVE-AI (ensemble) [22]	Melanoma detection	Prospective, multi-site dermoscopy (603 lesions; 95 melanomas)	Open-source AI validated prospectively	Sensitivity 96.8%; specificity 37.4%; AUC 0.857 at pre-specified high-sensitivity point	Shows feasibility of high-sensitivity triage; specificity may increase referrals—threshold choice is central.
Gregoor et al. (smartphone AI) [21]	Skin lesion diagnosis / triage thresholds	Real-world smartphone clinical images (primary care relevant)	Prospective evaluation of automated smartphone diagnosis	Accuracy reported with "meaningful-use" thresholds	Directly relevant to primary care image capture; emphasizes operating thresholds and workflow fit.
Menzies et al. (mobile phone AI) [23]	Pigmented lesion diagnosis and management	Multicentre prospective trial (clinical photos)	Humans vs mobile phone AI; novice vs specialist	Novices improved with AI; management accuracy slightly lower than specialists	Highlights decision-support potential for non-dermatologists; management recommendations need careful governance.
Papachristou et al. (decision-support app) [2]	Melanoma detection support	Primary care evaluation	Clinical evaluation of AI-based app	Diagnostic performance reported in primary care	Evidence closest to intended use; supports need for training and monitoring.
Marsden et al. (AI medical device) [38]	Triage on urgent suspected skin cancer pathway	UK teledermatology service images	Real-world evaluation of AI-as-a-medical-device	Sensitivity 91–92.5%; specificity 85.1	Service-level evidence; illustrates post-market surveillance and pathway integration.
Heinlein et al. (AI assistance) [8]	Dermoscopy-aided melanoma diagnosis	Prospective multicenter patient-care setting	AI assistance vs usual care	Improved diagnostic performance reported	Shows benefit of human–AI collaboration; workflow and automation-bias mitigation required.
Daneshjou et al. (equity study) [7]	Melanoma detection CNN	Multi-racial evaluation across skin tones	Subgroup performance assessment	Performance disparities across skin tones reported	Directly informs equity monitoring and dataset representativeness for primary care deployment.
Salinas et al. (meta-analysis) [6]	AI vs clinicians (skin cancer)	Pooled across heterogeneous datasets	Systematic review/meta-analysis	Pooled sensitivity 87.0%; specificity 77.1 (heterogeneity high)	Provides benchmark ranges; underscores heterogeneity and reporting gaps relevant to implementers.

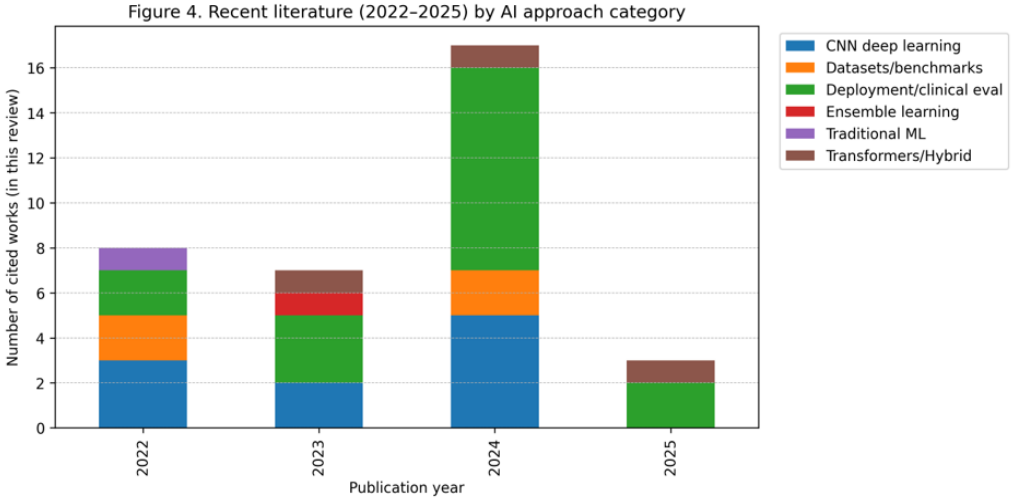

Figure 4. Histogram of Recent Research by ML Approach (2022–2025). Footnote: Counts are based on the studies discussed/cited in this review and are intended to summarize recent methodological direction rather than provide a systematic bibliometric census.

Administration (FDA) authorized a handheld spectroscopy device intended to support melanoma triage by non-dermatologists, and the UK National Health Service (NHS) has evaluated AI-enabled tele-dermatology triage with ongoing monitoring [4, 38, 39, 41, 42]. Key principles for safe integration include clearly defined intended use, clinician oversight, training, and continuous post-market performance monitoring, including evaluation across under-represented populations. As shown in Figure 3, AI outputs should feed into existing referral pathways with thresholds tuned for high sensitivity and with explicit safety-netting for equivocal cases.

Global research and adoption

Recent publication patterns (2022–2025) show a growing emphasis on deep learning and transformer-based approaches for lesion classification and triage (Figure 4). Across the studies discussed in this review, commonly used pretrained backbones include EfficientNet and MobileNet variants for CNN-based pipelines and Vision Transformer (ViT) families for transformer pipelines (Figure 5). Most AI evaluations target melanoma detection; however, primary-care triage typically encounters a broader case-mix including basal cell carcinoma (BCC), squamous cell carcinoma (SCC), and benign mimickers. Where models are trained on multi-class datasets, performance should be reported separately for melanoma vs keratinocyte cancers (BCC/SCC) and for benign lesions, because these distinctions affect referral thresholds and workload in primary care.

Documented performance gaps across skin tones and devices underscore the need for demographically representative training data, prospective subgroup reporting, and governance that mandates bias monitoring [3, 7]. Bibliometric analyses show accelerating output but also concentration in a handful of regions, suggesting opportunities for capacity building and data partnerships [43].

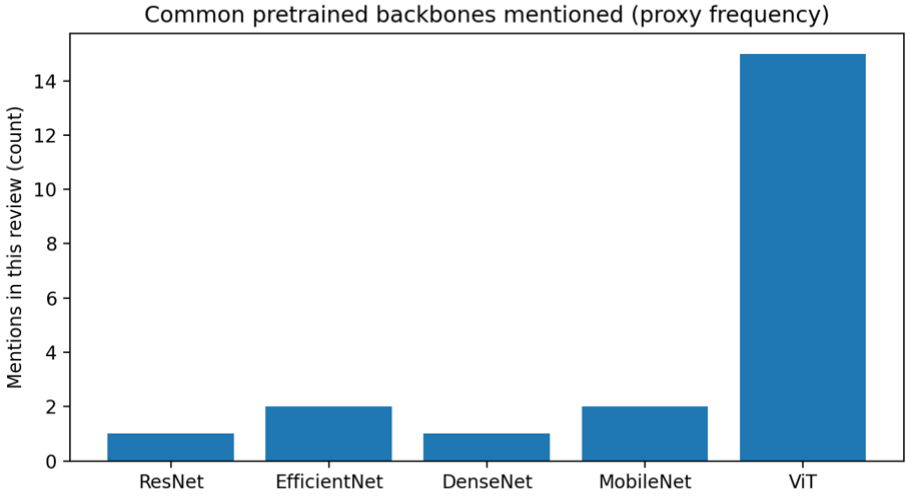

Figure 5. Common Pretrained Backbones Mentioned in This Review (Proxy Frequency). Footnote: Bars represent the number of times each backbone family is mentioned in the manuscript text; this is a proxy indicator and may not reflect true prevalence across all published studies.

Global research capacity has expanded with public datasets (ISIC challenges, PAD-UFES-20) and multi-site collaborations that enable benchmarking and external validation (International Skin Imaging Collaboration [44, 45]. High-income countries lead deployment pilots, often via tele-dermatology hubs, while low- and middle-income settings may benefit most from mobile, on-device AI coupled with remote expert review.

Regulatory and governance frameworks (e.g., NIHCE evaluations, FDA authorizations) increasingly emphasize claims tied to real-world intended use, human oversight and bias mitigation plans [4, 42]. Pragmatic trials in primary care and smartphone-based studies inform threshold selection and operational integration [1, 21].

Post-market monitoring is essential to detect calibration drift and performance changes after software updates or device mix shifts. Programmes should include periodic audits with histopathology-linked samples, dashboarding of sensitivity/specificity by site and skin tone, and decision-curve analyses to ensure net benefit remains positive as prevalence and referral flows change [38, 41]. Successful service adoption requires more than accuracy. Deployment plans should define the intended user and decision, referral thresholds aligned to service capacity, and safety-netting for uncertain or low-quality images. Training and competency check for image capture (lighting, focus, rulers/color cards) reduce avoidable retakes and false alarms [30].

Service design, monitoring and equity safeguards

Building on the points above, safe and effective adoption of AI skin-lesion triage in primary care requires a structured plan that links clinical workflow, thresholds, data quality, monitoring and governance. The following steps summarize a pragmatic blueprint derived from evaluations and guidance [1, 4, 21, 38, 41, 42].

1. Define success metrics: beyond ROC, track cancer detection rate, time to dermatologist, avoidable referrals, stage at diagnosis, and patient-reported outcomes. Review against baseline after 3–6 months.

2. Training and competency: provide brief scenario-based training and refreshers, certify competencies for image capture and AI-assisted decision steps, and collect user feedback in-app.

3. Privacy and security: prefer on-device inference where feasible. For cloud workflows ensure encryption, access controls and minimal retention, and record only metadata required for audit and model improvement.

4. Change management: Robust deployment of AI-based clinical systems requires systematic versioning of models, datasets, and decision thresholds, together with impact-limited rollout strategies (e.g., shadow mode or A/B evaluation) supported by predefined go/no-go criteria. Maintaining a documented safety case and release notes is also essential to ensure traceability, accountability, and controlled updates. Incremental learning approaches for skin lesion classification, such as the dynamically expandable representation framework proposed by Jang and Park [46], demonstrate how models can incorporate new data and extend their classification capabilities while maintaining performance across evolving datasets. However, continuous monitoring and governance mechanisms are still required to manage potential performance drift during real-world deployment.

5. Monitoring and audits: implement dashboards for volume, sensitivity/specificity, PPV/NPV, ECE/Brier, turnaround time and referral flows, and schedule periodic histopathology-linked audits to detect drift [38, 41].

6. Equity safeguards: monitor sensitivity/specificity and calibration by Fitzpatrick skin tone, age/sex and device; adopt conservative defaults where uncertainty is higher [3, 7].

7. Guard against automation bias: design the UI so AI suggestions are collapsible, and justification cues are shown after an initial clinician impression and require acknowledgement on high-risk overrides [10].

8. Calibrate and validate locally: apply simple post-hoc calibration (e.g., temperature scaling) on local data when feasible and verify discrimination and calibration across devices and sites [9].

9. Set thresholds to service capacity: prespecify sensitivity first operating points, tuned to local prevalence and referral slots, and documented expected changes in urgent/routine referrals and review intervals.

10. Standardize image acquisition: provide quick guides and in-app prompts for lighting, focus, scale/ruler and color card, and enable immediate quality checks and retakes [30].

11. Map the pathway and failure modes: draw the end-to-end flow (capture → AI → thresholds → referral/safety-net) and pre-agree responses to low-quality images, out-of-distribution cases, and discordant AI–clinician judgements.

12. Define intended use, user, and decision: specify who uses the tool (e.g., GP, nurse), on what images (clinical or dermoscopic), and for which decision (e.g., urgent two-week wait vs routine vs reassurance).

Putting these pieces together ensures that accuracy translates into clinical benefit while minimizing unintended harms. Real-world studies and evaluations emphasize the importance of predefined thresholds, calibration and subgroup monitoring, with human oversight and post-market surveillance embedded from day one [4, 21, 38, 41, 42].

In conclusion, AI for skin cancer detection has advanced from classical feature-based methods to deep learning systems that can rival expert clinicians in image-based diagnosis. Validated prospectively, AI demonstrates high sensitivity for melanoma detection and can augment primary-care triage. Translating laboratory accuracy into patient benefit requires attention to generalization (across devices, populations and skin tones), rigorous external and prospective validation, and integration within clinical workflows with ongoing monitoring. Future work should prioritize diverse training data, robust real-world trials, explainability, calibration, and clear regulatory pathways. With careful deployment and oversight, AI can support equitable, timely detection and improved outcomes.

Author Contribution Statement

Saeb Zakout: conceptualization, clinical framing,

literature screening, writing review and editing. Ihab Zaqout: review design, literature screening, writing original draft (technical sections), writing review and editing. Both authors approved the final manuscript.

Acknowledgements

Ethics approval

Not applicable. This is a narrative review and did not involve human participants, animals, or identifiable patient data. Therefore, no ethics committee approval was required.

Conflict of interest

The authors declare no conflict of interest.

References

1. Jones OT, Matin RN, Walter FM. Using artificial intelligence technologies to improve skin cancer detection in primary care. Lancet Digit Health. 2025;7(1):e8-e10. https://doi.org/10.1016/s2589-7500(24)00216-4.
2. Papachristou P, Söderholm M, Pallon J, Taloyan M, Polesie S, Paoli J, et al. Evaluation of an artificial intelligence-based decision support for the detection of cutaneous melanoma in primary care: A prospective real-life clinical trial. Br J Dermatol. 2024;191(1):125-33. https://doi.org/10.1093/bjd/ljae021.
3. Krakowski I, Kim J, Cai ZR, Daneshjou R, Lapins J, Eriksson H, et al. Human-ai interaction in skin cancer diagnosis: A systematic review and meta-analysis. NPJ Digit Med. 2024;7(1):78. https://doi.org/10.1038/s41746-024-01031-w.
4. National Institute for Health and Care Excellence (NICE). Artificial intelligence (AI) technologies for assessing and triaging skin lesions referred to the urgent suspected skin cancer pathway: early value assessment. 2025. Available from: https://www.nice.org.uk/guidance/hte24
5. Choy SP, Kim BJ, Paolino A, Tan WR, Lim SML, Seo J, et al. Systematic review of deep learning image analyses for the diagnosis and monitoring of skin disease. NPJ Digit Med. 2023;6(1):180. https://doi.org/10.1038/s41746-023-00914-8.
6. Salinas MP, Sepúlveda J, Hidalgo L, Peirano D, Morel M, Uribe P, et al. A systematic review and meta-analysis of artificial intelligence versus clinicians for skin cancer diagnosis. NPJ Digit Med. 2024;7(1):125. https://doi.org/10.1038/s41746-024-01103-x.
7. Daneshjou R, Vodrahalli K, Novoa RA, Jenkins M, Liang W, Rotemberg V, et al. Disparities in dermatology ai performance on a diverse, curated clinical image set. Sci Adv. 2022;8(32):eabq6147. https://doi.org/10.1126/sciadv.abq6147.
8. Heinlein L, Maron RC, Hekler A, Haggenmüller S, Wies C, Utikal JS, et al. Prospective multicenter study using artificial intelligence to improve dermoscopic melanoma diagnosis in patient care. Commun Med (Lond). 2024;4(1):177. https://doi.org/10.1038/s43856-024-00598-5.
9. Cassidy B, Kendrick C, Brodzicki A, Jaworek-Korjakowska J, Yap MH. Analysis of the isic image datasets: Usage, benchmarks and recommendations. Med Image Anal. 2022;75:102305. https://doi.org/10.1016/j.media.2021.102305.
10. Tschandl P, Rinner C, Apalla Z, Argenziano G, Codella N, Halpern A, et al. Human-computer collaboration for skin cancer recognition. Nat Med. 2020;26(8):1229-34. https://doi.org/10.1038/s41591-020-0942-0.
11. Bansal P, Garg R, Soni P. Detection of melanoma in dermoscopic images by integrating features extracted using handcrafted and deep learning models. Comput Ind Eng. 2022;168(C). https://doi.org/10.1016/j.cie.2022.108060.
12. Kawahara J, Daneshvar S, Argenziano G, Hamarneh G. 7-point checklist and skin lesion classification using multi-task multi-modal neural nets. IEEE J Biomed Health Inform. 2018;23(2): 538–546. https://doi.org/10.1109/jbhi.2018.2824327.
13. Musthafa MM, T RM, V VK, Guluwadi S. Enhanced skin cancer diagnosis using optimized cnn architecture and checkpoints for automated dermatological lesion classification. BMC Med Imaging. 2024;24(1):201. https://doi.org/10.1186/s12880-024-01356-8.
14. Sarıateş M, Özbay E. Transfer learning-based ensemble of cnns and vision transformers for accurate melanoma diagnosis and image retrieval. Diagnostics (Basel). 2025;15(15). https://doi.org/10.3390/diagnostics15151928.
15. Hussain S., Toscano E. An extensive investigation into the use of machine learning tools and deep neural networks for the recognition of skin cancer: Challenges, future directions, and a comprehensive review. Symmetry. 2024;16(3):366. https://doi.org/10.3390/sym16030366.
16. 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. https://doi.org/10.1038/sdata.2018.161.
17. Debelee TG. Skin lesion classification and detection using machine learning techniques: A systematic review. Diagnostics (Basel). 2023;13(19). https://doi.org/10.3390/diagnostics13193147.
18. Gessert N, Nielsen M, Shaikh M, Werner R, Schlaefer A. Skin lesion classification using ensembles of multi-resolution efficientnets with meta data. MethodsX. 2020;7:100864. https://doi.org/10.1016/j.mex.2020.100864.
19. Kurtansky NR, D'Alessandro BM, Gillis MC, Betz-Stablein B, Cerminara SE, Garcia R, et al. The slice-3d dataset: 400,000 skin lesion image crops extracted from 3d tbp for skin cancer detection. Sci Data. 2024;11(1):884. https://doi.org/10.1038/s41597-024-03743-w.
20. Primiero CA, Rezze GG, Caffery LJ, Carrera C, Podlipnik S, Espinosa N, et al. A narrative review: Opportunities and challenges in artificial intelligence skin image analyses using total body photography. J Invest Dermatol. 2024;144(6):1200-7. https://doi.org/10.1016/j.jid.2023.11.007.
21. Smak Gregoor AM, Sangers TE, Eekhof JA, Howe S, Revelman J, Litjens RJ, et al. Artificial intelligence in mobile health for skin cancer diagnostics at home (aim high): A pilot feasibility study. EClinicalMedicine. 2023;60:102019. https://doi.org/10.1016/j.eclinm.2023.102019.
22. Marchetti MA, Cowen EA, Kurtansky NR, Weber J, Dauscher M, DeFazio J, et al. Prospective validation of dermoscopy-based open-source artificial intelligence for melanoma diagnosis (prove-ai study). NPJ Digit Med. 2023;6(1):127. https://doi.org/10.1038/s41746-023-00872-1.
23. Menzies SW, Sinz C, Menzies M, Lo SN, Yolland W, Lingohr J, et al. Comparison of humans versus mobile phone-powered artificial intelligence for the diagnosis and management of pigmented skin cancer in secondary care: A multicentre, prospective, diagnostic, clinical trial. Lancet Digit Health. 2023;5(10):e679-e91. https://doi.org/10.1016/s2589-7500(23)00130-9.
24. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature. 2017;542(7639):115-8.

https://doi.org/10.1038/nature21056.
25. Wu Y, Chen B, Zeng A, Pan D, Wang R, Zhao S. Skin cancer classification with deep learning: A systematic review. Front Oncol. 2022;12:893972. https://doi.org/10.3389/fonc.2022.893972.
26. Han SS, Kim MS, Lim W, Park GH, Park I, Chang SE. Classification of the clinical images for benign and malignant cutaneous tumors using a deep learning algorithm. J Invest Dermatol. 2018;138(7):1529-38. https://doi.org/10.1016/j.jid.2018.01.028.
27. Chanda T, Hauser K, Hobelsberger S, Bucher TC, Garcia CN, Wies C, et al. Dermatologist-like explainable ai enhances trust and confidence in diagnosing melanoma. Nat Commun. 2024;15(1):524. https://doi.org/10.1038/s41467-023-43095-4.
28. Dagnaw G, el Mouhtadi M, Mustapha M. Skin cancer classification using vision transformers and explainable artificial intelligence. J Med Artif Intell. 2024;7:14-. https://doi.org/10.21037/jmai-24-6.
29. Khan S, Ali H, Shah Z. Identifying the role of vision transformer for skin cancer-a scoping review. Front Artif Intell. 2023;6:1202990. https://doi.org/10.3389/frai.2023.1202990.
30. Phung M, Muralidharan V, Rotemberg V, Novoa RA, Chiou AS, Sadée CY, et al. Best practices for clinical skin image acquisition in translational artificial intelligence research. J Invest Dermatol. 2023;143(7):1127-32. https://doi.org/10.1016/j.jid.2023.02.035.
31. Jain A, Way D, Gupta V, Gao Y, de Oliveira Marinho G, Hartford J, et al. Development and assessment of an artificial intelligence-based tool for skin condition diagnosis by primary care physicians and nurse practitioners in teledermatology practices. JAMA Netw Open. 2021;4(4):e217249. https://doi.org/10.1001/jamanetworkopen.2021.7249.
32. Abdul Gafoor SM, Nelson T, Woodcock E, Adityani B. Patient-led teledermatology for skin lesion triage: A service evaluation of the dyplenstm dermoscope. Clin Exp Dermatol. 2024;49(6):612-5. https://doi.org/10.1093/ced/llae031.
33. Han SS, Moon IJ, Kim SH, Na JI, Kim MS, Park GH, et al. Assessment of deep neural networks for the diagnosis of benign and malignant skin neoplasms in comparison with dermatologists: A retrospective validation study. PLoS Med. 2020;17(11):e1003381. https://doi.org/10.1371/journal.pmed.1003381.
34. Haenssle HA, Fink C, Schneiderbauer R, Toberer F, Buhl T, Blum A, et al. Man against machine: Diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. Ann Oncol. 2018;29(8):1836-42. https://doi.org/10.1093/annonc/mdy166.
35. Fujisawa Y, Otomo Y, Ogata Y, Nakamura Y, Fujita R, Ishitsuka Y, et al. Deep-learning-based, computer-aided classifier developed with a small dataset of clinical images surpasses board-certified dermatologists in skin tumour diagnosis. Br J Dermatol. 2019;180(2):373-81. https://doi.org/10.1111/bjd.16924.
36. Winkler JK, Kommoss KS, Vollmer AS, Blum A, Stolz W, Kränke T, et al. Computerizing the first step of the two-step algorithm in dermoscopy: A convolutional neural network for differentiating melanocytic from non-melanocytic skin lesions. Eur J Cancer. 2024;210:114297. https://doi.org/10.1016/j.ejca.2024.114297.
37. Yazdanparast T, Shamsipour M, Ayatollahi A, Delavar S, Ahmadi M, Samadi A, et al. Comparison of the diagnostic accuracy of teledermoscopy, face-to-face examinations and artificial intelligence in the diagnosis of melanoma. Indian J Dermatol. 2024;69(4):296-300. https://doi.org/10.4103/ijd.ijd_61_24.
38. Marsden H, Kemos P, Venzi M, Noy M, Maheswaran S, Francis N, et al. Accuracy of an artificial intelligence as a medical device as part of a uk-based skin cancer teledermatology service. Front Med (Lausanne). 2024;11:1302363. https://doi.org/10.3389/fmed.2024.1302363.
39. Zhao M, Kawahara J, Abhishek K, Shamanian S, Hamarneh G. Skin3d: Detection and longitudinal tracking of pigmented skin lesions in 3d total-body textured meshes. Med Image Anal. 2022;77:102329. https://doi.org/10.1016/j.media.2021.102329.
40. Khan MS, Alam KN, Dhruba AR, Zunair H, Mohammed N. Knowledge distillation approach towards melanoma detection. Comput Biol Med. 2022;146:105581. https://doi.org/10.1016/j.compbiomed.2022.105581.
41. Thomas L, Hyde C, Mullarkey D, Greenhalgh J, Kalsi D, Ko J. Real-world post-deployment performance of a novel machine learning-based digital health technology for skin lesion assessment and suggestions for post-market surveillance. Front Med (Lausanne). 2023;10:1264846. https://doi.org/10.3389/fmed.2023.1264846.
42. Venkatesh KP, Kadakia KT, Gilbert S. Learnings from the first ai-enabled skin cancer device for primary care authorized by fda. NPJ Digit Med. 2024;7(1):156. https://doi.org/10.1038/s41746-024-01161-1.
43. Yuan L, Jin K, Shao A, Feng J, Shi C, Ye J, et al. Analysis of international publication trends in artificial intelligence in skin cancer. Clin Dermatol. 2024;42(6):570-84. https://doi.org/10.1016/j.clindermatol.2024.09.012.
44. International Skin Imaging Collaboration. SIIM-ISIC 2020 Challenge Dataset. International Skin Imaging Collaboration https://doi.org/10.34970/2020-ds01
45. Pacheco AGC, Lima GR, Salomão AS, Krohling B, Biral IP, de Angelo GG, et al. Pad-ufes-20: A skin lesion dataset composed of patient data and clinical images collected from smartphones. Data Brief. 2020;32:106221. https://doi.org/10.1016/j.dib.2020.106221.
46. Jang BK, Park YR. Development and validation of adaptable skin cancer classification system using dynamically expandable representation. Healthc Inform Res. 2024;30(2):140-6. https://doi.org/10.4258/hir.2024.30.2.140.