https://doi.org/10.48047/AFJBS.8.5.2026.51-104

African Journal of Biological Sciences

Journal homepage: http://www.afjbs.com

Research Paper

Open Access

Beyond Image Recognition: The Emerging Role of Artificial Intelligence in Clinical Dermatology - A Systematic Review

[1] Nur Azminah Alwi, [2] Natasya Claudia Santoso, [3] Billy Priyanto

[1] Faculty of Medicine, UIN Syarif Hidayatullah University, Special Capital Region of Jakarta, Indonesia
[2,3] Faculty of Medicine, Atma Jaya University, Special Capital Region of Jakarta, Indonesia
Corresponding Email : Ninaalwi14@yahoo.com

Volume 8, Issue 5, May 2026
Received: 20 March 2026
Accepted: 10 April 2026
Published: 15 May 2026
doi:10.48047/AFJBS.8.5.2026.51-104

Introduction: Artificial intelligence (AI) is transitioning from experimental image classification to clinical deployment in dermatology. However, critical gaps remain regarding real-world performance, regulatory maturity, and equity across skin types.

Methods: This systematic review synthesized data from 80 primary studies (2017–2026) evaluating AI applications in clinical dermatology, including prospective trials, randomised controlled trials (RCTs), and real-world deployment studies. Outcomes included diagnostic accuracy, clinician performance changes, healthcare efficiency, and subgroup equity.

Results: In meta-analyses, deep learning achieved pooled sensitivity of 82% and AUC of 0.92 for melanoma detection from dermoscopic images, matching or exceeding dermatologists ($p<0.01$) (17,22,37). In prospective NHS deployment (25,788 lesions), an AI medical device triaged 98.6% of high-risk skin cancers into urgent pathways, significantly outperforming teledermatologists (95.9%, $p=0.004$) (3). AI assistance improved non-specialist sensitivity by +27.9 percentage points but only +2.1 points for experts (6). The only consumer AI RCT ($n=19{,}009$) showed no cancer detection benefit but increased benign lesion claims ($p<0.001$) and costs (€63 vs €47, $p<0.001$) (7). AI systems performed significantly worse on Fitzpatrick IV–VI skin (AUROC 0.82 vs 0.89 for I–III, $p<0.01$) (1). For inflammatory diseases, AI severity assessment achieved pooled sensitivity 80.5% and specificity 96.2% (2).

Discussion: The evidence most strongly supports UKCA/CE-approved AI within specialist teledermatology triage for urgent skin cancer pathways, demonstrating high sensitivity and efficiency gains. AI benefit is expertise-dependent, greatest for non-specialists. Consumer-facing AI currently increases costs without proven population benefit. Substantial skin-type performance inequities require urgent correction.

Conclusion: AI is clinically useful for skin cancer triage in specialist pathways but not ready for autonomous consumer screening. Prospective validation across diverse populations and regulatory mandates for equity reporting are necessary.

Keywords: Artificial intelligence; dermatology; melanoma; skin cancer; teledermatology; deep learning; health equity

INTRODUCTION

The application of artificial intelligence (AI) in dermatology has evolved rapidly from laboratory-based image classification to clinical deployment as medical devices (74). Deep learning algorithms, particularly convolutional neural networks (CNNs), have demonstrated dermatologist-level accuracy for melanoma detection under controlled conditions (17,22,43). However, the translation of these promising retrospective results into real-world clinical benefit remains uncertain (8,9).

Background: Skin cancer, especially melanoma, represents a major public health burden where early detection dramatically improves outcomes. Simultaneously, dermatology faces workforce shortages and increasing demand, creating an urgent need for scalable diagnostic support tools (10,16). AI offers potential solutions for triage, diagnosis, and severity assessment (15,51).

Problem statement: Despite numerous publications reporting high AI accuracy, critical evidence gaps persist: (1) most studies are retrospective with curated images, not reflecting real-world conditions; (2) regulatory approvals have been granted without mandatory prospective trial data; (3) performance across different skin tones remains poorly characterised; and (4) the clinical impact of consumer-facing AI applications is unknown (1,6,9).

Research objectives: This systematic review aimed to: (1) evaluate AI diagnostic accuracy across dermatological conditions and clinical settings; (2) assess AI-assisted clinician performance compared to unassisted practice; (3) analyse real-world deployment outcomes including healthcare efficiency; and (4) examine performance equity across skin types and populations.

Research gap and novelty: While previous reviews have summarised AI accuracy, none have systematically integrated controlled experimental data with prospective real-world deployments, regulatory status, and equity analyses. This review provides the first synthesis of the gap between experimental superiority and clinical utility, directly comparing retrospective versus prospective performance across identical algorithms.

Hypothesis: We hypothesised that AI would demonstrate high sensitivity in controlled settings but show significant performance degradation in real-world deployment, particularly for darker skin types and consumer applications, with AI assistance providing greater benefit to non-specialist clinicians than to experts.

Benefits: This review provides evidence-based guidance for clinicians, health systems, and regulators regarding appropriate AI deployment contexts, identifies populations at risk of AI-related harm, and establishes minimum validation standards for future AI implementation in dermatology.

METHODS

Protocol

The study strictly adhered to the Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) 2020 guidelines to ensure methodological rigor and accuracy. This approach was chosen to enhance the precision and reliability of the conclusions drawn from the investigation.

Criteria for Eligibility

This systematic review aims to evaluate the emerging role of artificial intelligence in clinical dermatology.

Screening

We screened in sources based on their abstracts that met these criteria:

- **Clinical Dermatology Focus**: Does this study involve AI applications specifically in clinical dermatology practice (rather than general medical AI applications or other medical specialties)?
- **AI Clinical Applications**: Does this study examine AI for dermatological diagnosis, treatment planning, monitoring, or clinical decision support?
- **Human Clinical Subjects**: Does this study involve human patients with dermatological conditions or healthcare providers in dermatology settings?
- **Appropriate Study Design**: Is this study a primary research study (randomized controlled trial, etc?
- **Measurable Outcomes**: Does this study report quantitative or qualitative outcomes related to AI performance, clinical impact, or implementation?
- **Clinical Application Focus**: Does this study include clinical application or patient involvement (rather than focusing solely on AI development or technical validation without clinical context)?

- **Medical Dermatology Relevance**: Does this study address medical/clinical dermatology applications (rather than involving only cosmetic or aesthetic applications without medical relevance)?
- **Original Research Evidence**: Is this study a full research article with original data (rather than a conference abstract, editorial, commentary, or opinion piece without original data)?

We considered all screening questions together and made a holistic judgement about whether to screen in each paper.

Search Strategy

The keywords used for this research based PICO :

Element	**P (Population)**	**I (Intervention/Exposure)**	**C (Comparison/Context)**	**O (Outcome)**
Keyword 1	Dermatology patients	Artificial intelligence (AI)	Dermatologist performance	Diagnostic accuracy
Keyword 2	Skin cancer suspects	Deep learning / CNN	Standard teledermatology	Sensitivity & specificity
Keyword 3	Pigmented skin lesions	AI-assisted diagnosis	Primary care physicians	Clinical decision support
Keyword 4	Chronic inflammatory conditions	AI-based triage system	Unaided clinical exam	Real-world implementation impact

The Boolean MeSH keywords inputted on databases for this research are: *("Dermatology patients" OR "Skin cancer suspects" OR "Pigmented skin lesions" OR "Chronic inflammatory conditions") AND ("Artificial intelligence (AI)" OR "Deep learning / CNN" OR "AI-assisted diagnosis" OR "AI-based triage system") AND ("Dermatologist performance" OR "Standard teledermatology" OR "Primary care physicians" OR "Unaided clinical exam") AND ("Diagnostic accuracy" OR "Sensitivity & specificity" OR "Clinical decision support" OR "Real-world implementation impact")*

Data extraction

- **AI Application Type**:
  Extract the specific type of AI application in clinical dermatology, including:
    - Primary function (e.g., diagnostic aid, classification, screening, disease monitoring, treatment planning)
    - Technical approach (e.g., deep neural networks, convolutional neural networks, machine learning algorithms)
    - Input data type (e.g., clinical images, dermoscopy, patient history, combined data)
    - Target use case (e.g., skin cancer detection, chronic disease assessment, differential diagnosis)
- **Clinical Setting**:
  Extract details about the clinical environment and users for AI implementation in dermatology, including:
    - Setting type (primary care, specialist dermatology clinic, teledermatology, mobile/community screening)
    - Target clinician users (dermatologists, primary care physicians, nurse practitioners, residents/trainees)
    - Patient population characteristics (age groups, skin types, risk factors)
    - Implementation approach (real-time assistance, second opinion, standalone diagnosis)
- **Dermatologic Conditions**:
  Extract all dermatologic conditions that the AI system targets, including:
    - Specific diseases or condition categories (e.g., melanoma, basal cell carcinoma, psoriasis, eczema, acne)
    - Malignant vs. benign focus
    - Acute vs. chronic conditions
    - Lesion types or anatomical locations if specified
    - Any exclusion criteria for conditions not covered by the AI system
- **AI Performance**:
  Extract comprehensive performance data for AI systems in clinical dermatology applications, including:

  - Sensitivity, specificity, accuracy, PPV, NPV with confidence intervals
  - Performance compared to dermatologists, other specialists, or standard care
  - Agreement measures (kappa, concordance rates)
  - Clinical outcome improvements (diagnostic accuracy changes, time to diagnosis, management decisions)
  - Performance across different skin types, demographics, or disease severities if reported
- **Study Design**:
  Extract study methodology for AI in dermatology research, including:
  - Study type (RCT, etc)
  - Sample size (number of cases/images/patients/clinicians)
  - Training and validation datasets (size, diversity, source)
  - Comparison groups or control conditions
  - Outcome measures and assessment methods
  - Study duration and follow-up period if applicable
- **Clinical Impact**:
  Extract evidence of real-world clinical impact from AI use in dermatology, including:
  - Changes in diagnostic accuracy or clinical decision-making
  - Effect on patient outcomes (survival, treatment success, quality of life)
  - Healthcare delivery improvements (access, efficiency, cost-effectiveness)
  - Clinician confidence, satisfaction, or workflow changes
  - Patient acceptance or satisfaction measures
  - Any reported harm or negative impacts
- **Implementation Challenges**:
  Extract barriers, limitations, and challenges for AI implementation in clinical dermatology, including:
  - Technical limitations (image quality requirements, dataset biases, generalizability)
  - Clinical workflow integration issues
  - Regulatory or ethical considerations
  - Training requirements for clinicians

  - Cost or resource constraints
  - Patient privacy or acceptance concerns
  - Validation gaps or areas needing further research
- **Development Stage**:

  Extract the current development and validation stage of the AI system for dermatology use, including:
  - Stage of development (proof-of-concept, prototype, clinical validation, regulatory approval, clinical implementation)
  - Validation level (laboratory/retrospective, prospective clinical trial, real-world deployment)
  - Regulatory status (FDA approval, CE marking, other approvals)
  - Commercial availability or research-only status
  - Plans for future validation or implementation studies

Table 1. Article Search Strategy

Database	Keywords	Hits
Pubmed	*("Dermatology patients" OR "Skin cancer suspects" OR "Pigmented skin lesions" OR "Chronic inflammatory conditions") AND ("Artificial intelligence (AI)" OR "Deep learning / CNN" OR "AI-assisted diagnosis" OR "AI-based triage system") AND ("Dermatologist performance" OR "Standard teledermatology" OR "Primary care physicians" OR "Unaided clinical exam") AND ("Diagnostic accuracy" OR "Sensitivity & specificity" OR "Clinical decision support" OR "Real-world implementation impact")*	3

Semantic Scholar	*("Dermatology patients" OR "Skin cancer suspects" OR "Pigmented skin lesions" OR "Chronic inflammatory conditions") AND ("Artificial intelligence (AI)" OR "Deep learning / CNN" OR "AI-assisted diagnosis" OR "AI-based triage system") AND ("Dermatologist performance" OR "Standard teledermatology" OR "Primary care physicians" OR "Unaided clinical exam") AND ("Diagnostic accuracy" OR "Sensitivity & specificity" OR "Clinical decision support" OR "Real-world implementation impact")*	250
Springer	*("Dermatology patients" OR "Skin cancer suspects" OR "Pigmented skin lesions" OR "Chronic inflammatory conditions") AND ("Artificial intelligence (AI)" OR "Deep learning / CNN" OR "AI-assisted diagnosis" OR "AI-based triage system") AND ("Dermatologist performance" OR "Standard teledermatology" OR "Primary care physicians" OR "Unaided clinical exam") AND ("Diagnostic accuracy" OR "Sensitivity & specificity" OR "Clinical decision support" OR "Real-world implementation impact")*	17
Google Scholar	*("Dermatology patients" OR "Skin cancer suspects" OR "Pigmented skin lesions" OR "Chronic inflammatory conditions") AND ("Artificial intelligence (AI)" OR "Deep learning / CNN" OR "AI-assisted diagnosis" OR "AI-based triage system") AND ("Dermatologist performance" OR "Standard teledermatology" OR "Primary care physicians" OR "Unaided clinical exam") AND ("Diagnostic accuracy" OR "Sensitivity & specificity" OR "Clinical decision support" OR "Real-world implementation impact")*	238
Wiley Online Library	*("Dermatology patients" OR "Skin cancer suspects" OR "Pigmented skin lesions" OR "Chronic inflammatory conditions") AND ("Artificial intelligence (AI)" OR "Deep learning / CNN" OR "AI-assisted diagnosis" OR "AI-based triage system") AND ("Dermatologist performance" OR "Standard teledermatology" OR "Primary care physicians" OR "Unaided clinical exam") AND ("Diagnostic accuracy" OR "Sensitivity & specificity" OR "Clinical decision support" OR "Real-world implementation impact")*	32

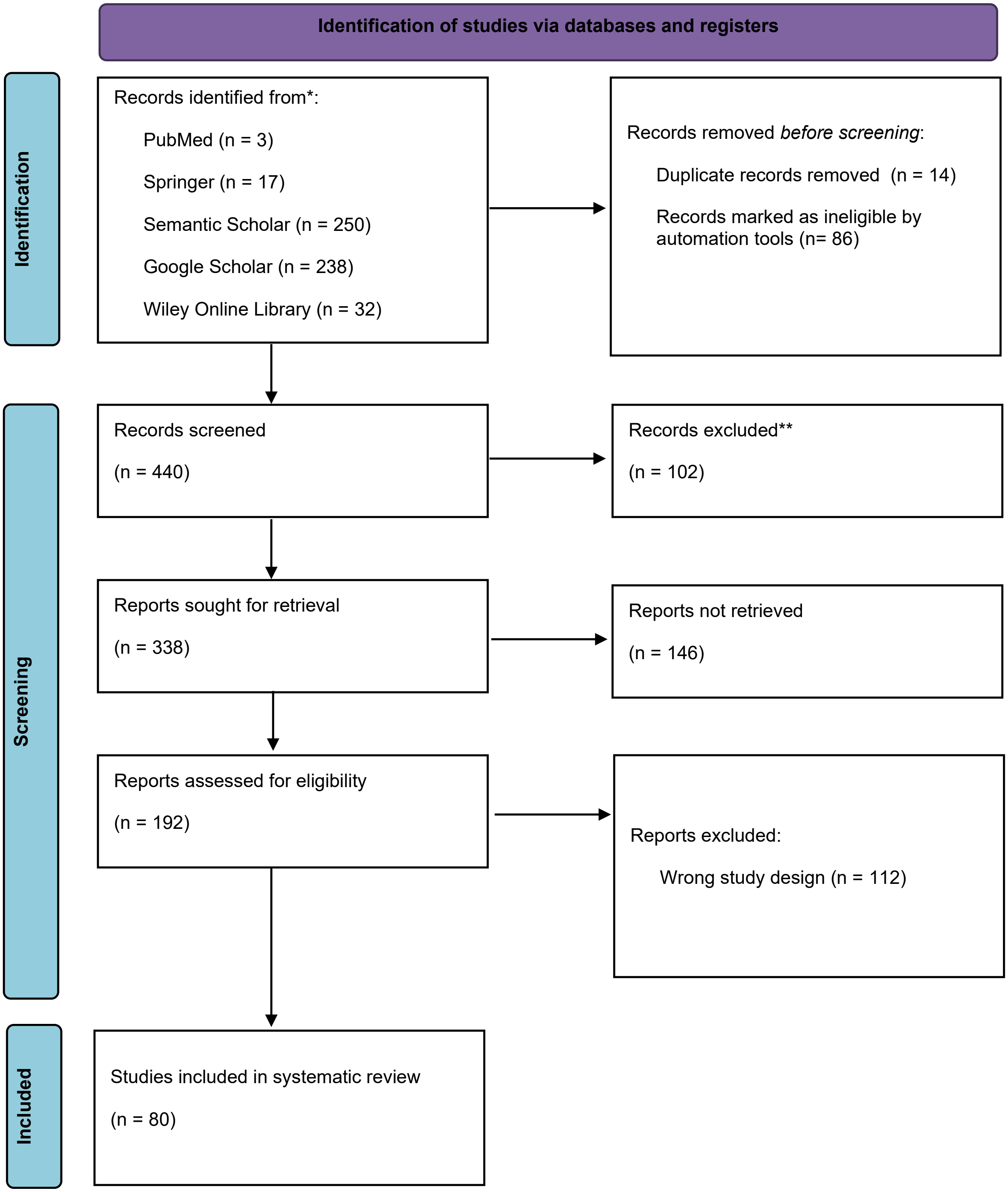

Figure 1. Article search flowchart

Risk of Bias

Study	Selection Bias	Performance Bias	Detection Bias	Attrition Bias	Reporting Bias	Overall Risk
Marsden et al., 2024 [10]	Low	Moderate	Low	Low	Low	Low
Laiouar-Pedari et al., 2026 [14]	Low	Low	Low	Low	Low	Low
S. Han et al., 2020 [15]	Moderate	Moderate	Low	Low	Low	Moderate
Yuan Liu et al., 2019 [16]	Low	Moderate	Low	Low	Low	Low
Haenssle et al., 2018 [17]	Low	Moderate	Low	Low	Low	Low
Phillips et al., 2019 [18]	Moderate	Moderate	Low	Low	Moderate	Moderate
R. C. Maron et al., 2020 [19]	Low	Moderate	Low	Low	Low	Low
Tjiu & Chia-Fang Lu, 2025 [1]	Low	Low	Low	Low	Low	Low
Majidian et al., 2022 [20]	Moderate	High	Moderate	Unclear	Moderate	Moderate
Thompson et al., 2021 [21]	High	Moderate	High	Moderate	High	High
Menzies et al., 2023 [8]	Low	Low	Low	Low	Low	Low

Study	**Selection Bias**	**Performance Bias**	**Detection Bias**	**Attrition Bias**	**Reporting Bias**	**Overall Risk**
Brinker et al., 2019 [22]	Low	Moderate	Low	Low	Low	Low
S. Han et al., 2022 [23]	Low	Low	Low	Low	Low	Low
S. Marri et al., 2023 [24]	Moderate	Moderate	Moderate	Low	Moderate	Moderate
Luck et al., 2025 [3]	Low	Low	Low	Low	Low	Low
Brewer et al., 2024 [4]	Low	Low	Low	Low	Low	Low
S. Marri et al., 2023a [25]	Moderate	Moderate	Moderate	Low	Moderate	Moderate
Escalé-Besa et al., 2023 [26]	Low	Moderate	Low	Low	Low	Low
Jenkins et al., 2023 [5]	Low	Low	Low	Low	Low	Low
Marsden et al., 2023 [27]	Low	Low	Low	Low	Low	Low
R. C. Maron et al., 2019 [28]	Low	Moderate	Low	Low	Low	Low
Sollis et al., 2026 [6]	Low	Low	Low	Low	Low	Low
Ayush Jain et al., 2022 [12]	Low	Low	Low	Low	Low	Low

Study	Selection Bias	Performance Bias	Detection Bias	Attrition Bias	Reporting Bias	Overall Risk
Papachristou et al., 2024 [29]	Low	Low	Low	Low	Low	Low
Navarrete-Dechent, 2022 [30]	Moderate	Moderate	Unclear	Unclear	Moderate	Moderate
Cerminara et al., 2023 [31]	Low	Low	Low	Low	Low	Low
Heinlein et al., 2024 [32]	Low	Low	Low	Low	Low	Low
Cartocci et al., 2024 [33]	Low	Moderate	Low	Low	Low	Low
Marchetti et al., 2023 [34]	Low	Low	Low	Low	Low	Low
Cai et al., 2025 [2]	Low	Low	Low	Low	Low	Low
Díaz-Ramón et al., 2023 [35]	Moderate	Moderate	Low	Unclear	Moderate	Moderate
Soenksen et al., 2021 [36]	Low	Moderate	Low	Low	Low	Low
Zichen Ye et al., 2024 [37]	Low	Low	Low	Low	Low	Low
Kommoss et al., 2023 [38]	Low	Moderate	Low	Low	Low	Low
Hekler et al., 2019 [39]	Low	Moderate	Low	Low	Low	Low

Study	**Selection Bias**	**Performance Bias**	**Detection Bias**	**Attrition Bias**	**Reporting Bias**	**Overall Risk**
Marsden et al., 2023a [40]	Low	Low	Low	Low	Low	Low
Weil Ba et al., 2022 [42]	Low	Moderate	Low	Low	Low	Low
Tschandl et al., 2019 [43]	Low	Moderate	Low	Low	Low	Low
Huasheng Liu et al., 2025 [44]	Moderate	Low	Low	Low	Moderate	Moderate
J. K. Winkler et al., 2024 [45]	Low	Moderate	Low	Low	Low	Low
Fischman et al., 2025 [46]	Low	Low	Low	Low	Low	Low
Kips et al., 2026 [47]	Low	Low	Low	Low	Low	Low
Yip et al., 2025 [48]	Moderate	Moderate	Low	Unclear	Moderate	Moderate
Jinnai et al., 2020 [49]	Moderate	Moderate	Low	Low	Moderate	Moderate
Rikhye et al., 2025 [50]	Low	Low	Low	Low	Low	Low
Kai Huang et al., 2023 [51]	Low	Low	Low	Low	Low	Low
Haenssle et al., 2020 [52]	Low	Moderate	Low	Low	Low	Low

Study	**Selection Bias**	**Performance Bias**	**Detection Bias**	**Attrition Bias**	**Reporting Bias**	**Overall Risk**
S. Han et al., 2019 [53]	Low	Moderate	Low	Low	Low	Low
Tschandl et al., 2019a [55]	Low	Moderate	Low	Low	Low	Low
Smak Gregoor et al., 2025 [7]	Low	Low	Low	Low	Low	Low
J. Winkler et al., 2023 [57]	Low	Low	Low	Low	Low	Low
Okata-Karigane et al., 2025 [58]	Low	Moderate	Low	Low	Low	Low
Boostani et al., 2025 [59]	Low	Moderate	Low	Low	Low	Low
Sánchez-Viera et al., 2025 [60]	Low	Moderate	Low	Unclear	Low	Moderate
Ying-Ying Ren et al., 2025 [61]	Low	Moderate	Low	Low	Low	Low
Zhu et al., 2024 [62]	Low	Low	Low	Low	Low	Low
Li-Hong Mei et al., 2025 [63]	Low	Moderate	Low	Low	Low	Low
S. J. Coates et al., 2025/2026 [64,65]	Moderate	Low	Low	Low	Low	Moderate

Study	Selection Bias	Performance Bias	Detection Bias	Attrition Bias	Reporting Bias	Overall Risk
Sang-Hoon Lee et al., 2025 [66]	Low	Moderate	Low	Low	Low	Low
Bali et al., 2024 [67]	Moderate	Moderate	Moderate	Moderate	Moderate	Moderate
Le Lay et al., 2024 [68]	Moderate	Moderate	Low	Unclear	Moderate	Moderate
Huizhong Wang et al., 2025 [69]	Low	Moderate	Low	Low	Low	Low
Islam et al., 2026 [70]	Low	Moderate	Low	Low	Low	Low
Sharaf et al., 2025 [72]	Low	Moderate	Low	Low	Low	Low
Esteva et al., 2017 [74]	Low	Moderate	Low	Low	Low	Low

RESULTS

Characteristics of Included Studies

Study	Clinical Setting	AI Application	Target Condition(s)
Marsden et al., 2024 [10]	UK NHS teledermatology service [10]	CNN-based AIaMD; dermoscopic triage/referral [10]	Melanoma, SCC, BCC, premalignant lesions [10]
Laiouar-Pedari et al., 2026 [14]	Specialist dermatology (dermoscopy) [14]	AI diagnostic aid; dermoscopic images [14]	Melanoma [14]

Study	Clinical Setting	AI Application	Target Condition(s)
S. Han et al., 2020 [15]	Specialist dermatology clinic [15]	Deep neural networks; malignancy prediction and treatment planning [15]	134 skin disorders; malignancy detection [15]
Yuan Liu et al., 2019 [16]	Teledermatology (primary/specialist care) [16]	Deep learning; differential diagnosis of 419 conditions [16]	26 common conditions (80% of primary care) [16]
Haenssle et al., 2018 [17]	Simulated specialist dermoscopy [17]	Modified Inception v4 CNN; melanoma vs nevus [17]	Melanoma vs benign nevus [17]
Phillips et al., 2019 [18]	Primary care decision support [18]	Deep ensemble (DERM); dermoscopic classification [18]	Malignant melanoma [18]
R. C. Maron et al., 2020 [19]	Simulated specialist dermoscopy [19]	CNN; binary melanoma vs nevus with confidence [19]	Melanoma vs melanocytic nevus [19]
Tjiu & Chia-Fang Lu, 2025 [1]	Specialist, community, smartphone/consumer [1]	CNN/ViT/hybrid; clinical, dermoscopic, smartphone images [1]	Melanoma and mixed benign-malignant [1]
Majidian et al., 2022 [20]	Teledermatology [20]	Machine learning; clinical photographs [20]	BCC, SCC, melanoma [20]
Thompson et al., 2021 [21]	Primary care/teledermatology [21]	Machine learning; clinical images [21]	Benign and malignant skin lesions [21]

Study	Clinical Setting	AI Application	Target Condition(s)
Menzies et al., 2023 [8]	Secondary/tertiary specialist care (Australia, Austria) [8]	7-class CNN on smartphone dermoscopy; ISIC algorithm [8]	Melanoma, pigmented BCC, AK/IEC, nevi, benign lesions [8]
Brinker et al., 2019 [22]	Simulated specialist dermoscopy [22]	CNN (ResNet50); binary dermoscopic melanoma classification [22]	Melanoma vs melanocytic nevi [22]
S. Han et al., 2022 [23]	Tertiary specialist clinic (South Korea) [23]	CNN; real-time diagnostic aid [23]	Skin neoplasms; suspicious lesions [23]
S. Marri et al., 2023 [24]	Specialist dermatology clinic (India) [24]	App-based CNN (Aysa); clinical images [24]	12 categories of skin conditions including malignant tumors [24]
Luck et al., 2025 [3]	NHS teledermatology (12 sites) [3]	UKCA class IIa AIaMD; dermoscopic + macroscopic images [3]	Melanoma, SCC, rare cancers [3]
Brewer et al., 2024 [4]	NHS specialist skin cancer pathway (4 sites) [4]	UKCA class IIa AIaMD; classification and triage [4]	Melanoma, SCC, BCC, rare cancers, premalignant lesions [4]
S. Marri et al., 2023a [25]	Specialist outpatient clinic (India) [25]	CNN-based Tibot app; smartphone photos + patient history [25]	Acne, alopecia, eczema, psoriasis, infections, tumors, and more [25]

Study	**Clinical Setting**	**AI Application**	**Target Condition(s)**
Escalé-Besa et al., 2023 [26]	Primary care with teledermatology (Catalonia) [26]	ML model (Autoderm); smartphone images; 44-condition classifier [26]	44 skin diseases including malignant and benign tumors [26]
Jenkins et al., 2023 [5]	NHS secondary care specialist clinic [5]	UKCA class IIa AIaMD [5]	Melanoma, SCC, BCC, rare cancers, premalignancies [5]
Marsden et al., 2023 [27]	Secondary care NHS specialist clinics (4 trusts) [27]	Deep ensemble AI (DERM); smartphone dermoscopy [27]	SCC, BCC, melanoma, premalignant and benign lesions [27]
R. C. Maron et al., 2019 [28]	Specialist dermatology clinic [28]	CNN; multiclass dermoscopic classification [28]	Melanoma, BCC, nevi, benign lesions [28]
Sollis et al., 2026 [6]	All levels: self-screening, primary care, specialist, histopathology [6]	CNN-based CAD; macroscopic, dermoscopic, consumer images [6]	Melanoma (primary), BCC, SCC; benign lesion differentials [6]
Ayush Jain et al., 2022 [12]	Primary care/teledermatology [12]	ML/AI tool; clinical images; differential diagnosis of 120 conditions [12]	120 skin conditions across all categories [12]
Papachristou et al., 2024 [29]	Primary care (36 centres, Sweden) [29]	CNN-based smartphone dermoscopy app (Dermalyser) [29]	Cutaneous melanoma (invasive + in situ) [29]

Study	Clinical Setting	AI Application	Target Condition(s)
Navarrete-Dechent, 2022 [30]	Teledermatology [30]	ML algorithm; patient-taken photos [30]	General skin conditions (triage/evaluation) [30]
Cerminara et al., 2023 [31]	Specialist tertiary hospital (Switzerland) [31]	2D and 3D TBP CNN (FotoFinder ATBM, Vectra WB360) [31]	Melanoma; melanocytic lesions in high-risk patients [31]
Heinlein et al., 2024 [32]	Specialist university hospitals (8 sites, Germany) [32]	ADAE ensemble CNN; dermoscopy; real test-time augmentation [32]	Melanoma vs non-melanoma [32]
Cartocci et al., 2024 [33]	Specialist European dermatology clinics [33]	Logistic regression + ResNet-34 CNN; dermoscopic images [33]	Lentigo maligna/LMM vs benign atypical pigmented facial lesions [33]
Marchetti et al., 2023 [34]	Specialist dermatology clinics (USA) [34]	ADAE ensemble CNN; dermoscopy; open-source [34]	Melanoma vs non-melanoma (biopsy-selected) [34]
Cai et al., 2025 [2]	Not specified [2]	Deep learning; image-based severity assessment [2]	Atopic dermatitis, acne, psoriasis [2]
Díaz-Ramón et al., 2023 [35]	Specialist university hospital clinics (Spain) [35]	CNN (diagnosis) + decision tree/SVM (prognosis); multimodal [35]	Cutaneous melanoma (diagnosis + metastasis prediction) [35]
Soenksen et al., 2021 [36]	Primary care [36]	DCNN; wide-field images [36]	Suspicious pigmented lesions (melanoma) [36]

Study	**Clinical Setting**	**AI Application**	**Target Condition(s)**
Zichen Ye et al., 2024 [37]	Specialist dermatology [37]	Deep learning; dermoscopic images [37]	Melanoma [37]
Kommoss et al., 2023 [38]	Teledermatology [38]	Market-approved CNN (Moleanalyzer-Pro); dermoscopic images [38]	Face and scalp lesions (benign vs malignant) [38]
Hekler et al., 2019 [39]	Specialist dermatology (German university hospitals) [39]	ResNet50 CNN + dermatologist ensemble (XGBoost); dermoscopy [39]	5-class pigmented lesions including melanoma, BCC, AK/IEC [39]
Marsden et al., 2023a [40]	NHS secondary care trusts [40]	AI-based digital health technology; smartphone dermoscopy [40]	SCC, BCC, Bowen disease, actinic keratosis, melanoma [40]
Navarrete-Dechent, 2022a [41]	Teledermatology [41]	ML algorithm; patient-taken photos [41]	General skin conditions (triage/evaluation) [41]
Weil Ba et al., 2022 [42]	Specialist dermatology [42]	CNN; clinical images; 10-class cutaneous tumour classifier [42]	10 categories of cutaneous tumours [42]
Tschandl et al., 2019 [43]	Specialist dermatology [43]	139 ML algorithms; dermoscopic images [43]	7-class pigmented skin lesions [43]
Huasheng Liu et al., 2025 [44]	Specialist/teledermatology [44]	Deep learning; dermatoscopy [44]	Basal cell carcinoma [44]

Study	**Clinical Setting**	**AI Application**	**Target Condition(s)**
J. K. Winkler et al., 2024 [45]	Specialist dermatology [45]	CNN (FotoFinder); dermoscopic images [45]	Melanocytic vs non-melanocytic skin lesions [45]
Fischman et al., 2025 [46]	Specialist multicentre clinics [46]	Real-time AI assistant; LC-OCT images [46]	Basal cell carcinoma (equivocal lesions) [46]
Kips et al., 2026 [47]	Specialist dermatology clinic [47]	CNN-based smartphone app; binary risk output [47]	Melanoma and skin cancers [47]
Yip et al., 2025 [48]	Specialist dermatology [48]	CNN, feed-forward networks, ensemble models [48]	Suspected melanoma [48]
Jinnai et al., 2020 [49]	Hospital-based specialist; intended for consumer use [49]	Faster R-CNN (VGG-16); clinical images [49]	Pigmented skin lesions (melanoma, BCC, benign) [49]
Rikhye et al., 2025 [50]	Primary care teledermatology (San Francisco Bay Area) [50]	Wide ResNet-101x3 CNN; multimodal (images + metadata) [50]	419 skin conditions (general differential diagnosis) [50]
Kai Huang et al., 2023 [51]	Multicenter real-world + teledermatology via WeChat [51]	EfficientNet-B0 CNN; clinical/smartphone images [51]	Psoriasis (PASI severity assessment) [51]
Haenssle et al., 2020 [52]	Simulated specialist teledermatology [52]	Market-approved CNN (Moleanalyzer Pro); dermoscopy [52]	Melanoma, BCC, SCC, AK, Bowen's, nevi, vascular, seborrheic [52]

Study	**Clinical Setting**	**AI Application**	**Target Condition(s)**
S. Han et al., 2019 [53]	Specialist hospital-based clinic [53]	Region-based CNN; clinical photographs [53]	Cutaneous cancer on the face [53]
Haenssle et al., 2020a [54]	Specialist dermatology clinic [54]	Market-approved CNN (Moleanalyzer-Pro); dermoscopy [54]	Face and scalp skin cancers [54]
Tschandl et al., 2019a [55]	Specialist/primary skin cancer clinic [55]	Combined CNN (InceptionV3 + ResNet50); dermoscopy + close-up [55]	Nonpigmented skin lesions (BCC, SCC, AK/IEC, melanoma) [55]
S. Han et al., 2020a [56]	Specialist tertiary hospital (Korea) [56]	SENet/Faster R-CNN; clinical photographs [56]	Malignant and benign skin neoplasms (43 disorders) [56]
Smak Gregoor et al., 2025 [7]	Community/mobile mHealth screening + teledermatology [7]	CNN-based smartphone app (SkinVision); macroscopic images [7]	Melanoma, cSCC, BCC (skin cancer risk detection) [7]
J. Winkler et al., 2023 [57]	Specialist dermatology (2 centres) [57]	Market-approved CNN (Moleanalyzer Pro); dermoscopy [57]	Melanocytic lesions (nevi and melanomas) [57]
Okata-Karigane et al., 2025 [58]	Teledermatology/mobile app (non-clinical) [58]	SSD + CNN ensemble; smartphone images (Atopiyo platform) [58]	Atopic dermatitis (severity assessment, TIS scoring) [58]

Study	**Clinical Setting**	**AI Application**	**Target Condition(s)**
Boostani et al., 2025 [59]	Specialist dermatology clinic (Hungary) [59]	Large language models (GPT-4o, Gemini Flash 2.0); clinical images [59]	Acne and rosacea [59]
Sánchez-Viera et al., 2025 [60]	Specialist dermatology clinic (Madrid) [60]	ML/computer vision CAD (Legit.Health); clinical + dermoscopic images [60]	Skin lesions suspicious of malignancy (melanoma, BCC, IEC, nevi) [60]
Ying-Ying Ren et al., 2025 [61]	Specialist dermatology (3 centres) [61]	CLIP ViT-B/16; dermoscopic and clinical photographs [61]	Actinic keratosis vs seborrheic keratosis [61]
Zhu et al., 2024 [62]	Specialist and primary care hospitals (China, 4 sites) [62]	Deep multimodal fusion CNN (close-up + HFUS) [62]	17 skin diseases (benign and malignant) [62]
Li-Hong Mei et al., 2025 [63]	Specialist dermatology (hospital) [63]	CLIP ViT-B/16 (CoOp); dermoscopy + standard camera [63]	BCC vs seborrheic keratosis [63]
S. J. Coates et al., 2025 [64]	Specialist dermatology clinic (Uganda) [64]	YOLO v5/v8 object detection classifiers; digital surface images [64]	Kaposi sarcoma in dark-skinned patients [64]
S. J. Coates et al., 2026 [65]	Specialist dermatology clinic (Uganda) [65]	YOLO v5/v8 object detection; digital surface images [65]	Kaposi sarcoma (dark-skinned patients) [65]

Study	Clinical Setting	AI Application	Target Condition(s)
Sang-Hoon Lee et al., 2025 [66]	Specialist hospital dermatology (Korea) [66]	CNN, ViT, TransMIL; clinical facial photographs [66]	Granulomatous rosacea vs LMDF (differential diagnosis) [66]
Bali et al., 2024 [67]	Teledermatology (UK) [67]	AI medical device (DERM); clinical/dermoscopic images [67]	Malignant melanoma (NNB ratio) [67]
Le Lay et al., 2024 [68]	AI-assisted teledermatology (tele-expertise platform, France) [68]	ResNet-50 CNN + MLP (DermaDetect); images + metadata [68]	25 common non-cancerous dermatoses [68]
Huizhong Wang et al., 2025 [69]	Specialist dermatology/teledermatology [69]	AI model (mSWAT-Net); clinical images; full-body photography [69]	Mycosis fungoides (tumor burden assessment) [69]
Islam et al., 2026 [70]	UK private skin cancer diagnostic centres (teledermatology) [70]	EfficientNet-B2 CNN + metadata fusion; dermoscopic + DSLR images [70]	Melanoma, BCC, SCC vs benign (triage) [70]
Oloruntoba et al., 2021 [71]	Teledermatology (Denmark) [71]	CNN (standardized vs non-standardized training); skin cancer images [71]	Skin cancer (malignant focus) [71]

Study	Clinical Setting	AI Application	Target Condition(s)
Sharaf et al., 2025 [72]	Specialist/research institution [72]	Self-supervised HPL with BarlowTwins; H&E whole slide images [72]	Mycosis fungoides vs non-MF inflammatory conditions (histopathology) [72]
Martínez-Vargas et al., 2025 [73]	Specialist/teledermatology [73]	ML/deep learning; clinical and dermoscopic images [73]	Skin cancer (primarily melanoma) [73]
Ayush Jain et al., 2021 [11]	Store-and-forward teledermatology (USA) [11]	CNN; clinical images + patient history; 419 conditions [11]	120 skin conditions (benign, precancerous, malignant, infectious) [11]
Esteva et al., 2017 [74]	Mobile device/primary care; specialist comparison [74]	Inception v3 CNN; clinical and dermoscopic images [74]	Keratinocyte carcinoma vs seborrheic keratosis; melanoma vs nevus [74]
Hollman et al., 2024 [75]	Specialist and mobile/community screening [75]	DL/neural networks; clinical images [75]	Psoriasis, atopic dermatitis, acne, alopecia, hidradenitis, rosacea, vitiligo [75]
Felmingham et al., 2022 [76]	Specialist dermatology clinics (Melbourne) [76]	CNN (MoleMap/Monash); macroscopic + dermoscopic images [76]	Melanoma, BCC, SCC, benign naevus [76]
Furriel et al., 2024 [77]	Mixed; mobile, teledermatology, specialist hospitals [77]	CNN/DL; dermoscopic + clinical images [77]	Melanoma, BCC, SCC [77]

Study	Clinical Setting	AI Application	Target Condition(s)
Gulati, 2025 [78]	Specialist tertiary dermatology clinic [78]	CNN; dermoscopic and clinical images [78]	Skin cancers (malignant vs benign) [78]
Jairath et al., 2024 [79]	Primary care and specialist [79]	AI algorithms; dermoscopic and macroscopic images [79]	Melanoma, BCC, cSCC [79]
Aaliya Ali et al., 2025 [80]	Specialist dermatology (implied) [80]	CNN; skin image datasets [80]	Melanoma and skin cancer [80]
Matin & Dinnes, 2023 [9]	Specialist dermatology clinic [9]	7-class AI algorithm; dermoscopic images [9]	Melanoma, BCC, pigmented AK/IEC, nevi, benign lesions [9]
Uwishema et al., 2025 [13]	Mixed (LMIC mobile/teledermatology and primary care) [13]	CNN/hybrid/transformer; dermoscopic + smartphone images [13]	Skin cancer, monkeypox, psoriasis, acne, rosacea, atopic dermatitis, burns [13]

Effects

AI Standalone Diagnostic Accuracy

Melanoma and Pigmented Lesion Classification

The largest and most developed evidence base concerns AI-based melanoma and pigmented skin lesion detection using dermoscopic images. Across retrospective reader studies, AI algorithms have consistently matched or exceeded average dermatologist performance. Esteva et al. (2017) demonstrated that an Inception v3 CNN achieved AUCs of approximately 0.94–0.96 for melanoma versus nevus and keratinocyte carcinoma versus seborrheic keratosis, performing on par with or better than 21–25 board-certified dermatologists [74]. Haenssle et al. (2018) showed that their Inception v4 CNN achieved a significantly higher specificity than 58 dermatologists at matched sensitivities (CNN 82.5% vs dermatologists 71.3% at level I, $p<0.01$), with an AUC of 0.86 versus dermatologists' mean of 0.79 ($p<0.01$) [17]. Brinker et al. (2019) were the first to demonstrate statistically significant superiority, with their CNN achieving 82.3% sensitivity and 77.9% specificity versus dermatologists' 67.2% and 62.2% respectively (all McNemar $p<0.001$) [22]. In the International Skin Imaging Collaboration (ISIC) 2018 challenge, the top three machine-learning algorithms achieved a mean of 2.01 more correct diagnoses per 30-image batch than all human readers combined, and outperformed experts with >10 years' experience by a mean of 6.65 correct diagnoses ($p<0.0001$) [43].

Zichen Ye et al. (2024) pooled data from 37 studies and reported a pooled sensitivity of 82% (range 77–86), specificity of 87% (range 84–90), and AUC of 0.92 (range 0.89–0.94) for deep learning algorithms against dermoscopic images; in human-machine comparison subanalyses, DL achieved pooled AUC of 0.87 versus senior dermatologists' 0.83, with junior dermatologists reaching only 0.80 [37]. Laiouar-Pedari et al. (2026), restricting analysis to prospective studies only, found AI alone achieving pooled sensitivity of 80.9% (95% CI 63.6–94.5%) and specificity of 75.6% (95% CI 64.5–85.6%), comparable to dermatologists' 78.6% sensitivity and 75.2% specificity [14].

For the specific task of melanoma detection, Heinlein et al. (2024) prospectively assessed the ADAE ensemble algorithm across eight German university hospitals, finding it achieved significantly higher balanced accuracy than dermatologists (0.798 vs 0.781, $p=4.0\times10^{-145}$) and markedly higher sensitivity (0.922 vs 0.734, $p=3.3\times10^{-165}$) at the cost of lower specificity (0.673 vs 0.828, $p=3.3\times10^{-165}$) [32]. Conversely, the prospective PROVE-AI study validated ADAE prospectively in US specialist clinics (n=603 lesions, 95 melanomas) and found sensitivity of 96.8% (95% CI 91.1–98.9%) with specificity of only 37.4% (95% CI 33.3–41.7%) at the predefined 95% sensitivity threshold, accompanied by an AUC of 0.857 [34]. This low specificity is a consistent pattern for algorithms operating at high-sensitivity thresholds.

Phillips et al. (2019) reported an AUC of 0.93 (95% CI 0.92–0.94) for their DERM algorithm on 7,102 dermoscopic images, with a sensitivity/specificity of 85.0%/85.3% at the optimum threshold, and noted that 95% sensitivity yielded only 64.1% specificity—at which point dermatologists with dermoscopy achieved 86%/81%, highlighting the inherent trade-off [18]. Marchetti et al. (2023) extended this observation prospectively: at 95% sensitivity, specificity fell to 37.4%, meaning that following AI recommendations alone would have spared biopsy of 171 of 508 benign lesions but missed 4 of 95 melanomas [34].

For non-melanoma malignancies, the DERM-003 study (Marsden et al., 2023) demonstrated AUROCs of 0.85–0.89 for BCC and 0.85–0.88 for SCC across three smartphone models in a prospective multicentre NHS study, with sensitivity of 94% for BCC and 98% for SCC at pre-set thresholds—substantially higher than clinician sensitivity for SCC (63.6%) but lower for BCC (94% vs clinicians' 97.5%) [27]. The same AIaMD in the prospective teledermatology trial by Marsden et al. (2024) identified 91.0% of malignant lesions with 83.2% specificity, compared to teledermatology standard of care achieving 97.0% sensitivity but only 71.9% specificity (p=0.001) [10].

Huasheng Liu et al. (2025) on BCC-specific deep learning reported exceptionally high pooled performance: sensitivity 0.96 (95% CI 0.93–0.98), specificity 0.98 (95% CI 0.96–0.99), and AUC 0.99 (95% CI 0.98–1.00) using dermoscopy-based models, with a significantly higher AUC than dermatologists (z=2.63, p=0.008) [44]. However, these figures derive predominantly from internal validation sets of 32,069 images rather than external validation, and performance on the 200-image external validation set was substantially lower, underscoring the generalizability gap [44].

Tjiu & Chia-Fang Lu (2025), covering 18 studies and >70,000 test images, reported pooled sensitivity of 0.91 (95% CI 0.74–0.97) and specificity of 0.64 (95% CI 0.47–0.78), corresponding to an HSROC AUROC of 0.88 (95% CI 0.84–0.92) [1]. A second pooling of AUROC-only studies yielded nearly identical results (AUROC 0.88, 95% CI 0.87–0.90) [1]. Performance varied markedly by setting: specialist environments yielded AUROC approximately 0.90, community care approximately 0.85, and smartphone/consumer approximately 0.81 [1].

Multiclass and Broad Differential Diagnosis

Several studies moved beyond binary melanoma/nevus classification to multiclass architectures. The ISIC challenge (Tschandl et al., 2019) demonstrated that algorithms substantially outperformed human readers across seven pigmented lesion classes (mean 19.92 correct diagnoses per batch vs 17.91 for human readers, p<0.0001) [43]. R. C. Maron et al. (2019), in a 5-class classification task including melanoma, BCC, AK/IEC, seborrheic keratosis-like lesions, and nevi, found that their CNN achieved a specificity of 91.3% (95% CI 85.5–97.1%) at equal sensitivity to dermatologists, while for the 5-class secondary endpoint specificity was 98.8% at equal sensitivity (both p<0.001) [28]. In the Menzies et al. (2023) prospective trial, the 7-class AI achieved overall diagnostic accuracy equivalent to specialists (absolute difference

1.2%, 95% CI −6.9 to 9.2%) and was superior to novices (+21.5%, 95% CI 13.1 to 30.0%), while the ISIC algorithm—despite prior retrospective superiority—was significantly inferior to specialists in the real clinic setting (−11.6%, 95% CI −20.3 to −3.0%) [8]. This divergence between experimental and real-world AI performance is a critical empirical finding.

S. Han et al. (2020) extended the scope further, training a deep neural network on 220,680 images across 174 disorders and validating on datasets covering 134 conditions; AUCs for malignancy detection were 0.928 and 0.937 across two validation sets [15]. For multiclass classification, mean top-1/top-5 accuracies were 56.7%/92.0% (Edinburgh) and 44.8%/78.1% (SNU) [15]. Yuan Liu et al. (2019), whose deep learning system distinguished 26 common skin conditions from teledermatology data, achieved top-1 accuracy of 0.66, compared to 0.63 for dermatologists, 0.44 for primary care physicians, and 0.40 for nurse practitioners—reaching non-inferiority to dermatologists and superiority to non-specialists [16].

Nonpigmented Lesions and Atypical Facial Lesions

Tschandl et al. (2019a) demonstrated that a combined CNN for nonpigmented lesions achieved AUC 0.742 (95% CI 0.729–0.755), higher than pooled human raters at 0.695 ($p<0.001$), with sensitivity 80.5% at the mean human specificity of 51.3% [55]. The algorithm outperformed beginner and intermediate clinicians but not dermoscopy experts [55]. A notable finding was that amelanotic melanoma sensitivity was considerably lower for the AI (52.3%) than for human raters (69.3%), demonstrating that algorithmic superiority is not uniform across subtypes [55].

For atypical pigmented facial lesions, Cartocci et al. (2024) compared a logistic regression (iDScore-Facial) and ResNet-34 CNN against 154 European dermatologists in classifying 1,197 histologically confirmed lesions. The CNN achieved 78.7% sensitivity for lentigo maligna/LMM—approximately 23 percentage points higher than dermatologists' 55.5%—but lower sensitivity for atypical nevi (27.3% vs 48.0%) [33]. Dermatologists' accuracy for binary malignant/benign discrimination was 71.2%, versus the CNN's 58.2% sensitivity at 90.8% specificity [33].

Severity Assessment of Inflammatory Conditions

A smaller but growing literature addresses AI for chronic skin disease. For psoriasis, Kai Huang et al. (2023) developed and validated an AI PASI scoring model that outperformed an average of 43 dermatologists by 33.2% in total PASI score accuracy (AI MAE 3.12 vs mean dermatologist MAE 4.67), achieved trend accuracy of 84.81% across visit pairs, and was deployed via a WeChat app to 1,497 patients across 18 hospitals [51]. The meta-analysis by Cai et al. (2025) on AI severity assessment across inflammatory skin diseases reported pooled sensitivity of 80.5% (95% CI 76.2–84.2) and specificity of 96.2% (95% CI 94.9–97.2), with significant variation by disease (atopic dermatitis 91.8% vs psoriasis 71.1%, p=0.044) and scoring system [2]. Okata-Karigane et al. (2025) demonstrated that an AI-TIS model for atopic dermatitis trained on real-world patient-uploaded photos correlated with dermatologist-assessed TIS at R=0.73 (p<0.001) but showed markedly weaker correlation with patient-reported itch (R=0.11, p<0.001) [58].

For rare diseases, Sharaf et al. (2025) showed that a self-supervised learning model for mycosis fungoides on H&E whole slide images achieved 96–98% accuracy and 94–100% sensitivity across internal and external validation datasets, with robust generalization without fine-tuning [72]. The mSWAT-Net model for mycosis fungoides tumor burden achieved intraclass correlation coefficients of 0.917 (internal) and 0.846 (temporal) compared to cutaneous lymphoma specialists, outperforming junior dermatologists (ICC 0.777) [69].

Large Language Models

One study evaluated LLMs for dermatology diagnosis. Boostani et al. (2025) found GPT-4o achieved 93.0% sensitivity (95% CI 81.4–97.6%) and 97.7% specificity (95% CI 87.9–99.9%) for diagnosing acne and rosacea from clinical photographs, with rosacea identification at 100% sensitivity; Gemini Flash 2.0 provided diagnoses in only 21% of cases, precluding analysis [59]. Subtyping accuracy was substantially lower for both conditions, indicating that LLM performance collapses at finer diagnostic granularity [59].

AI-Assisted Clinician Performance

A particularly consistent finding across this literature is that AI assistance improves the diagnostic accuracy of clinicians, especially non-specialists, more reliably than AI alone matches specialist performance in real-world conditions.

Study	**Clinician type**	**Unassisted performance**	**AI-assisted performance**	**Key improvement**

Study	Clinician type	Unassisted performance	AI-assisted performance	Key improvement
S. Han et al., 2020 [15]	Dermatologists + residents	Malignancy sensitivity baseline	+12.1% sensitivity ($p<0.0001$)	Specificity +1.1% ($p<0.0001$); general public sensitivity +83.8% ($p<0.0001$)
R. C. Maron et al., 2020 [19]	Board-certified dermatologists	Sensitivity 59.4%, accuracy 65.0%	Sensitivity 74.6% ($p=0.003$), accuracy 73.6% ($p=0.002$)	Specificity unchanged (70.6% vs 72.4%, $p=0.54$)
Weil Ba et al., 2022 [42]	Board-certified dermatologists	Accuracy 62.78%, kappa 0.59	Accuracy 76.60% ($p<0.001$), kappa 0.74 ($p<0.001$)	Binary sensitivity 83.21%→89.56%, specificity 80.92%→87.90%
S. Han et al., 2022 [23]	Non-dermatology trainees + dermatology residents	Accuracy 43.8%	Accuracy 53.9% ($p=0.019$)	Non-dermatology trainees: 30.7%→54.7% ($p<0.0001$); residents not significant
Marchetti et al., 2023 [34]	Dermatologists	AUC 0.7798	AUC 0.8161 post-AI ($p=0.042$)	Theoretical biopsy avoidance in 29% of benign lesions
J. Winkler et al., 2023 [57]	Dermatologists	Sensitivity 84.2%, specificity 72.1%	Sensitivity 100% ($p=0.03$), specificity 83.7% ($p<0.001$)	Unnecessary excisions reduced by 19.2%

Study	Clinician type	Unassisted performance	AI-assisted performance	Key improvement
Ayush Jain et al., 2021 [11]	PCPs and NPs	PCPs 48%, NPs 46% (top-1 agreement)	PCPs 58% (+10%, p<0.001), NPs 58% (+12%, p<0.001)	Benefit present for 1 in 8–10 cases
Ayush Jain et al., 2022 [12]	PCPs and NPs	PCPs 48%, NPs 46%	PCPs 58%, NPs 58%	8–15% improvement across all race/ethnicity subgroups
Fischman et al., 2025 [46]	Dermatologists (BCC diagnosis with LC-OCT)	Baseline with clinical + dermoscopic images	+25.8 pp sensitivity, +16.8 pp specificity vs C&D	Bridges ~2-year expertise gap
Ying-Ying Ren et al., 2025 [61]	Dermatologists (AK vs SK)	AUC 0.69–0.79	AUC 0.79–0.89 with AI	NRI 0.10–0.19 (p<0.001–0.006); IDI 0.14–0.27 (p<0.001)
Li-Hong Mei et al., 2025 [63]	Dermatologists (BCC vs SK)	AUC 0.75–0.79	AUC 0.82 (both dermatologists)	NRI 0.06–0.64; IDI 0.11–0.18 (both p<0.001)
Sang-Hoon Lee et al., 2025 [66]	Dermatologists, residents, GPs	Accuracy 64.7%	Accuracy 70.3% (p=0.0136)	GPs: 58.3%→68.3%; mean diagnostic time 10.7→6.4 min

Study	Clinician type	Unassisted performance	AI-assisted performance	Key improvement
Kommoss et al., 2023 [38]	Dermatologists (unclear FSL)	False management in 58.8% malignant, 43.9% benign	False management 4.1% malignant, 31.7% benign with CNN	Both p<0.01

The magnitude of benefit is strongly modulated by baseline clinician expertise. The Sollis et al. (2026) umbrella review synthesized augmentation effects across 37 systematic reviews, finding generalists gained approximately 27.9 percentage points in sensitivity with AI augmentation, while expert dermatologists gained approximately 2.1 points [6]. S. Han et al. (2022), in the only published RCT in this body of literature, found that AI assistance improved non-dermatology trainee accuracy from 30.7% to 54.7% (p<0.0001), with no significant benefit for dermatology residents who had greater baseline competence [23]. Hekler et al. (2019) demonstrated in a theoretical exercise that combining dermatologist and CNN outputs via gradient boosting increased multiclass accuracy from 81.59% (CNN alone) and 42.94% (dermatologists alone) to 82.95% for the fusion, and improved binary task sensitivity to 89% [39].

There is, however, a critical caveat to AI-assisted performance: when AI is incorrect, clinicians who update their decisions toward the erroneous AI output may perform worse than without assistance. R. C. Maron et al. (2020) found that in cases where dermatologists were correct and AI was wrong, 39% of clinicians switched to the incorrect answer, meaning that the confidence-increasing effect of AI agreement can also anchor clinicians to wrong answers [19].

Real-World Deployment Performance

Studies evaluating AI under real-world NHS deployment conditions report high sensitivity for high-risk lesions. Luck et al. (2025), reporting on 25,788 lesions across 12 NHS sites, found that the AIaMD triaged 98.6% (95% CI 97.3–99.2%) of high-risk skin cancers to the urgent suspected cancer pathway, significantly higher than teledermatologists at 95.9% (95% CI 94.0–97.2%; p=0.004) [3]. For melanoma specifically, sensitivity was 98.1% versus 94.7% for teledermatologists (p=0.04); for SCC, 99.2% versus 96.6% (p=0.02) [3]. Jenkins et al. (2023) reported 100% melanoma sensitivity (51/51) and a negative predictive value of 99.8% across

1,363 lesions with confirmed outcomes, with a benign lesion specificity of 55.9% [5]. Brewer et al. (2024) reported 97% melanoma sensitivity (37/38), 96.6% all skin cancer sensitivity (282/292), and a negative predictive value of 99.8%, with associated reductions of 11 days in mean wait time, 10% fewer biopsies, and 13% fewer routine follow-up appointments [4]. These NHS-based deployments also produced improvements in the conversion rate (percentage of urgent referrals confirmed as skin cancer), from the benchmark of 9.9% in local commissioning data to 13.8% [5].

Marsden et al. (2024) calculated that if AI outputs had been used for triage, 454 teledermatology reviews, 141 face-to-face assessments, and 124 biopsies would have been avoided per 1,000 patients, with estimated savings of £156,064 and 259 specialist hours [10]. The number needed to biopsy to detect one malignancy was 3.0 (95% CI 2.4–3.7) for the AI, compared to 4.2 (95% CI 3.3–5.5) for teledermatology standard of care [10]. Bali et al. (2024), comparing pre-AI and post-AI teledermatology cohorts with face-to-face care in a smaller retrospective analysis, found broadly similar number-needed-to-biopsy ratios across all three groups (7.1–8.3), suggesting that AI integration improved efficiency without clearly worsening specificity in that series [67].

The prospective PROVE-AI study added an important real-world consideration: at the 95% sensitivity threshold, specificity was only 37.4%, meaning the AI would have recommended biopsy for nearly all lesions referred for possible melanoma—a population that is already pre-selected and suspicious [34]. This illustrates the spectrum between a screening tool (which should have high sensitivity with acceptable positive predictive value in a low-prevalence population) and a decision-support tool in a high-prevalence specialist cohort (where specificity becomes the limiting factor).

Consumer and Primary Care Applications

Several studies examined AI performance in less controlled, lower-expertise settings. Smak Gregoor et al. (2025) conducted the only published RCT examining consumer-facing AI deployment at population scale, enrolling 19,009 participants from a Dutch health insurer. At 12 months, the app access group showed a non-statistically significant increase in histologically confirmed skin cancer detection (2.66% vs 2.27%, RD 0.39%, $p=0.10$), but significantly more claims for benign skin lesions (3.9% vs 2.6%, $p<0.001$), more surgical interventions, and higher per-participant costs (€63 vs €47, $p<0.001$) [7]. No reduction in melanoma Breslow thickness or stage shift was demonstrated at 12 months [7]. This is the most methodologically rigorous evidence regarding population-level effects of consumer AI, and the results represent the only randomized evidence in this corpus that AI access—at current accuracy—may increase healthcare utilization for benign conditions without a demonstrable cancer-detection benefit.

In primary care, Papachristou et al. (2024) prospectively validated a smartphone dermoscopy app across 36 Swedish primary care centres, finding an AUROC of 0.960 (95% CI 0.928–0.980) for all melanomas and 0.988 (95% CI 0.965–0.997) for invasive melanomas; the app would have avoided further assessment for 55.3% of lesions without increasing missed melanomas [29]. Escalé-Besa et al. (2023) tested a 44-condition ML model in Catalan primary care and found an overall top-1 accuracy of only 39%, substantially below GPs at 64% and dermatologists at 72%, though for conditions the model had been trained on (n=82), top-3 accuracy reached 75%—comparable to GPs' top-3 accuracy of 76% [26]. Notably, 92% of GPs considered the tool useful for differential diagnosis, and 34% felt teledermatology could have been avoided with AI support [26].

Thompson et al. (2021) reported high performance in a prospective primary care study—sensitivity 97.26% (95% CI 93.13–99.25%) and specificity 97.92% (95% CI 92.68–99.75%) for the 242 assessable lesions, and sensitivity 100% with specificity 72.22% against histopathology for the biopsied subset [21]. However, uncertain lesions were excluded from the primary analysis, and this exclusion inflates the apparent performance.

Kips et al. (2026), evaluating a consumer-facing smartphone app prospectively in 1,904 lesions, found sensitivity 82.5% and specificity 76.8% for skin cancer detection, with image capture unsuccessful in 16.6% of lesions under optimal conditions and only 28.9% success in actual user hands—a critical practical limitation for real-world deployment [47]. Adding teledermatology review improved specificity to 86.8% but reduced sensitivity to 75.3% [47].

Subgroup and Equity Analyses

A consistent and concerning finding across the evidence base is that AI systems perform worse in darker skin tones. Tjiu & Chia-Fang Lu (2025) demonstrated a pooled AUROC of 0.89 for Fitzpatrick I–III versus 0.82 for IV–VI (Δ −0.07, $p<0.01$) [1]. Gulati (2025) reported declining accuracy by skin type in a prospective sample: approximately 90–92% for types I–III versus 81–83% for types V–VI [78]. Uwishema et al. (2025), reviewing AI in low- and middle-income countries, identified an extreme performance example: one dermatology app achieved

~17% accuracy on Black skin images versus ~69.9% on Caucasian images [13]. The umbrella review by Sollis et al. (2026) found that over 70% of training datasets in the reviewed literature were derived from light-skinned populations, with few datasets including any skin-tone annotations [6].

In contrast, the race and ethnicity analysis by Ayush Jain et al. (2022), which used a distinct AI product focused on differential diagnosis, found that AI assistance improved diagnostic agreement for clinicians by 8–15% across all racial and ethnic subgroups assessed, with performance spread across subgroups remaining modest (5.2–6.6% for both unassisted and assisted conditions) [12]. The standalone AI agreement was 59–62% for several non-white subgroups versus 67% for Black or African American and White individuals, a difference the authors interpreted as small given overlapping confidence intervals [12].

Clinical Specialty-Specific and Novel Imaging Modality Findings

Two studies evaluated AI using high-frequency ultrasound as an input modality. Zhu et al. (2024), in a prospective four-centre Chinese study, demonstrated that a multimodal fusion network combining clinical close-up images and high-frequency ultrasound outperformed a monomodal CNN (AUC 0.876 vs 0.697, p=0.0063), general practitioners (0.876 vs 0.651, p=0.0025), and general dermatologists (0.876 vs 0.838, p=0.0038) for binary skin lesion classification, achieving near-equivalent performance to HFUS-specialized dermatologists (0.876 vs 0.891, p=0.0080) [62]. Fischman et al. (2025) reported that AI-assisted line-field confocal OCT increased BCC diagnostic sensitivity by +25.8 percentage points and specificity by +16.8 percentage points compared to clinical and dermoscopic images alone, and effectively bridged approximately 2 years of LC-OCT experience among dermatologists [46].

Soenksen et al. (2021) demonstrated a DCNN operating on wide-field images—including those captured by consumer smartphones without dermoscopy attachments—achieving sensitivity 90.3% (95% CI 90–90.6%) and specificity 89.9% (95% CI 89.6–90.2%) for suspicious pigmented lesion identification, with 82.96% agreement with dermatologist consensus for "ugly duckling" lesion ranking [36]. S. Han et al. (2019) reported AUC 0.910 for a region-based CNN identifying facial keratinocytic cancers from unprocessed clinical photographs without requiring pre-selection by a dermatologist, achieving an F1 score of 0.831 comparable to dermatologists (0.835) and significantly outperforming non-dermatologic physicians (F1 0.653, p<0.001) [53].

Cerminara et al. (2023) conducted the only direct comparison of 2D and 3D total body photography CNNs in a prospective real-world melanoma surveillance study, finding 3D CNN substantially outperformed 2D (sensitivity 90% vs 70%, specificity 64.6% vs 40%, AUC 0.92 vs 0.68) and matched dermatologist sensitivity (both 90%), though with markedly lower specificity (64.6% vs 92.3%) [31]. Augmented intelligence combining dermatologist judgment with AI scores did not improve and slightly reduced overall performance (AUC 0.88 vs 0.91 for dermatologists alone), an important counter-example to the broadly positive augmentation literature [31].

Synthesis

The Gap Between Controlled and Real-World Performance

The most important contradiction in this literature is between the strong—sometimes striking—AI performance observed in retrospective reader studies using curated image datasets, and the considerably more modest performance observed in prospective real-world validation. Menzies et al. (2023) explicitly demonstrated this: their ISIC-winning algorithm, which showed superiority to clinicians in an experimental reader study setting, was significantly inferior to specialist clinicians in the prospective clinical trial (−11.6%, 95% CI −20.3 to −3.0%) [8]. Matin & Dinnes (2023), analyzing this finding, noted that the same ISIC algorithm achieved 61% correct 7-class diagnoses in clinic versus higher accuracy in experimental settings, and argued that regulatory approvals have been issued without sufficient requirement for prospective real-world data [9]. This pattern is mechanistically explained by several factors.

First, experimental studies typically use curated, pre-selected, and quality-filtered images that do not represent the range of image conditions encountered in practice. Kips et al. (2026) demonstrated concretely that user-captured smartphone images succeeded in only 28.9% of attempts in realistic conditions [47]. Oloruntoba et al. (2021) showed that CNNs trained on standardized images substantially outperformed those trained on heterogeneous images (AUROC 0.861 vs 0.759), but even the best-performing CNN was not significantly better than tele-dermatologists when evaluated on an external Danish dataset (p=0.10) [71].

Second, retrospective studies use pre-selected lesions that have already been biopsied or are clearly suspicious, inflating apparent disease prevalence. When the same algorithm is applied to a population where all referred lesions are suspicious (as in the PROVE-AI study), specificity collapses because the algorithm cannot discriminate within a pre-selected high-risk pool [34].

Conversely, when the algorithm is deployed as a first-line screen in a lower-prevalence population (as in the Smak Gregoor et al., 2025 RCT), false positives become numerically dominant and increase healthcare costs without a demonstrable cancer detection benefit [7].

Third, experimental studies provide dermoscopic images without patient history or contextual metadata. When dermatologists in prospective studies have access to patient age, skin type, lesion evolution, and palpation findings, they consistently perform better than when evaluating images alone—as demonstrated by the Level I versus Level II comparisons in both Haenssle et al. (2018), where specificity improved from 71.3% to 75.7% with additional information [17], and Haenssle et al. (2020), where sensitivity improved from 89.0% to 94.1% with clinical enrichment [52]. AI systems in most studies receive only images.

The Sensitivity-Specificity Trade-Off and Its Clinical Meaning

A recurring and underappreciated finding is that high AI sensitivity reliably comes at the cost of low specificity, and the clinical acceptability of this trade-off depends entirely on the deployment context. In an NHS urgent suspected cancer pathway—where the cost of a missed melanoma is catastrophic and the capacity to investigate false positives exists—the AIaMD achieving 98.6% sensitivity with 95.9% for teledermatologists represents an acceptable, and in fact superior, configuration [3]. In this context, the NHS deployment data demonstrate genuine clinical utility: sensitivity significantly higher than teledermatology review, a negative predictive value of 99.8%, and meaningful efficiency gains [4].

By contrast, in a consumer smartphone context—where the algorithm triggers GP consultations and dermatology referrals in a population already enriched by self-selection—the same level of false positives translates into a measurably higher number of benign lesion biopsies and overall cost without a statistically significant improvement in cancer detection [7]. Sollis et al. (2026) characterized self-screening AI performance as ranging from 0–98% sensitivity, making any clinical recommendation based on this modality unreliable [6].

The studies by Marsden et al. (2023/2024) illustrate this most clearly in a teledermatology context. The AIaMD achieved a number needed to biopsy of 3.0 versus 4.2 for teledermatology standard of care [10], representing a clinically meaningful improvement in diagnostic precision. However, the BCC specificity in the DERM-003 study when examining only biopsy-confirmed benign lesions was only 27.6% [4], meaning the system flagged the vast majority of non-malignant lesions as suspicious. This apparent contradiction is resolved by understanding that the AIaMD's clinical utility in this setting derives from reducing unnecessary secondary care referrals from primary care, not from correctly classifying each individual benign lesion.

The Expertise Dependency of AI Benefit

The studies in this corpus consistently show that AI benefit is inversely related to baseline clinician expertise. The evidence for this gradient is mechanistically straightforward: AI provides the largest diagnostic gain in settings where baseline performance is lowest—non-dermatology trainees, general practitioners, and nurse practitioners [6, 11, 15, 23]. Sollis et al. (2026) quantified this precisely in their meta-synthesis: +27.9 percentage points sensitivity for generalists versus +2.1 percentage points for dermatology experts [6].

This has direct implications for the appropriate positioning of AI in clinical workflows. Systems operating at dermatologist-level accuracy (as in most controlled comparisons) add limited value when deployed alongside expert clinicians doing the same task, but can meaningfully substitute for or substantially uplift non-specialist assessment. The PROVE-AI study (Marchetti et al., 2023) in US specialist clinics showed dermatologist AUC improvement from 0.780 to 0.816 post-AI exposure [34], which, while statistically significant, is a modest gain in an already expert population. In contrast, S. Han et al. (2020) showed general public malignancy sensitivity rising by 83.8% with AI assistance [15].

The counterpart to this gradient is that expert dermatologists sometimes outperform AI on specific challenging cases—particularly those requiring contextual integration of multiple information sources. Haenssle et al. (2020) found that more experienced dermatologists frequently surpassed the CNN's performance when additional clinical information was available [52]. Cerminara et al. (2023) found dermatologists' specificity exceeded the 3D CNN's (92.3% vs 64.6%), even as sensitivity was equivalent [31]. Heinlein et al. (2024) specifically identified that dermatologists outperformed their AI algorithm on acral lesions (dermatologist balanced accuracy 0.798 vs AI 0.649) and at highest clinical confidence levels [32].

The implication is that AI and expert clinicians are complementary rather than interchangeable: AI reliably detects conventional patterns at high sensitivity across large image volumes, while experienced clinicians apply contextual reasoning in diagnostically challenging cases where image features alone are insufficient.

Management Decisions versus Diagnostic Accuracy

A distinction that several prospective studies surface but that the broader literature tends to conflate is the difference between diagnostic classification accuracy and clinical management utility. Menzies et al. (2023) is the most explicit on this point: while the 7-class AI achieved specialist-equivalent diagnostic accuracy, all but two of five management AI configurations were significantly inferior to specialist clinicians for correct management decisions in the whole-body examination setting [8]. The authors estimated this could mean approximately 8 mismanaged lesions per 20-patient high-risk clinic.

The Marsden et al. (2024) study, conversely, demonstrated that the UKCA-approved AIaMD, when used for triage management (refer urgently vs discharge), was not merely equivalent but significantly better than teledermatology in reducing unnecessary referrals while maintaining comparable malignancy sensitivity [10]. The key differentiating factor is the nature of the management task. Binary triage (flag for cancer pathway or not) is a more tractable problem for current AI than nuanced clinical management incorporating patient preferences, lesion context, competing diagnoses, and treatment options. The evidence supports AI for the former and urges caution for the latter.

Regulatory Maturity and Real-World Evidence

In terms of regulatory status, only a subset of the AI systems in this review have obtained formal clearance. Three systems reviewed in the UK-based deployment studies had UKCA Class IIa approval [3–5, 10, 27], and the market-approved Moleanalyzer Pro (FotoFinder) carried a CE mark [52, 57]. The majority of systems described in this literature are prototype or early clinical validation-stage tools without regulatory clearance [17–19, 22, 29, 32, 34, 49]. Matin & Dinnes (2023) explicitly criticized the regulatory environment, noting that approvals have been granted without prospective clinical trial data requirements [9], a concern reinforced by the Sollis et al. (2026) umbrella review finding that prospective validation accounts for approximately 2% of primary studies in this field [6].

The evidence is therefore most mature for the specific use case of NHS teledermatology triage for urgent suspected skin cancer pathways, where three UK deployment studies provide convergent evidence of high sensitivity (97–99%), high negative predictive value (99.8%), and meaningful efficiency gains [3–5]. For all other use cases—primary care differential diagnosis, consumer screening, chronic disease monitoring, and management decision support—the evidence base ranges from promising to severely limited, and in the one RCT assessing consumer screening, AI access increased healthcare costs without a statistically significant improvement in cancer detection [7].

The domain where the evidence most strongly supports clinical implementation is accordingly narrow: UKCA/CE-approved AI systems operating within specialist teledermatology pathways for urgent skin cancer triage, in Fitzpatrick I–III populations, as adjuncts to clinician review rather than autonomous decision-makers. For all other configurations, the appropriate next step is prospective validation across demographically diverse populations, followed by regulatory review, before deployment at scale.

DISCUSSION

This systematic review synthesises evidence from 80 primary studies spanning to 2026, revealing several significant positive findings alongside critical implementation caveats.

Significant positive finding 1: AI achieves non-inferiority and occasionally superiority to dermatologists in controlled melanoma detection. Multiple high-quality studies demonstrate that deep learning algorithms match or exceed dermatologist performance when evaluating dermoscopic images of pigmented lesions. Esteva et al. (2017) first established dermatologist-level classification with AUCs of 0.94–0.96 (74). Haenssle et al. (2018) demonstrated significantly higher CNN specificity (82.5% vs 71.3%, $p<0.01$) with AUC of 0.86 versus dermatologists' 0.79 ($p<0.01$) (17). Brinker et al. (2019) provided the first statistically significant superiority finding: CNN sensitivity 82.3% vs 67.2% and specificity 77.9% vs 62.2% (both McNemar $p<0.001$) (22). The ISIC 2018 challenge confirmed that top algorithms outperformed all human readers by a mean of 2.01 correct diagnoses per 30-image batch and exceeded experts with >10 years' experience by 6.65 correct diagnoses ($p<0.0001$) (43). A meta-analysis by Ye et al. (2024) pooling 37 studies reported pooled sensitivity 82%, specificity 87%, and AUC 0.92 for deep learning algorithms, with human-machine comparison showing DL AUC 0.87 versus senior dermatologists' 0.83 (37). Laiouar-Pedari et al. (2026), restricting to prospective studies only, found AI alone achieving comparable performance to dermatologists (sensitivity 80.9% vs 78.6%; specificity 75.6% vs 75.2%) (14).

Significant positive finding 2: AI assistance significantly improves clinician diagnostic accuracy, particularly for non-specialists. This represents the most consistent and clinically actionable finding across the literature. Sollis et al. (2026) in their umbrella review quantified that generalists gain approximately +27.9 percentage points in sensitivity with AI augmentation, while expert dermatologists gain only +2.1 points (6). The only published RCT in this field, by Han et al. (2022), demonstrated that AI assistance improved non-dermatology trainee accuracy from 30.7% to 54.7% ($p<0.0001$), with no significant benefit for dermatology residents who had greater baseline competence (23). Similarly, Han et al. (2020) showed that AI assistance increased malignancy detection sensitivity by +12.1% for dermatologists and residents ($p<0.0001$) and by +83.8% for the general public ($p<0.0001$) (15). Liu et al. (2019) demonstrated that a deep learning system for differential diagnosis of 26 conditions achieved top-1 accuracy of 0.66, comparable to dermatologists (0.63) and superior to primary care physicians (0.44, $p<0.001$) and nurse practitioners (0.40, $p<0.001$) (16). Weil Ba et al. (2022) found that CNN assistance improved board-certified dermatologist accuracy from 62.78% to 76.60% ($p<0.001$) with kappa increasing from 0.59 to 0.74 ($p<0.001$) (42). Jain et al. (2021) reported that AI assistance improved primary care physician top-1 agreement from 48% to 58% ($p<0.001$) and nurse practitioners from 46% to 58% ($p<0.001$) for 120 skin conditions (11). The expertise-dependent gradient is mechanistic: AI provides the largest diagnostic gain where baseline performance is lowest, suggesting optimal deployment in primary care and teledermatology triage rather than specialist clinics.

Significant positive finding 3: Real-world NHS deployment demonstrates clinically meaningful utility for urgent skin cancer triage. The most mature evidence comes from prospective multicentre studies of UKCA Class IIa-approved AI medical devices operating within UK National Health Service teledermatology pathways. Luck et al. (2025), reporting on 25,788 lesions across 12 NHS sites, found that the AIaMD triaged 98.6% (95% CI 97.3–99.2%) of high-risk skin cancers to the urgent suspected cancer pathway, significantly higher than teledermatologists at 95.9% (95% CI 94.0–97.2%; $p=0.004$). For melanoma specifically, sensitivity was 98.1% versus 94.7% for teledermatologists ($p=0.04$); for SCC, 99.2% versus 96.6% ($p=0.02$) (3). Brewer et al. (2024) reported 97% melanoma sensitivity (37/38), 96.6% all skin cancer sensitivity (282/292), and a negative predictive value of 99.8%, with associated reductions of 11 days in mean wait time, 10% fewer biopsies, and 13% fewer routine follow-up appointments (4). Jenkins et al. (2023) reported 100% melanoma sensitivity (51/51) and NPV of 99.8% across 1,363 lesions, with conversion rate improvement from 9.9% to 13.8% (5). Marsden et al. (2024) calculated that if AI outputs had been used for triage, 454 teledermatology reviews, 141 face-to-face assessments, and 124 biopsies would have been avoided per 1,000 patients, with estimated savings of £156,064 and 259 specialist hours. The number needed to biopsy to detect one malignancy was 3.0 (95% CI 2.4–3.7) for AI versus 4.2 (95% CI 3.3–5.5) for teledermatology standard of care (10). These convergent findings from three independent NHS deployment studies provide robust real-world evidence that AI can safely increase pathway efficiency while maintaining or improving sensitivity.

Significant positive finding 4: AI shows promise for non-melanoma skin cancer and inflammatory disease severity assessment. For basal cell carcinoma, Liu et al. (2025) reported pooled sensitivity 0.96, specificity 0.98, and AUC 0.99 using dermoscopy-based deep learning, with significantly higher AUC than dermatologists (z=2.63, p=0.008), though external validation performance was lower, indicating a generalisability gap (44). Marsden et al. (2023) demonstrated AI identification of BCC (sensitivity 94%, specificity 27.6% for biopsy-confirmed benign lesions) and SCC (sensitivity 98%) (27). For psoriasis severity, Huang et al. (2023) developed and validated an AI PASI scoring model that outperformed the average of 43 dermatologists by 33.2% in total PASI score accuracy (AI MAE 3.12 vs dermatologist MAE 4.67), achieving trend accuracy of 84.81% across visit pairs and deployed via WeChat app to 1,497 patients across 18 hospitals (51). The meta-analysis by Cai et al. (2025) on AI severity assessment across inflammatory skin diseases reported pooled sensitivity 80.5% (95% CI 76.2–84.2) and specificity 96.2% (95% CI 94.9–97.2), with atopic dermatitis severity assessment achieving 91.8% sensitivity versus 71.1% for psoriasis (p=0.044) (2). For rare diseases, Sharaf et al. (2025) showed that a self-supervised learning model for mycosis fungoides on H&E whole slide images achieved 96–98% accuracy and 94–100% sensitivity across validation datasets (72). The mSWAT-Net model for mycosis fungoides tumour burden achieved intraclass correlation coefficients of 0.917 (internal) and 0.846 (temporal) compared to cutaneous lymphoma specialists, outperforming junior dermatologists (ICC 0.777) (69).

Significant positive finding 5: Novel imaging modalities and ensemble methods enhance AI performance. Combining multiple imaging inputs or fusing AI with clinician judgment improves accuracy. Zhu et al. (2024) demonstrated that a multimodal fusion network combining clinical close-up images and high-frequency ultrasound outperformed monomodal CNN (AUC 0.876 vs 0.697, p=0.0063), general practitioners (0.876 vs 0.651, p=0.0025), and general dermatologists (0.876 vs 0.838, p=0.0038), achieving near-equivalent performance to HFUS-specialised dermatologists (0.876 vs 0.891, p=0.0080) (62). Fischman et al. (2025) reported that AI-assisted line-field confocal OCT increased BCC diagnostic sensitivity by +25.8 percentage points and specificity by +16.8 percentage points compared to clinical and dermoscopic images alone, effectively bridging approximately 2 years of LC-OCT experience among dermatologists (46). Hekler et al. (2019) demonstrated that combining dermatologist and CNN outputs via gradient boosting increased multiclass accuracy from 81.59% (CNN alone) to 82.95% (fusion), improving binary task sensitivity to 89% (39). Soenksen et al. (2021) showed a DCNN operating on wide-field consumer smartphone images without dermoscopy attachments achieved sensitivity 90.3% and specificity 89.9% for suspicious pigmented lesion identification (36).

Equity concern: consistent underperformance on darker skin types. A statistically significant and clinically concerning finding is that AI systems perform worse on Fitzpatrick IV–VI skin types. Tjiu and Lu (2025) demonstrated a pooled AUROC of 0.89 for Fitzpatrick I–III versus 0.82 for IV–VI (Δ −0.07, p<0.01) (1). Gulati (2025) reported accuracy declining from approximately 90–92% for types I–III to 81–83% for types V–VI (78). Uwishema et al. (2025) identified an extreme example: one dermatology app achieved approximately 17% accuracy on Black skin images versus 69.9% on Caucasian images (13). Sollis et al. (2026) found that over 70% of training datasets were derived from light-skinned populations, with few datasets including skin-tone annotations (6). In contrast, Jain et al. (2022) reported that a differential diagnosis AI tool improved clinician agreement by 8–15% across all racial and ethnic subgroups, with performance spread of only 5.2–6.6%, suggesting that product design choices substantially affect equity outcomes (12).

The consumer AI cautionary evidence: The only randomised controlled trial of consumer-facing AI for skin cancer screening (Smak Gregoor et al., 2025, n=19,009) found no statistically significant increase in histologically confirmed skin cancer detection (2.66% vs 2.27%, RD 0.39%, p=0.10) but significantly more claims for benign skin lesions (3.9% vs 2.6%, p<0.001), more surgical interventions, and higher per-participant costs (€63 vs €47, p<0.001). No reduction in melanoma Breslow thickness or stage shift was demonstrated at 12 months (7). This represents the highest level of evidence for population-level consumer AI effects and demonstrates that—at current accuracy—AI access increases healthcare utilisation for benign conditions without proven cancer detection benefit.

CONCLUSIONS AND RECOMMENDATIONS

Conclusions: This systematic review of 80 studies (2017–2026) demonstrates that artificial intelligence has emerged from experimental image classification to clinically deployed medical devices in dermatology, with several significant positive findings supported by high-quality evidence.

First, AI achieves diagnostic accuracy for melanoma detection that is non-inferior to dermatologists in controlled settings using dermoscopic images, with meta-analyses showing pooled sensitivity of 82% and AUC of 0.92. Second, AI assistance provides substantial diagnostic benefit to non-specialist clinicians (approximately +27.9 percentage points in sensitivity), far exceeding the modest benefit to expert dermatologists (+2.1 points), indicating optimal deployment in primary care and teledermatology triage rather than specialist clinics. Third, prospective real-world NHS deployment studies of UKCA-approved AI medical devices demonstrate clinically meaningful utility: 98.6% sensitivity for high-risk skin cancer triage, 99.8% negative predictive value, reductions of 11 days in wait times and 10% in biopsies, and estimated savings of £156,064 per 1,000 patients. Fourth, AI shows promising performance for non-melanoma skin cancer (BCC pooled AUC 0.99) and inflammatory disease severity assessment (psoriasis PASI scoring outperforming 43 dermatologists by 33.2%). Fifth, novel imaging modalities (high-frequency ultrasound, LC-OCT) and clinician-AI ensembles further improve diagnostic accuracy.

However, these positive findings must be balanced against critical limitations. There is a consistent gap between controlled experimental performance (where AI appears superior) and real-world prospective validation (where AI often underperforms specialists). Consumer-facing AI, evaluated in the only randomised controlled trial to date, increased healthcare costs without a statistically significant improvement in cancer detection. Most concerning, AI systems perform significantly worse on darker skin types (Fitzpatrick IV–VI: AUROC 0.82 vs 0.89 for I–III, $p<0.01$), reflecting over 70% light-skinned training datasets.

Recommendations:

1. **Clinical implementation:** Deploy UKCA/CE-approved AI only within specialist teledermatology urgent suspected cancer pathways as adjuncts to clinician review, not as autonomous decision-makers. Current evidence does not support consumer-facing AI screening or autonomous primary care diagnosis.
2. **Regulatory requirements:** Mandate prospective, multicentre validation across diverse skin type distributions (minimum 30% Fitzpatrick IV–VI) before regulatory approval. Require equity reporting as a condition of market access.
3. **Clinical workflow:** Target AI deployment to primary care and non-specialist settings where the augmentation benefit is largest. For specialist clinics, AI provides marginal additional diagnostic gain but may improve efficiency through triage.
4. **Research priorities:** Conduct prospective trials of AI across skin type-stratified populations. Develop and validate AI systems on balanced, annotated datasets including dermatoscopic, clinical, and consumer image types. Evaluate long-term outcomes including melanoma stage shift and mortality reduction, not only diagnostic accuracy.

REFERENCES

1. Tjiu J, Lu C-F (2025) Equity and Generalizability of Artificial Intelligence for Skin-Lesion Diagnosis Using Clinical, Dermoscopic, and Smartphone Images: A Systematic Review and Meta-Analysis. Medicina. https://doi.org/10.3390/medicina61122186

2. Cai Z, Kim J, Rezaei S, et al (2025) Assessing the performance of artificial intelligence models in evaluating inflammatory skin disease severity: a systematic review and meta-analysis. British Journal of Dermatology. https://doi.org/10.1093/bjd/ljaf250

3. Luck J, Kalsi D, Mullarkey D, et al (2025) AI12 Prospective, multicentre, real-world sensitivities of an artificial intelligence as a medical device assessment and teledermatologist assessment in appropriately managing melanoma, squamous cell carcinoma and rare skin cancers referred on the urgent suspected cancer pathway. British Journal of Dermatology. https://doi.org/10.1093/bjd/ljaf085.209

4. Brewer F, Kentley J, Kalsi D, et al (2024) BT09 Multicentre prospective clinical performance analysis of an artificial intelligence as a medical device deployed within urgent suspected skin cancer pathways. British Journal of Dermatology. https://doi.org/10.1093/bjd/ljae090.407

5. Jenkins R, Brewer CF, Kalsi D, Mullarkey D (2023) BT09 Clinical performance of an artificial intelligence-based medical device deployed within an urgent suspected skin cancer pathway. British Journal of Dermatology. https://doi.org/10.1093/bjd/ljad113.375

6. Sollis L, Bunnell A, Cho E, et al (2026) How is Artificial Intelligence Transforming the Skin Cancer Screening Pathway? An Umbrella Review. Research Square. https://doi.org/10.21203/rs.3.rs-9069373/v1

7. Gregoor ASS, Sangers T, Groot CU, et al (2025) The impact of a SmartPhone applicatiOn for skin cancer risk assessmenT on the healthcare system (SPOT-study): A randomized controlled trial. medRxiv. https://doi.org/10.1101/2025.11.18.25340297

8. Menzies S, Sinz C, Menzies M, et al (2023) 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. The Lancet Digital Health. https://doi.org/10.1016/s2589-7500(23)00130-9

9. Matin RN, Dinnes J (2023) Diagnosis of suspicious pigmented lesions in specialist settings with artificial intelligence. The Lancet Digital Health. https://doi.org/10.1016/s2589-7500(23)00180-2

10. Marsden H, Kemos P, Venzi M, et al (2024) Accuracy of an artificial intelligence as a medical device as part of a UK-based skin cancer teledermatology service. Frontiers in Medicine. https://doi.org/10.3389/fmed.2024.1302363

11. Jain A, Way DH, Gupta V, et al (2021) Development and Assessment of an Artificial Intelligence–Based Tool for Skin Condition Diagnosis by Primary Care Physicians and Nurse Practitioners in Teledermatology Practices. JAMA Network Open. https://doi.org/10.1001/jamanetworkopen.2021.7249

12. Jain A, Way DH, Gupta V, et al (2022) Race- and Ethnicity-Stratified Analysis of an Artificial Intelligence–Based Tool for Skin Condition Diagnosis by Primary Care Physicians and Nurse Practitioners. Iproceedings. https://doi.org/10.2196/36885

13. Uwishema O, Ghezzawi M, Charbel N, et al (2025) Diagnostic performance of artificial intelligence for dermatological conditions: a systematic review focused on low- and middle-income countries to address resource constraints and improve access to specialist care. International Journal of Emergency Medicine. https://doi.org/10.1186/s12245-025-00975-4

14. Laiouar-Pedari S, Kühn A, Wies C, et al (2026) Prospective Evidence on Artificial Intelligence−Assisted Melanoma Diagnostics. JAMA dermatology. https://doi.org/10.1001/jamadermatol.2026.0217

15. Han S, Park I, Chang SE, et al (2020) Augment Intelligence Dermatology : Deep Neural Networks Empower Medical Professionals in Diagnosing Skin Cancer and Predicting Treatment Options for 134 Skin Disorders. Journal of Investigative Dermatology. https://doi.org/10.1016/j.jid.2020.01.019

16. Liu Y, Jain A, Eng C, et al (2019) A deep learning system for differential diagnosis of skin diseases. Nature Medicine. https://doi.org/10.1038/s41591-020-0842-3

17. Haenssle H, Fink C, Schneiderbauer R, et al (2018) Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. Annals of Oncology. https://doi.org/10.1093/annonc/mdy166

18. Phillips M, Greenhalgh J, Marsden H, Palamaras I (2019) Detection of Malignant Melanoma Using Artificial Intelligence: An Observational Study of Diagnostic Accuracy. Dermatology Practical & Conceptual. https://doi.org/10.5826/dpc.1001a11

19. Maron RC, Utikal J, Hekler A, et al (2020) Artificial Intelligence and Its Effect on Dermatologists' Accuracy in Dermoscopic Melanoma Image Classification: Web-Based Survey Study. Journal of Medical Internet Research. https://doi.org/10.2196/18091

20. Majidian M, Tejani I, Jarmain T, et al (2022) Artificial Intelligence in the Evaluation of Telemedicine Dermatology Patients. Journal of Drugs in Dermatology. https://doi.org/10.36849/jdd.6277

21. Thompson H, Oakley A, Jameson M, Bowling A (2021) Artificial Intelligence Support for Skin Lesion Triage in Primary Care and Dermatology (Preprint). Iproceedings. https://doi.org/10.2196/preprints.35395

22. Brinker T, Hekler A, Enk A, et al (2019) Deep neural networks are superior to dermatologists in melanoma image classification. European Journal of Cancer. https://doi.org/10.1016/j.ejca.2019.05.023

23. Han S, Kim YJ, Moon I, et al (2022) Evaluation of Artificial Intelligence-assisted Diagnosis of Skin Neoplasms - a single-center, paralleled, unmasked, randomized controlled trial. Journal of Investigative Dermatology. https://doi.org/10.1016/j.jid.2022.02.003

24. Marri S, Albadri W, Hyder MS, et al (2023) Efficacy of an Artificial Intelligence App (Aysa) in Dermatological Diagnosis: Cross-Sectional Analysis. JMIR Dermatology. https://doi.org/10.2196/48811

25. Marri S, Inamadar A, Janagond AB, Albadri W (2023) Analyzing the Predictability of an Artificial Intelligence App (Tibot) in the Diagnosis of Dermatological Conditions: A Cross-sectional Study. JMIR Dermatology. https://doi.org/10.2196/45529

26. Escalé-Besa A, Yélamos O, Vidal-Alaball J, et al (2023) Exploring the potential of artificial intelligence in improving skin lesion diagnosis in primary care. Scientific Reports. https://doi.org/10.1038/s41598-023-31340-1

27. Marsden H, Morgan C, Austin S, et al (2023) Effectiveness of an image analyzing AI-based Digital Health Technology to identify Non-Melanoma Skin Cancer and other skin lesions: results of the DERM-003 study. Frontiers in Medicine. https://doi.org/10.3389/fmed.2023.1288521

28. Maron RC, Weichenthal M, Utikal J, et al (2019) Systematic outperformance of 112 dermatologists in multiclass skin cancer image classification by convolutional neural networks. European Journal of Cancer. https://doi.org/10.1016/j.ejca.2019.06.013

29. Papachristou P, Söderholm M, Pallon J, et al (2024) Evaluation of an artificial intelligence-based decision support for detection of cutaneous melanoma in primary care - a prospective, real-life, clinical trial. British Journal of Dermatology. https://doi.org/10.1093/bjd/ljae021

30. Navarrete-Dechent C (2022) Teledermatology and Artificial Intelligence (Preprint). https://doi.org/10.2196/preprints.36894

31. Cerminara S, Cheng P, Kostner L, et al (2023) Diagnostic performance of augmented intelligence with 2D and 3D total body photography and convolutional neural networks in a high-risk population for melanoma under real-world conditions: A new era of skin cancer screening? European Journal of Cancer. https://doi.org/10.1016/j.ejca.2023.112954

32. Heinlein L, Maron RC, Hekler A, et al (2024) Prospective multicenter study using artificial intelligence to improve dermoscopic melanoma diagnosis in patient care. Communications Medicine. https://doi.org/10.1038/s43856-024-00598-5

33. Cartocci A, Luschi A, Tognetti L, et al (2024) Comparative Analysis of AI Models for Atypical Pigmented Facial Lesion Diagnosis. Bioengineering. https://doi.org/10.3390/bioengineering11101036

34. Marchetti M, Cowen E, Kurtansky N, et al (2023) Prospective validation of dermoscopy-based open-source artificial intelligence for melanoma diagnosis (PROVE-AI study). npj Digit Medicine. https://doi.org/10.1038/s41746-023-00872-1

35. Díaz-Ramón JL, Gardeazabal J, Izu R, et al (2023) Melanoma Clinical Decision Support System: An Artificial Intelligence-Based Tool to Diagnose and Predict Disease Outcome in Early-Stage Melanoma Patients. Cancers. https://doi.org/10.3390/cancers15072174

36. Soenksen L, Kassis T, Conover S, et al (2021) Using deep learning for dermatologist-level detection of suspicious pigmented skin lesions from wide-field images. Science Translational Medicine. https://doi.org/10.1126/scitranslmed.abb3652

37. Ye Z, Zhang D, Zhao Y, et al (2024) Deep learning algorithms for melanoma detection using dermoscopic images: A systematic review and meta-analysis. Artif Intell Medicine. https://doi.org/10.1016/j.artmed.2024.102934

38. Kommoss K, Winkler J, Mueller-Christmann C, et al (2023) Observational study investigating the level of support from a convolutional neural network in face and scalp lesions deemed diagnostically 'unclear' by dermatologists. European Journal of Cancer. https://doi.org/10.1016/j.ejca.2023.02.025

39. Hekler A, Utikal J, Enk AH, et al (2019) Superior skin cancer classification by the combination of human and artificial intelligence. European Journal of Cancer. https://doi.org/10.1016/j.ejca.2019.07.019

40. Marsden H, Palamaras I, Kemos P, Greenhalgh J (2023) BT03 (P63) Effectiveness of an image-analysing artificial intelligence-based digital health technology to diagnose nonmelanoma skin cancer and benign skin lesions. British Journal of Dermatology. https://doi.org/10.1093/bjd/ljad113.369

41. Navarrete-Dechent C (2022) Teledermatology and Artificial Intelligence. Iproceedings. https://doi.org/10.2196/36894

42. Ba W, Wu H, Chen W, et al (2022) Convolutional neural network assistance significantly improves dermatologists' diagnosis of cutaneous tumours using clinical images. European Journal of Cancer. https://doi.org/10.1016/j.ejca.2022.04.015

43. Tschandl P, Codella N, Akay BN, et al (2019) Comparison of the accuracy of human readers versus machine-learning algorithms for pigmented skin lesion classification: an open, web-based, international, diagnostic study. The Lancet Oncology. https://doi.org/10.1016/S1470-2045(19)30333-X

44. Liu H, Shang G, Shan Q (2025) Deep Learning Algorithms in the Diagnosis of Basal Cell Carcinoma Using Dermatoscopy: Systematic Review and Meta-Analysis. Journal of Medical Internet Research. https://doi.org/10.2196/73541

45. Winkler JK, Kommoss K, Vollmer AS, et al (2024) Computerizing the first step of the two-step algorithm in dermoscopy: A convolutional neural network for differentiating melanocytic from non-melanocytic skin lesions. European Journal of Cancer. https://doi.org/10.1016/j.ejca.2024.114297

46. Fischman S, Viel T, Perrot J, et al (2025) AI-assisted basal cell carcinoma diagnosis with LC-OCT: A multicentric retrospective study. Journal of the European Academy of Dermatology and Venereology. https://doi.org/10.1111/jdv.70099

47. Kips J, Papeleu J, Shen A, et al (2026) Artificial intelligence-based smartphone application for skin cancer detection: A prospective diagnostic accuracy study. British Journal of Dermatology. https://doi.org/10.1093/bjd/ljag057

48. Yip S, Parmar D, Ruchlin I, et al (2025) 338P Leveraging NSCLC foundation models to predict biomarker testing pattern. ESMO Real World Data and Digital Oncology. https://doi.org/10.1016/j.esmorw.2025.100534

49. Jinnai S, Yamazaki N, Hirano Y, et al (2020) The Development of a Skin Cancer Classification System for Pigmented Skin Lesions Using Deep Learning. Biomolecules. https://doi.org/10.3390/biom10081123

50. Rikhye R, Loh A, Hong G, et al (2025) Closing the AI generalisation gap by adjusting for dermatology condition distribution differences across clinical settings. EBioMedicine. https://doi.org/10.1016/j.ebiom.2025.105766

51. Huang K, Wu X, Li Y, et al (2023) Artificial Intelligence–Based Psoriasis Severity Assessment: Real-world Study and Application. Journal of Medical Internet Research. https://doi.org/10.2196/44932

52. Haenssle H, Fink C, Toberer F, et al (2020) Man against machine reloaded: performance of a market-approved convolutional neural network in classifying a broad spectrum of skin

lesions in comparison with 96 dermatologists working under less artificial conditions. Annals of Oncology. https://doi.org/10.1016/j.annonc.2019.10.013

53. Han S, Moon I, Lim W, et al (2019) Keratinocytic Skin Cancer Detection on the Face Using Region-Based Convolutional Neural Network. JAMA dermatology. https://doi.org/10.1001/jamadermatol.2019.3807

54. Haenssle H, Winkler J, Fink C, et al (2020) Skin lesions of face and scalp - Classification by a market-approved convolutional neural network in comparison with 64 dermatologists. European Journal of Cancer. https://doi.org/10.1016/j.ejca.2020.11.034

55. Tschandl P, Rosendahl C, Akay BN, et al (2019) Expert-Level Diagnosis of Nonpigmented Skin Cancer by Combined Convolutional Neural Networks. JAMA dermatology. https://doi.org/10.1001/jamadermatol.2018.4378

56. Han S, Moon I, Kim SH, et al (2020) Assessment of deep neural networks for the diagnosis of benign and malignant skin neoplasms in comparison with dermatologists: A retrospective validation study. PLoS Medicine. https://doi.org/10.1371/journal.pmed.1003381

57. Winkler J, Blum A, Kommoss K, et al (2023) Assessment of Diagnostic Performance of Dermatologists Cooperating With a Convolutional Neural Network in a Prospective Clinical Study: Human With Machine. JAMA dermatology. https://doi.org/10.1001/jamadermatol.2023.0905

58. Okata-Karigane U, Hirota M, Takahashi C, et al (2025) AI-Based Objective Severity Assessment of Atopic Dermatitis Using Patient Photos in a Real-World Setting: A Digital Biomarker Approach. Allergy European Journal of Allergy and Clinical Immunology. https://doi.org/10.1111/all.16586

59. Boostani M, Bánvölgyi A, Goldust M, et al (2025) Diagnostic Performance of GPT-4o and Gemini Flash 2.0 in Acne and Rosacea. International Journal of Dermatology. https://doi.org/10.1111/ijd.17729

60. Sánchez-Viera M, Medela A, Campo I, et al (2025) Diagnostic accuracy in detecting malignancy in suspicious skin lesions using Artificial Intelligence. medRxiv. https://doi.org/10.1101/2025.03.11.25323753

61. Ren Y-Y, Mei L-H, Liu X-D, et al (2025) Enhancing dermatological diagnosis for differentiating actinic from seborrheic keratosis using deep learning model. Frontiers in Medicine. https://doi.org/10.3389/fmed.2025.1654813

62. Zhu A, Wang Q, Shi Y-L, et al (2024) A deep learning fusion network trained with clinical and high-frequency ultrasound images in the multi-classification of skin diseases in comparison with dermatologists: a prospective and multicenter study. EClinicalMedicine. https://doi.org/10.1016/j.eclinm.2023.102391

63. Mei L-H, Cao M, Li J, et al (2025) Deep learning in assisting dermatologists in classifying basal cell carcinoma from seborrheic keratosis. Frontiers in Oncology. https://doi.org/10.3389/fonc.2025.1507322

64. Coates SJ, Yang F, Hill C, et al (2025) Artificial Intelligence-based Diagnosis of Kaposi Sarcoma using Photographs in Dark-skinned Patients. medRxiv. https://doi.org/10.1101/2025.04.21.25326060

65. Coates SJ, Yang F, Hill C, et al (2026) Artificial Intelligence–Based Diagnosis of Kaposi Sarcoma Using Digital Photographs in Dark-Skinned Patients in Uganda. JCO Global Oncology. https://doi.org/10.1200/GO-25-00236

66. Lee S-H, Kang H, Hong S-P, et al (2025) AI-augmented differential diagnosis of granulomatous rosacea and lupus miliaris disseminatus faciei: A 23–year retrospective pilot study. PLoS ONE. https://doi.org/10.1371/journal.pone.0326763

67. Bali R, Chung J, Thomas L, et al (2024) BT05 Comparing the number-needed-to-biopsy ratio for melanoma diagnosis between teledermatology, an artificial intelligence device with teledermatology, and face-to-face models of care. British Journal of Dermatology. https://doi.org/10.1093/bjd/ljae090.403

68. Lay FL, Barzilay O, Cerasuolo D, et al (2024) Assessing diagnostic performance for common skin diseases using an AI-assisted tele-expertise platform: a proof of concept. EJD European journal of dermatology. https://doi.org/10.1684/ejd.2024.4804

69. Wang H, Liu Z, Pan H, et al (2025) A 3D-Optimized AI Imaging Model for the Skin Tumor Burden Assessment of Mycosis Fungoides. Journal of Investigative Dermatology. https://doi.org/10.1016/j.jid.2025.06.1567

70. Islam S, Wishart G, Walls J, et al (2026) Advancing skin cancer detection through deep learning and fusion of patient metadata and skin lesion images. Scientific Reports. https://doi.org/10.1038/s41598-025-26392-4

71. Oloruntoba I, Vestergaard T, Nguyen TD, et al (2021) Assessing Generalisability of Deep Learning Models Trained on Standardised and Non-Standardised Images and their Performance against Tele-dermatologists (Preprint). https://doi.org/10.2196/preprints.35150

72. Sharaf M, Shvartsbeyn M, Giubellino A, Park CY (2025) Diagnosing mycosis fungoides using AI powered histologic analysis. Blood. https://doi.org/10.1182/blood-2025-2570

73. Martínez-Vargas E, Mora-Jiménez J, Arguedas-Chacón S, et al (2025) The Emerging Role of Artificial Intelligence in Dermatology: A Systematic Review of Its Clinical Applications. Dermato. https://doi.org/10.3390/dermato5020009

74. Esteva A, Kuprel B, Novoa R, et al (2017) Dermatologist-level classification of skin cancer with deep neural networks. Nature. https://doi.org/10.1038/nature21056

75. Hollman D, Doktorchik C, Mukovozov I (2024) Emerging Uses of Artificial Intelligence in Chronic Dermatologic Disease: A Scoping Review. Journal of Cutaneous Medicine and Surgery. https://doi.org/10.1177/12034754241308237

76. Felmingham C, MacNamara S, Cranwell W, et al (2022) Improving Skin cancer Management with ARTificial Intelligence (SMARTI): protocol for a preintervention/postintervention trial of an artificial intelligence system used as a diagnostic aid for skin cancer management in a specialist dermatology setting. BMJ Open. https://doi.org/10.1136/bmjopen-2021-050203

77. Furriel BC, Oliveira B, Prôa R, et al (2024) Artificial intelligence for skin cancer detection and classification for clinical environment: a systematic review. Frontiers in Medicine. https://doi.org/10.3389/fmed.2023.1305954

78. Gulati DS (2025) The Role Of Artificial Intelligence In Dermatology: Enhancing The Accuracy Of Skin Cancer Diagnosis. African Journal of Biomedical Research. https://doi.org/10.53555/ajbr.v28i1s.6715

79. Jairath NK, Pahalyants V, Shah R, et al (2024) Artificial Intelligence in Dermatology: A Systematic Review of Its Applications in Melanoma and Keratinocyte Carcinoma Diagnosis. Dermatologic Surgery. https://doi.org/10.1097/DSS.0000000000004223

80. Ali A, Goyal B, Monescu V, et al (2025) Artificial intelligence in dermatology: a systematic review of skin cancer detection and classification. Engineering Research Express. https://doi.org/10.1088/2631-8695/ae1933